DOCUMENT RESUME

ED 036 664 24 AA 000 502

AUTHOR Bereiter, Carl; And Others
TITLE Curriculum Development and Evaluation: Research and Development Program on Preschool Disadvantaged Children. Final Report. (Volume II of III Volumes).
INSTITUTION Illinois Univ., Urbana. Inst. of Research for Exceptional Children.; Ontario Inst. for Studies in Education, Toronto.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-5-1181
PUB DATE May 69
CONTRACT OEC-6-10-235
NOTE 139p.

EDRS PRICE EDRS Price MF-$0.75 HC-$7.05
DESCRIPTORS Arithmetic Curriculum, Auditory Discrimination, *Compensatory Education Programs, Culturally Disadvantaged, Curriculum Design, *Curriculum Development, Exceptional Child Research, *Experimental Programs, Followup Studies, Intelligence Quotient, Kindergarten, Language Programs, Mathematical Models, Preschool Children, *Preschool Programs, Reading Programs, Verbal Learning
IDENTIFIERS Academically Oriented Preschool Program, Bereiter Engelmann, Direct Verbal Preschool, Stanford Binet

ABSTRACT

Seven studies were undertaken to further extend the development and testing of an academically-oriented preschool program for disadvantaged children. The studies investigated (1) Curricula Development and Testing in Bereiter-Engelmann Program, (2) Dual Kindergarten, (3) Follow-Up Data on the Achievement of Disadvantaged Children Who Participated in an Academically-Oriented Preschool Program, (4) Achievement Components of Stanford-Binet Performance, (5) A Model for the Interpretation of IQ Changes, (6) Verbal and Nonverbal Factors in Cultural Deprivation: Evidence from Children with Sensory Handicaps, and (7) The Performance of Advantaged and Disadvantaged Preschool Children on Tests of Sound Pattern and Speech Sound Auditory Discrimination. The hypothesis, procedure, population, method, and results are recorded for each study. Appendixes include a test on part-whole relationships, criterion tests, an achievement test, and sound pattern and speech sound discrimination tests. (DO)

FINAL REPORT
Volume II of III Volumes
Project No. 5-1181
Contract No. OE 6-10-235

RESEARCH AND DEVELOPMENT PROGRAM ON PRESCHOOL DISADVANTAGED CHILDREN

Curriculum Development and Evaluation

Principal Investigators:

Carl Bereiter
Ontario Institute for Studies in Education
Toronto, Ontario
Canada

and

Ernest D. Washington,
Rutgers University
New Brunswick, New Jersey

in collaboration with

Siegfried Engelmann

and

Jean Osborn
University of Illinois
Urbana, Illinois

May, 1969

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Merle B. Karnes, Project Director
Institute for Research on Exceptional Children
University of Illinois
Urbana, Illinois

PREFACE

The studies and data presented in this report aim at two objectives. First, is curriculum development and evaluation; second is research on social and psychological factors in the intellectual functioning of culturally disadvantaged children.

Under the overall direction of Dr. Merle Karnes a number of highly qualified investigators have pursued these objectives with diligence and ingenuity. Furthermore, they have, in important instances, pursued objectives to conclusions which are at the very least provocative. To many they will be startling and disturbing. Thus the sociologists, Farber, Lewis and Harvey conclude in Volume III:

> Technical emphasis in educational reform (particulary that which is intended for the dispossessed) may preclude any possibility of educators making a positive contribution to the obliteration of the social and economic injustices which victimize millions of Americans...Technical emphasis in education, as it is in welfare services, is a symptom of a condition which may be termed progressive status-quoism.

Volumes I and II deal largely with the first objective, curriculum development and evaluation, and as such are excellent examples of the highest quality of the "technical emphasis" to which the sociologists on the team refer. In Volume I, Karnes, Hodgins and Teska attack such concerns as the relative effectiveness of five differing methods of preschool educational intervention with the disadvantaged child. Other concerns are to determine how long such special intervention must be continued, the optimum age for intervention, and, how much can be done by paraprofessionals in the classroom and by mothers in the home.

In Volume II, Bereiter, Engelmann, Washington and colleagues describe efforts to burrow deeper into the processes and products of educational intervention on behalf of the disadvantaged. Taking the view that the Stanford-Binet may be considered as an achievement test for the "hidden curriculum" of the middle-class home, they boldly set about to construct a compensatory curriculum geared to the Binet, and to test the curriculum. In so doing they throw new light on the criticism that substantial I.Q. gains in programs for the disadvantaged are merely a result of "teaching for the test." In another section Bereiter grapples with the theoretical complexities of interpreting changes in I.Q.

Volume III deals almost exclusively with the description and analysis of family and kinship, neighborhood and community variables that bear on children's readiness and competence to enter into formal education. Farber examines this transition from home to school in the perspective of the necessity of articulation and accommodation of private and public cultures. He posits that where private and public culture clash those families and individuals whose way of life is incompatible with the public culture are superfluous population. Harvey describes life in a white, lower class, semi-rural community. Because his frame of reference is the same as that of Farber and Lewis, his findings extend the implications of the total report beyond the question of racial differences. Lewis presents a sociologically derived model and definition of "competence." For him, competence is a social dimension and in that perspective input from the family, neighborhood, and community sets severe limits on the part that formal schooling can play in the development of competence.

This is a multi-disciplinary multi-volume work which on the one hand undertook, with success, to add to our knowledge of educational curricula and techniques which enhance the academic performance of culturally disadvantaged children. On the other hand, an equally important objective was to inquire into factors which underlay the intellectual functioning of children. In these volumes we are confronted with the cruel paradox that acceptance of conclusions arrived at in pursuit of the second objective, raises grave doubts as to the value of present day endeavors aimed at the first objective. Resolution of this paradox will not be for the timid.

William P. Hurder
Director, Institute for
Research on Exceptional Children

ACKNOWLEDGEMENTS

Lest this become an itemlized extolment of all our friends or recent years, we must pass over the contributions of a number of dedicated and talented people who served valuably as teachers, students, supporting staff, or helpful colleagues and confine these acknowledgements to the several people who had a direct and major influence on the course and character of the project. They are Siegfried Engelmann (primum mobile), Jean Osborn, Elaine C. Bruner, Valerie Anderson, and Jessica Daniel.

Carl Bereiter
Ernest D. Washington

CONTENTS

Volume II

Tables. ix

INTRODUCTION 1

SUMMARY. 5

THE INDIVIDUAL STUDIES

1. Curricula Development and Testing in Bereiter-Engelmann Program by Siegfried Engelmann 13

2. The Dual Kindergarten by Ernest D. Washington and Jean Osborn 39

3. Follow-Up Data on the Achievement of Disadvantaged Children who Participated in an Academically-Oriented Preschool Program by Ernest D. Washington and Helen B. Bereiter 55

4. Achievement Components of Stanford-Binet Performance by Ernest D. Washington, Siefried Engelmann, and Carl Bereiter 69

5. A Model for the Interpretation of IQ Changes by Carl Bereiter. 91

6. Verbal and Nonverbal Factors in Cultural Deprivation: Evidence from Children with Sensory Handicaps by Carl Bereiter. 103

7. The Performance of Advantaged and Disadvantaged Preschool Children on Tests of Sound Pattern and Speech Sound Auditory Discrimination by Girvin E. Kirk and Carl Bereiter. 107

APPENDIXES 113

REFERENCES 153

TABLES

Volume II

Curriculum Development and Testing in Bereiter-Engelmann Program

1. Arithmetic achievement of groups I, II, and III after two years of instruction (Wide Range Achievement Test) 22

2. Reading achievement of groups I, II, and III after one and two years of instruction (Wide Range Achievement Test) 24

3. Stanford-Binet IQ scores for groups I, II, and III 28

The Dual Kindergarten

1. Group composition. 45

2. Unadjusted and adjusted mean grade placement scores on California Achievement Tests. 47

3. Means and standard deviation of the raw scores of the four subtests of reading vocabulary. 49

4. Unadjusted and adjusted mean scores from Stanford-Binet and ITPA 51

Follow-Up Data on the Achievement of Disadvantaged Children who Participated in an Academically-Oriented Preschool Program

1. California Achievement Tests, Lower Primary. 57

2. Stanford-Binet mental age. 57

3. ITPA Auditory-Vocal Automatic. 58

4. ITPA Visual Decoding 58

5. ITPA Motor Encoding. 59

6. ITPA Auditory-Vocal Association. 59

7. ITPA Visual-Motor Sequencing 60

8. ITPA Vocal Encoding. 60

9. ITPA Auditory-Vocal Sequencing 61

10. ITPA Visual-Motor Association. 61

11. ITPA Auditory Decoding 62

12. ITPA Total 62

Achievement Components of Stanford-Binet Performance

1. Similarities, differences, and incongruities 74

2. Initial group composition. 78

3. Testing schedule 79

4. Mean Stanford-Binet scores for the Binet group at the four intervals 80

5. WPPSI, pre- and post-test scores 80

6. IQ gains by intelligence strata. 82

7. Paired t tests on verbal, non-verbal, and total scores of the WPPSI 82

8. Results of prediction of success on Stanford-Binet items from performance on achievement test items 86

A Model for the Interpretation of IQ Changes

1. Hypothetical data illustrating IQ changes following a simple MA increments model. 93

2. Actual and predicted mean MAs and IQs for three preschool groups. 98

Verbal and Nonverbal Factors in Cultural Deprivation: Evidence from Children with Sensory Handicaps

1. Comparative data on deaf, blind, and partially-sighted children. 105

INTRODUCTION

The seven studies reported here grew out of the earlier development of "an academically-oriented preschool for disadvantaged children," described in the final report for Project Number 2129 (Contract No. OE 4-10-008), titled, Acceleration of Intellectual Development in Early Childhood. The first report describes subsequent modifications of the academically-oriented program and effects thereof. The second report deals with an administrative variation of the program, in which it was employed as a supplement to a regular kindergarten program rather than as a preliminary to or substitute for it. The third report presents follow-up findings through second grade for the original pilot group whose preschool attainments were reported in the earlier study. The fourth study employed a different experimental program designed to investigate achievement components in Stanford-Binet performance. The fifth report presents a simple mathematical model to account for IQ gains and losses during and following educational interventions. An earlier study had found a low correlation between child IQ and parent education among deaf children and the sixth study was intended to explore this relationship among visually-handicapped children. The expectation was that these children, being more than normally dependent on the language medium, would show unusually high correlations of IQ with parental education level. The final study attempts to clarify, in a pedagogically relevant way, the auditory discrimination difficulties that disadvantaged children have been claimed to possess.

SUMMARY

This project comprised seven studies having the general purpose of further extending development and testing of the academically-oriented preschool program for disadvantaged children and investigating questions related thereto. Procedures and results are summarized separately for each study.

1. Curricula Development and Testing in Bereiter-Engelmann Program.

Procedure: Over a three-year period programs in language, reading, and arithmetic were substantially revised through close observation of difficulties encountered by more advanced groups and try-out of changes on groups lagging behind them. Evaluation was by day-to-day curriculum-specific testing and by standardized end-of-course tests.

Results: There was a general tendency over the three years toward higher mean end-of-course achievement and toward a smaller number of low achievers. The development program led to many suggestions for more effective teaching.

2. The Dual Kindergarten

Procedure: Ten disadvantaged children who were given the academically-oriented program as a supplement to a regular kindergarten program were compared with 10 similar children who had only the regular program.

Results: Experimental children performed significantly better on measures of scholastic achievement, but not on IQ or total ITPA score.

3. Follow-Up Data on the Achievement of Disadvantaged Children Who Participated in an Academically-Oriented Preschool Program.

Procedure: Twelve surviving members of the original pilot group used in development of the academically-oriented program were studied in second grade by means of parent and teacher interviews, classroom observations, administration of the California Achievement Test, and readministration of the Stanford-Binet and ITPA.

Results: Five of the children were performing above grade level, three somewhat below grade level, and four well below grade level on the California Achievement Tests. Data from readministrations of the Stanford-Binet and ITPA showed that children in the lower achieving groups had not maintained their gains as well as the high achievers.

4/5/67

4. Achievement Components of Stanford-Binet Performance.

Procedure: In order to test the hypothesis that IQ gains were the result of learning of conceptual content tapped by the IQ test, a curriculum devised by working backward from Stanford-Binet items to define a universe of content for which the Stanford-Binet could serve as a content-valid achievement test was taught to 20 four-year-old disadvantaged children.

Results: IQ gains on the Stanford-Binet were of about the same magnitude as those obtained from the academically-oriented program and as those obtained on the WPPSI, a test whose content was kept hidden from teachers and curriculum writers. Attempts to predict IQ test performance from an inventory of specific curriculum attainments, while successful, suggested that the relation was general rather than specific. Results were thus contrary to the hypothesis of the study.

5. A Model for the Interpretation of IQ Changes.

Procedure: A mathematical model was developed and tested on data from current preschool experiments, which treats IQ gains during treatment and losses following termination of treatment as a function of expected IQ for the group, effectiveness of treatment, the extent to which induced IQ gains carry forward into increased potential for future growth, and the extent to which induced mental age increments are offset by subsequent reduced increments.

Results: The model provided a good fit to available data, a parsimonious interpretation of losses following termination of treatment, a pessimistic prognosis for the long-term survival of IQ gains, and a promising index of effectiveness of programs.

6. Verbal and Nonverbal Factors in Cultural Deprivation: Evidence from Children with Sensory Handicaps.

Procedure: A previous study had found a very low correlation between child IQ and parent education among deaf children, leading to the inference that language was the primary medium through which social-class related differences in intelligence are transmitted. The present study was conducted as a further check on this hypothesis, examining comparable correlations for visually-handicapped children. The expectation was that these children, being more than normally dependent on the language medium, would show unusually high correlations of IQ with parental education level.

Results: Correlations in five groups revealed no significant differences from the correlation obtained among the deaf. The appearance of high percentages of children with IQs below 70 suggested that organic intellectual impairment may have been responsible for the low correlations. It was concluded that children with sensory handicaps were not suitable subjects for testing hypotheses concerning experiential factors in mental growth.

7. The Performance of Advantaged and Disadvantaged Preschool Children on Tests of Sound Pattern and Speech Sound Auditory Discrimination.

Procedure: Fifteen disadvantaged Negro kindergarten children were compared with fifteen white children of similar age on their ability to discriminate speech and non-speech sounds, using a simplified test procedure.

Results: Advantaged children did better on discrimination of speech sounds than non-speech sounds; the reverse was true of disadvantaged children. Disadvantaged children had relatively much more difficulty discriminating final consonants than beginning consonants. Results were attributed to Negro dialect characteristics rather than to a generalized auditory discrimination deficit.

THE INDIVIDUAL STUDIES

Curricula Development and Testing in Bereiter-Engelmann Program

Siegfried Engelmann

The purposes of the curricula development component of the project were:

1. To provide teachers with more immediate feedback about teaching techniques.

2. To allow for the study of how children learn specific skills in greater detail.

3. To allow for a more articulate expression of how the general learning performance of a child (as measured by such instruments as the Stanford-Binet) relates to more specific performance on tasks taught in the program.

4. To ascertain which skills in the program are sequenced and presented relatively less efficiently than other skills in the program (which information would derive from the amount of time necessary for a child to reach a specific criterion of performance on a particular teaching objective).

The investigation, in summary, focused on two primary areas: one concerned with the behavior of the children in mastering specific objectives, the other concerned with the effectiveness of the curriculum, as determined by the relative difficulty of specific tasks, compared to other tasks.

PROCEDURE

General: The investigation centered around the three primary academic skills taught in the program -- reading, arithmetic, and language -- over the three-year period. During the school year of 1967-68, a more intensive attempt at developing systematic procedures for program evaluation and revision was conducted on the entering four-year-old children. The investigation focused on language skills. The amount of testing and observation time needed to carry out this undertaking mitigated against a broad investigation of all skills in reading, arithmetic, and language. It was felt that the investigators could do a more creditable study by working on the language component and conducting a more detailed evaluation, rather than attempting a more ambitious evaluation and sacrificing detailed data.

The evaluation of the "trouble shooting" program revisions was primarily informal but had several formal components. Formal instruments were used both to give a gross indication of how many skills the program taught during a given period and to document specifically how well children were taught specific tasks. The assumption was that if programs were relatively improved, there would be a performance improvement on achievement tests and on IQ tests (which tests would reflect the effectiveness of the language program). More specific measures (constructed for the experiment) were needed, however, to document the teaching and to provide the investigators with data about the variables associated with the process of teaching children so that they achieve a specific criterion of performance on a specific set of skills.

SUBJECTS

Subjects, as noted above, were four-year-old disadvantaged children (group III in the B-E program), who attended classes for 2½ hours a day in the mornings for two years. Children were grouped homogeneously, initially on the basis of IQ scores. They were then re-grouped on the basis of performance within the language program. One teacher worked **exclusively** with each group.

Informal Evaluation:

The informal evaluation consisted primarily of "brain storming" sessions in which the teachers compared observations on the relative difficulty of particular tasks included in the program. The teachers were not primarily responsible for making curricula changes, merely for noting problem spots. If the four teachers seemed to agree that a particular task or series of tasks was "difficult," their judgement was taken as prima facie evidence that the skill was poorly programmed.

A curriculum writer then made changes that were designed to remedy the situation. He made changes either by (1) lengthening the sequence and breaking it into smaller steps that were to be distributed over a longer period of time, or (2) developing a new method of demonstrating the concept (new instances of the concept, changes in the statements used in connection with the demonstration, changes in the type of responses required from the children, etc.)

Whether or not the change was relatively effective was determined on an "eye ball" basis. The variation in performance between the top performing disadvantaged children and the lower performing children is substantial. The top performing children usually achieve specific instructional objectives at least two times as fast as the lower performing children at the beginning of the year. By the end of the year, the difference decreases such that the top performing group is now learning at about one and a half times the rate of the bottom performers. We used this observation as the basis for evaluating the effectiveness of a given change. The problem is encountered first with the top performing children (since they are proceding through the program more rapidly and reach a particular trouble spot before the low performing groups reach that part of the program). A change is made in the program.

The assumption is that if the change is effective, the lower performing children will achieve the objective relatively faster than would be expected. If, for example, it took the top performing group 30 minutes to achieve a desired criterion of performance on a particular task (one that was deemed troublesome) one would expect (in the early part of the year) that the lower performing children would require at least 60 minutes to achieve the same objective, unless a change was made. If a change were made and the low performing children still required at least 60 minutes to achieve the objective, the change would be considered ineffective. If the low performing children required only 40 minutes, for example, to achieve the objective, the change was deemed effective. One could conclude (in a very rough manner) that the higher performing group, had it been exposed to the same change, would have achieved the objective in 20 minutes or less, a saving of perhaps 10 minutes, and more important, an indication that the objective was achieved with fewer errors and, consequently, perhaps greater enthusiasm.

The curriculum changes were not quantified because the procedure above was not always (and could not always be) followed rigidly. The procedure seems to have a great deal of potential for evaluating performance of children, but the problem is complicated by three factors:

1. Sometimes, the fault was not with the curriculum but rather with the method in which the concept under consideration was presented to the children. The remedy for a faulty or weak presentation was to change the presentation.

2. Sometimes, the remedy was obvious and could be offered very quickly (either by the other teachers or by the curriculum specialist). In some cases, the teacher wasn't presenting in the appropriate manner. In other cases, there was a minor fault in the program (perhaps a confusing example). The change was made immediately and the group with which the teacher was working at that time benefited from the change. In other words, the top group did not always go through the program "the wrong way" and the lower performing group "the right way." Sometimes changes were made so that the top group benefited from the change.

3. Sometimes, the factor of $2\frac{1}{2}$ early in the year and $1\frac{1}{2}$ later in the year does not apply to disadvantaged groups. It seems to apply only if the original program is reasonably well sequenced. If one were to try to use very poor teaching techniques to try to teach such concepts as colors or sequencing events to a high and low group; the difference between the groups would be far

more than the $2\frac{1}{2}$ factor. The difference might be more accurately expressed by a factor of 5.

Formal Criterion Testing:

The formal testing was based on the premise that intelligence is the sum of what a child had been taught during his lifetime. Translated into the instructional situation, the premise would hold that a program responsible for an increase in IQ or achievement must have achieved very specific objectives -- a number of them. The most accurate way to express what the program had done is to detail everything that had been taught. A test that samples the behavior of the children cannot specify precisely what had been taught. Furthermore, it cannot account for generalizations that were not taught or demonstrate how these generalizations were a function of what had been taught. The most accurate evaluation of a program is one that demonstrates everything that had been taught in the program, with an indication of the amount of time required to teach each concept. The number of things actually taught gives an indication of what the program did to change the behavior of the children. (Note that skills are not necessarily taught if the teacher merely goes through a teaching routine. They are taught only if the children can demonstrate after instruction that they can perform on the skills that were set out to be taught. Incidentally, it helps if one can demonstrate that they had not mastered the skills before entering the program.) The rate required for children to meet specific instructional objectives gives a more detailed, fine-grained profile of the program. The check list shows what the program has achieved for each child. The rate analysis indicates in greater detail where the program is relatively strong and where it is relatively weak. A skill that requires four times the amount of teaching may be indicative of specific weakness in the program.

The formal procedure used to evaluate the children's performance was to develop check lists of each item taught in the program. For example, one of the broad objectives of the program was to teach children the names and functions of parts of familiar objects. Real objects and pictures of objects were used to teach these concepts. The part-whole part of the curricula was organized and systematized into a set of 132 lessons. Examples of the instructional materials and accompanying illustrations are given in Appendix A. Every object and every part that was supposed to be taught in the program was included in a test. This test employed the language that the teacher used when teaching parts. "What is this?" (shovel) "A shovel has a ... (tester points to handle) and a shovel has a ... (tester points to scoop)." The test for parts was administered after the teacher indicated that she had completed the parts unit with her group. The results of the test were then fed back to the teacher. She was instructed to bring every child in the group to perfect performance. After the teacher indicated that she had worked with the children who failed particular items, the entire test was re-administered, not merely the items that the children had missed. The teacher was again told about the performance of her children and was again instructed to bring them to criterion. After the second test, only the items missed on the last test were re-administered.

The criterion tests (one on prepositions and one on parts) are explained and the check lists included in Appendix B. Because of the length of these tests, they were extremely time-consuming and sometimes difficult to schedule in the children's relatively short school day. The tests, however, provided a great deal of qualitatively valuable information.

1. The tests provided a basis for evaluating the performance of different teachers. As noted above, the investigators had an expectation of the relative time required for the different groups of children to reach the desired criterion of performance on any unit. When the performance of the children was not consistent with the expectation (when, for example, the second or third group beat the first group in reaching a particular criterion of performance), the discrepency was interpreted as a function of the relative effectiveness of the two teachers in question. The interpretation was that if one teacher's group consistently lagged behind that of another whom they should beat, this teacher was not teaching effectively. This conclusion could be tested by "switching" teachers from one group to another (letting the teacher of the top group work with the second group for a while) and noting whether or not the performance discrepency continued. If not, the initial discrepency was a function of poor teaching skills. The testing allowed the investigators to study "good teaching" in some detail.

2. The study allowed for inferences about the "hidden curriculum" in the middle-class home and the mysterious ways that children perform on sampling tests, such as the Stanford-Binet. It allowed the investigators to see what the children had been taught. The study of time and amount of teaching required to teach the set of skills under investigation allowed for at least some qualitative generalization about what would be required to teach a more elaborate set of skills, such as those sampled on a Binet IQ test. The assumption is that middle-class children have been taught a greater range of skills. The study of what it takes to teach these skills gives at least a rough idea of how much teaching goes on in the middle-class home -- how many hours and the variety of concepts. Since the format of instruction is quite different in the middle-class home than it is in the classroom (individual, informal versus group, formal instruction) there is no basis for extrapolation, merely a suggestion about the intensity of the education in the middle-class home.

3. The study provided a model for more effective appraisals of curricula. Unlike most achievement tests, the tests designed

for the study implied very specific instruction for the child that failed. Often the teacher who is given the results of an achievement test cannot see precisely which skills have not been taught to the children. The items are often complex and contain a number of possible causes of failure. Sometimes, the items included in the test are not specific items that have been taught in the course. Often the teacher does not view the test with the idea that if the children fail items on the test, they fail because they haven't been taught the skills necessary to master the item. (Teachers sometimes interpret test performance in terms of the mental adequacy of the children. "He's just a slow learner.")

The tests used for the present study involved simple items, presented according to the language conventions of the program and embodying the content words taught in the program. Each failed item, therefore, reflected a weakness in the program (in the way the concept presentation was designed, or in the way the presentation was executed by the teacher, or in both.) In any case, a teaching remedy was implied. The child had not been taught sufficiently well. The remedy was to teach him.

4. The study provided further inference about the relative effectiveness of the program. If certain items were failed rather consistently across the four groups of children, the conclusion was drawn that the items in question were not taught well. For example, if the children had trouble naming the parts of a sailboat, the investigators concluded that the method of instruction used to teach the parts of the sailboat left something to be desired. The qualitative nature of the children's responses often provided further information about how the teaching presentation was weak. The fault in the program may have been one of sequencing, which would mean that the objects taught immediately before the sailboat was taught may have been similar to the sailboat and may have induced some confusion. (This type of confusion would be reflected by the children's performance on the sailboat and on the objects that preceded it in the teaching sequence). The confusion may have been one of time. (Regardless of where the sailboat appeared in the teaching sequence, it perhaps required relatively more time to be taught than other four-part objects). The confusion may have been one of generalization. (For example, the children may have learned in connection with such objects as wagons, beavers, etc., that the main part of the object is called the body. The body of the sailboat is not called a body, however, but a hull.)

5. The study provided specific clues about how to change troublesome parts of the program. For example, if children consistently called the hull of a sailboat "body," a variation of the teaching demonstration for sailboat would be implied. The teacher would anticipate the mistake and alert the children to the mistake that they might make. "Look at this part of the sailboat. You know what some children call it? The body. That's silly, isn't it. We know what it is. It's the hull. Oh, you

have to be smart to know that. What part of the sailboat is this?" Other types of changes were prompted by (1) consistent failure of an item, and (2) the nature of the problems children had on the tests.

Qualitative analysis was not handled in a formal manner simply because the techniques used were basically exploratory. It was felt that adherence to a rigid procedure might force the investigators to ignore possibly useful qualitative data. As a result of the somewhat haphazard attempts to use the children's test performance results as the basis for drawing inferences about weaknesses in teacher presentation and in the structure of the program, the investigators were left with many interesting leads but little in the way of hard data about the relative effectiveness of the remedies that were implemented. The leads were for the most part qualitative since it was not practical under the circumstances to control for the different variables that might influence effectiveness of program revision.

RESULTS AND DISCUSSION

Three types of results emanated from the study. The first has to do with the changes in the program. The next is some kind of documentation that the changes were for the better, supported by improved achievement and IQ-test scores. The third type of results has to do with the procedures developed for "trouble shooting" and revising curricula. Because the results vary from objective test scores to impressionistic observations, reporting of results and discussion of them are combined.

Changes in the Program and Effects of Changes:

Changes were made in all three curricula areas -- arithmetic, reading, and language. Most of the changes were extremely specific; however, some of them proved to be rather extensive.

Arithmetic: The original conception of the arithmetic program was to relate the statements of arithmetic to counting operations and to give the children an operational understanding of each "symbol" used in the statements of arithmetic.

The problem with the original program was that it was dangerous for the slower performing children. The slower performer did not know how to count. He did not know the "names" of the numerals. And he did not know what the statements of arithmetic purported to tell him about the world. In brief, the entire game of arithmetic was new to him. He learned the word three in connection with counting exercises. At the same time, he learned three in association with a symbol. And he was exposed to a statement, "This number is three."

The three in the counting context, in the symbol identification context, and in the statement context often resulted in total confusion. He knew something about identifying objects, but his object called "three" was something that was quite different from the other objects to which he had been exposed. One didn't simply identify the object.

The slower-performing children had similar difficulties with the statements of arithmetic, such as, "One plus zero equals one." They did not adequately learn the meaning of this statement. Their great failure was to learn that one plus zero is the same as one. We discovered the faster performing children also had a great deal of difficulty with this concept, and that we had assumed that the concept is far easier for children to learn than it actually is.

The major changes that resulted from our tryouts of the arithmetic program were:

1. The symbol identification tasks were changed. Children were not required to give full statements in response to numerals. "What's this?...A three." The rate at which symbols were introduced was also slowed, so that the children would acquire a solid base in counting before they were exposed extensively to symbols.

2. The emphasis on production of arithmetic statements (one plus zero equals one) was reduced in favor of increased emphasis on demonstrations and practice with the operations (assembling one, adding no more, and noting the total). Emphasis on statement repetition and production was delayed until the children had had considerable experience with the operations.

3. Emphasis was placed on the meaning of "equal." The children were exposed to a series of graded exercises, beginning with the rule, "As many as you have on one side of the equal sign, you have to have on the other side." Children were initially given such exercises as, "What equals seven?...What's the same as eight? What equals eight?...If you have eight on one side of the equal sign, how many do you have to have on the other side?" The children progressed to more complicated statements, always with the emphasis on the equal sign. For example, the problem $4 + 2 = \square$ was solved by figuring out (by following the operation specified) how many are on the left of the equal sign. "So what equals six?... And how many do you have to have on the other side of the equal sign?...And what numeral goes in the box?..."

4. The entire arithmetic program was "slowed." A firm grounding in the meaning of "equal" was traded for possible acceleration of the faster performing children into such areas as fractions, rules of converting equations, etc. What this means is that the higher performing children were not able to progress

as far as they perhaps would have been able to in the original program. It was felt, however, that they would have a more solid foundation in their understanding of arithmetic statements and therefore would be able to progress more smoothly in the less structured school situation.

The effects of the changes in the arithmetic program are vaguely reflected in the performance of the children. The arithmetic achievement scores for the first and second classes (Groups I and II) of Bereiter-Engelmann children (after two years of instruction) is given in Table 1. The mean scores are nearly identical. Note, however, that the lower performing children in the 1967 graduating class scored substantially higher than the lower performers in the 1966 graduating class.

Reading: The reading program has undergone more change than any of the others. The original program was linguistically oriented. Children were taught that words had beginnings and endings. The word beginnings consisted of single sounds, such as, s, m, b, etc. The endings were ed, ad, am, etc. Different beginnings were programmed with different endings. The children were taught both the letter names and the sounds the letters made in the words that were formed by coupling a beginning and an ending.

The program proved to be only modestly successful. The mean reading achievement score of the children who were taught according to this program was .99, based on the Wide-Range Achievement Test, first edition. A number of the children failed to learn to read. (See Table 2).

In 1965, the ITA program was substituted for the original. It was obvious that the ITA had an advantaged; however, it did not succeed well with the low performing children, primarily because these children had a great deal of difficulty learning to combine sounds. They could learn to sound out a word, such as d-ooo-g, but they couldn't blend the sounds to form the word <u>dog</u>.

In 1966, a new program was introduced for the lower performing children. This program concentrated on the skills that are involved in sequencing sounds and blending them to form a word. Children were given verbal exercises in which they were given words verbally, such as, mmmmaaaannn. They were then told to "say it fast." The children were also taught to say words slowly. Words were presented at a normal speaking rate (man). The children were then to, "say it slow." Finally, a convention for blending was introduced. The first letters the children were taught were continuous sound letters--mm, a, u, s, etc. These were identified by the sound they make, not by the name. When the children began to read simple words formed by these letters, they were taught to hold each sound until they produced the next. Instead of sounding

Table 1

Arithmetic Achievement of Groups I, II, and III
After Two Years of Instruction
(Wide Range Achievement Test*)

Subject	Grade Level
Group I	
1	2.5
2	1.8
3	1.4
4	2.0
5	3.6
6	2.9
7	3.1
8	1.8
9	2.3
10	3.3
11	3.6
12	3.1
13	2.5
Group I Mean	2.61
Group II	
1	2.2
2	2.3
3	3.3
4	3.1
5	2.9

6	2.5
7	3.3
8	1.4
9	2.2
10	2.7
11	2.2
12	2.0
Group II Mean	2.51
Group III	
1	2.4
2	1.6
3	.9
4	.4
5	2.2
6	1.9
7	.9
8	2.2
9	1.6
10	2.1
11	2.2
12	1.4
Group III Mean	1.65

*Changes occurring in the Wide Range Achievement Test prior to its use to test Group III make these scores not strictly comparable to those of Groups I and II.

Table 2

Reading Achievement of Groups I, II, and III
After One and Two Years of Instruction
(Wide Range Achievement Test*)

Subject	Grade Level End of First Year	Grade Level End of Second Year
Group I		
1	1.0	1.3
2	1.0	.9
3	1.0	1.0
4	1.1	1.1
5	1.4	1.7
6	.9	1.3
7	1.3	2.3
8	.5	1.3
9	1.1	1.3
10	1.1	1.9
11	1.2	2.1
12	1.1	1.8
13	1.0	1.3
Group I Mean	.99	1.48
Group II		
1	1.1	2.7
2	1.1	1.6
3	1.2	3.1
4	1.3	3.7
5	1.2	2.7

6	1.2	3.6
7	1.2	3.1
8	1.2	1.6
9	1.3	2.0
10	1.2	3.1
11	1.2	1.7
12	1.2	2.3
Group II Mean	1.20	2.60
Group III		
1	1.4	1.4
2	1.5	1.6
3	1.6	1.6
4	.7	.5
5	1.8	2.3
6	1.3	2.3
7	1.2	.8
8	1.3	2.8
9	1.2	1.3
10	1.2	1.6
11	1.2	1.3
12	1.4	1.5
Group III Mean	1.32	1.58

*Changes occurring in the Wide Range Achievement Test prior to its use to test Group III make these scores not strictly comparable to those of Groups I and II.

out the word man as mm--aa--nn, they were taught to sound it out, mmmaaannn. They were then told to "say it fast."

During the school year, 1966-67, the new reading program was introduced to the slower-performing children only. During the school year of 1967-68, it became the standard program for all of the children. The program enabled the investigators to teach reading to the slower-performing children for the first time.

The lowest reading achievement score in the 1967 graduating class (Group II) was grade level 1.6, compared with grade level .9 for the 1966 graduates (Group I). The difference in reading instruction seems to be reflected both in the mean reading scores and in the range of scores, with the 1967 class having the advantage in both cases.

Because a new edition of the Wide-Range Achievement Test was used to evaluate performance of the 1968 class (Group III), the scores are not comparable. However, the achievement scores for this class are also included in Table 2. Although the 1968 scores are lower than those of 1967, the absolute performances of the children in these groups are quite similar (and considerably better than that of the 1966 class).

Language: The goal of the original language program was to teach children the language of instruction, that is, the concepts and constructions that would be used in a range of future teaching demonstrations. The original program called for a tight sequence of activities, beginning with work on the "identity" statement, and then a variety of "second-order statements." The identity statement is used to identify objects or actions, "This is ______ ." The second-order statement is designed to tell something about the object (or group) that has been identified, "This _____ is _____."

The concepts taught using the second-order statement were polars (opposites), class concepts, colors, and prepositions. After the children had mastered the first- and second-order concepts, they were introduced to if-then constructions, statements involving and or, etc. Four primary changes have occurred in the language program.

1. The range of skills has been expanded.

2. Better demonstrations for teaching language concepts have been introduced.

3. As in the arithmetic, the emphasis moved from the repetition and production of statements to the operations that underlie statements.

4. The sequence of skills was altered to give the lower-performing children a more errorless introduction to the second-order concepts.

The first class of children was introduced to color relatively early in instruction. The introduction of color was delayed for subsequent classes. The rationale behind the delay was that the slower-performing disadvantaged child does not have a clear idea that objects can be instances of various concepts. The slower performer may have the notion that an object is an instance of a single concept (conveyed by the name of the object), but he may not understand that after an object has been labelled, a number of attributes (such as color, shape, position, etc.) can be noted. The first introduction to these other attributes should be as obvious as possible. The notion of color is not particularly obvious, because of demonstration limitations associated with the concept. It is not possible to "remove" the color from an object and show the child a not-instance of a particular color. The introduction of color concepts was therefore delayed until the child had learned a number of other "attributes" which were more easily demonstrated than color. Parts, polars, and prepositions were used to provide the introduction to attributes. Only after these had been mastered were color concepts presented.

The demonstrations associated with the presentation of concepts were improved on a trial-and-error basis. If a particular presentation proved to be difficult for the slower-performing children (as noted by the length of time required for them to master the concept and by the number of persistent errors they made), the format of the presentation was changed. In many cases, the original presentation was broken into a series of tasks that would lead to terminal performance in a more errorless fashion.

The effectiveness of a language program is reflected in the IQ achievement of children. If the range of language skills that is taught includes a wide variety of applications and a relatively extensive vocabulary, the child will have been taught more language concepts than children his age are normally taught. This increased language-skill repetoire should be reflected on an IQ test.

The IQ performance of the 1966, 67, and 68 classes (Groups I, II, and III), is summarized in Table 3. The improvement in the language program seems to be reflected in the increased IQ scores. Note, however, that these scores are relatively independent of the achievement scores in other areas. Since the reading program was improved at the same time the language program was improved, there is some ambiguity about the relationship between reading performance and IQ level; however, the arithmetic performance and the spelling performance of the children in the three Bereiter-Engelmann classes seems to be uncorrelated with IQ performance. The arithmetic achievement scores have not changed substantially over the three

Table 3

Stanford - Binet IQ Scores
for Groups I, II, and III

Subjects	At Time of Entry	After First Year	After Second Year
Group I			
1	121	101	126
2	94	96	100
3	95	109	117
4	98	101	100
5	92	100	107
6	105	110	108
7	96	100	99
8	95	103	100
9	91	110	103
10	105	116	114
11	99	114	114
12	88	98	102
13	90	96	99
Group I Mean	97.21	104.07	106.93
Group II			
1	92	113	123
2	93	94	103
3	105	112	121
4	89	101	131
5	99	116	119

6	86	105	112
7	119	130	139
8	90	107	112
9	84	100	108
10	109	125	138
11	99	114	129
12	101	123	118
Group II Mean	97.17	111.67	121.08
Group III			
1	101	123	94
2	83	96	94
3	96	100	108
4	84	94	102
5	78	103	112
6	93	96	118
7	81	102	88
8	111	121	121
9	90	97	105
10	111	113	117
11	85	95	107
12	95	109	122
Group III Mean	92.33	104.08	107.33

years under consideration. Yet, as Table 3 indicates, the IQ scores have changed dramatically.

Perhaps the most significant aspect of the IQ performance of the three classes is the increased number of children who achieved substantially above 100 in IQ performance. The investigators interpret this tendency as an indication that the language program allowed the faster-performing children to learn language concepts at a substantially above-normal rate. The same rate was not possible with the slower-performing children, since a great deal of time had to be devoted to the programming of basic concepts.

Observations About "Trouble Shooting" for Program Improvement

The observations divide into two categories: observations about the curricula-revision process, and observations about the dynamics of effective teaching techniques.

Observations about the curricula-revision process: The following observations were somewhat facilitated by the procedure of testing each child individually on the skills that were supposed to be taught in the program and by the procedure of then revising the program where it appeared that error rate was high or time required to reach criterion was generally excessive. Since adequate controls were not present during the investigation, these observations will be discussed qualitatively.

1. Different formats for teaching a particular concept may result in substantially different numbers of errors produced by the children and the time required for the children to reach criterion performance. This observation may seem patently obvious, but it is quite important. The difference between a carefully constructed program designed to teach a particular skill and one that is not as carefully sequenced is reflected both in the amount of time it takes for children to learn a particular skill and the degree to which the skill is retained over a two-month period. The better program not only teaches the skill more rapidly. It also results in better retention.

In several cases, skills were re-programmed in such a way that the lower performing children (group 4) achieved criterion in almost as little time as higher performing children (group 1).

Programs that have the potential for reduced error production are particularly important when the first instance of a concept is presented. All part-whole tasks are the same in that all have the same conceptual structure. In all, the child must learn that the object is the sum of the parts and that if all of the parts are in place, the whole object is complete. If the whole object is not complete, some of the parts are missing. The name of the object is not the same as the name of a part of the object. If one is asked, "What is this?" and is presented with an object, the answer is the name of the object. If one is asked to "Tell me the parts a _______ has," a different type of response is required. Telling the name of the object is not called for. The respondent must now examine the object not as a whole, but as an aggregate of parts.

The structure above applies to all part-whole relationships. A child must learn this general structure. However, he must learn it in order to handle the <u>first part-whole task he encounters</u>. The child then applies what he learns to all other part-whole tasks. He does not have to re-learn the structure of the concept of part-whole relationships. Conversely, there is no way that he can consistently learn to handle the tasks that are presented in connection with the first part-whole object without learning the structure. What this means, translated into performance expectations, is that the child who has not learned part-whole relationships in connection with an object will make a greater number of mistakes on the first object presented than he will on other objects. If the objects that are presented after the first are similar in degree of difficulty (number of parts, length of name of part, familiarity with whole object), we would expect a typical learning curve in which the child made progressively fewer mistakes on each new object until he reached a "plateau" in speed and number of errors. His performance would then remain relatively constant.

From the standpoint of curriculum development, it is important to flatten the curve. The reason is that when a child makes a high number of mistakes on the first "instance" of a concept, retention will be relatively poor for that instance. Not only will the subject require many more trials to reach criterion, but (1) the subject will not retain the skill as well as a subject who has the same number of trials in which he made fewer errors, (2) the subject will not retain the skill as well as a subject who made fewer errors but who made the same number of correct responses, (3) the subject will not retain the skill as well as subject who made fewer errors and fewer correct responses.

In brief, one of the major differences in material that is well programmed and material that is not so well programmed shows up most dramatically in the rate at which children master the first instances of a new skill. The more poorly sequenced program does not have the potential for acceleration because it takes longer to bring children to an initial criterion. If the program introduces new instances before the children master the first instance (or first set of instances), then the performance of the children may be severely retarded (compared to the program in which the skills are sequenced to achieve fewer initial errors). A good teaching demonstration is needed for the first instance if the performance of the children is to be accelerated. The advantage of the well programmed sequence is not as evident after the children in both groups have mastered the initial set of instances. The well sequenced program, however, will achieve both faster learning of the entire set (primarily because of the saving on the first several instances) and it will account for better retention of the entire set (primarily because of the better retention on the first instances taught).

The difference in performance on well programmed sequences and relatively poorly programmed sequences is most noticeable among the lower performing children. The learning curve for the higher performing children on most of the tasks presented in the language program is flatter, probably because these children have already been taught some of the components of the criterion tasks. The lower performing children must be taught more, and the importance of a good demonstration is therefore increasingly important.

2. **The minimum time in which new** skills can be taught to four-year-olds (given that the skill is not worked on for more than five minutes a day) is three days for the higher performing children and about twice that time for the lower performing children. This observation is based on the performance of children over a broad range of skills.[1] For some skill acquisition, the

---

[1]It is interesting to note how far some existing instructional programs fall short of this minimum time per unit of learning. Some reading programs, for example, presume to teach the children the alphabet in the matter of a few minutes. Science programs expect children to learn a principle from one experimental demonstration of that principle. Language programs and English programs assume that children learn about the parts of speech and about contractions, for example, from one lecture. If the children are to retain what they have been taught, they will probably require at least three days of five-minute-a-day instruction on the skill in question. (It may be that older children who are more familiar with the programming game will not require this much time; however, younger children, even very bright children, do.) The greatest mistake a program can make is to present a compelling demonstration, provide for the children to make responses, and then expect the children to have learned the skill well enough to be able to retain what they have learned or to use it in other situations. The children will most certainly fail. The simplest two-part object (such as a shovel) requires instruction over three days if the knowledge is to be retained. Surprisingly enough, the three-day minimum holds reasonably well across various tasks (prepositions, parts, letter sounds, sequencing skills, statement repetition skills, etc.)

difference between the high and low performing children is even greater. On symbol recognition, it is not uncommon for naive, high performing children to learn new symbols at the rate of one every two days, and for lower performing children to require about seven days for a symbol, after the first two symbols have been introduced. On other tasks, the differential is not so great.

3. According to the concept analysis procedures used in the program (Engelmann, Concept and Problem Solving, 1969), a teaching presentation must be consistent with one and only interpretation. If it is possible to derive two interpretations from the presentations of examples and the language the teacher uses during the presentation, some children will learn the incorrect interpretation. Although no formal attempt was made to design tasks that were consistent with more than one interpretation by virtue of either language ambiguity or ambiguity of the examples selected, tasks that presented problems were scrutinized. In every case, a reasonable source of confusion was found to be implied by the presentation. Perhaps the words employed were similar to words used to describe a superficially similar concept. Perhaps the examples were poorly selected. Perhaps the teacher did not rule out all of the irrelevant cues (always presenting the letter c after the letter b, for example), etc.

The intensive study of rough spots in the program gave the investigators greater conviction that if tasks and teaching demonstrations could be constructed analytically sound, they would work. The children would learn relatively quickly and errorlessly.

4. The investigation confirmed that a successful program is one that (a) quickly demonstrates what the concept is and what it is not, (b) uses a carefully selected set of examples during the initial demonstration and then quickly expands the set, introducing a wide variety of examples, to firm the generalization, (c) uses "formats" or a similar form for tasks that are similar, so that the teacher's language is controlled and so that the children learn what general types of responses are required of them not only for the present tasks but for all tasks that take the same "format", (d) uses simplified language and places emphasis on the demonstration through examples rather than on verbal explanations from the teacher that are designed to appeal to the children's logic, and (e) requires a great number of responses from the children. Responses are not a substitute for adequate teaching demonstrations. Responses, however, help the children get used to the new language introduced and help them become comfortable with the new concept. Good programs are also designed so that they do not "spiral," that is, they do not "sort of" teach a concept the first time the concept is introduced and then, at prescribed intervals, "sort of" teach a little more about the concept. Such programs are designed to create confusion, especially among the lower-performing children. The evaluation of the present curricula would indicate that a good program teaches the concept as it will be used during the

initial teaching demonstrations (using enough examples of the concept to assure generalization). The concept is then incorporated in more complex tasks. Unless the concepts are used, normal forgetting will occur. Such forgetting may be prevented by continually review of the concepts (which is sometimes necessary if the skills are not neatly structured). It is far more efficient, howeve:, to teach the concept and then require the children to perform on tasks in which they must use the concept.

Observations about the Role of Teacher Behavior in Inducing Most Rapid Learning

The teachers in our sample differed considerably in their efficiency in teaching children, even when adjustments were made for the expected rate of performance of the different groups. Switching teachers from group to group verified the differences. These differences were only slightly noticeable on "short" tasks, that is, tasks that involved no more than a few examples presented by the teacher and a simple statement response from the children, a yes-no response, or a pointing response. On more complicated tasks (tasks requiring a more elaborate demonstration or a series of responses), the difference was more apparent. The investigators observed that this "difference" is actually the produce of very specific behaviors. These behaviors could be taught to teachers who did not initially have them, resulting in improved performance of the children. These behaviors are listed below. Note that these are not all of the behaviors that make for good teaching. The investigators felt that all of the teachers in the investigation would be rated at least "good" by outside observers. These are the behaviors that make a difference on tasks that typically present problems to some or most of the children. The investigation focu'ed on these tasks and attention was paid to the behaviors that lea' one teacher into trouble and the behaviors of another teacher tha' butt-tress against this trouble.

1. The "good teacher" was more pessimistic abo' the preparation of her children for a test. She was also more acc'.ate in predicting how well they would do on the test. The good te.cher apparently tended to overteach. She hesitated to move o. to another task until all of the children in her group were perf.rming adequately. The teacher who was not as good did not get as much .eedback from the children. She did not seem to have the burning desire to teach every child. She let the children get by with perf.rmances that would not be acceptable to the good teacher. In o.e sense, the good teachers reminded one of Helen Keller's teacher, as she was portrayed in The Miracle Worker. They felt that the children could perform and should perform if the teacher knew how to reach them. The teacher who was not so good seemed to have a more mechanical view of the teaching process. It did not seem to bother her if the children did not perform well. This is not to say that the good teachers showed their feelings and concern to the children. Quite the contrary. However, discussions with them revealed their frustrations and desires.

All teachers failed to judge completely accurately when they had "taught" a concept to all of the children, as judged by the performance of the children. On almost every test, one of the children who was supposed to be taught a set of items failed one or more items. The good teachers had fewer failures, however.

2. All teachers, good and not quite so good, tended to create a set for responses on particular kinds of tasks, especially among the lower-performing children. The tester could present the item and a child might fail it. The teacher could present the same item, giving no prompts, and the child might respond appropriately. It was difficult to determine whether good teachers created a stronger "set" than the other teachers. The writer tends to think that they do. They are such strong sources of information and direction that the children rely on them. If this is the case, a good teacher would have to change the instructional set after the children had mastered a concept, perhaps presenting in a low-key manner, giving the children worksheets, etc.

3. The good teacher tended to require 100% performance from the children. When a good teacher pointed to a picture and said, "What's this?" she expected all children to respond. If they didn't respond, she would perhaps smile and say, "I didn't hear you. What's this?" By now all of the children were responding. She would smile, cock her head, and say, "I didn't hear you." Now the children let out with a veritable roar. The teacher would acknowledge, "Now I heard you," and procede with the next task. It was quite noticeable that the children performed well on the next task, with virtually 100% of them responding. Basically her approach was to stop and introduce some kind of gimmick if the children -- all of them -- were not responding or paying attention. She did not bludgeon the children. She "conned" them. It seemed obvious that they understood her rules. She would not go on until they performed. It seemed that they liked performing, because when they performed well she acted pleased.

A similar pattern was noted in the manner in which she would reinforce. She would tend to present a series of quick tasks together, without interrupting herself at the end of each task, except perhaps to say, "Here's another one." Occasionally, she would urge the children, "Come on, this is tough." Throughout the series of tasks, she would not reinforce the children. Then, at the end of the series, she would stop, smile and perhaps say something like, "Wow, you guys are really smart. How did you do that?"

The pattern of presentation for the not-so-good teacher was quite different. She tended not to require 100% performance from the children. As a result, it seemed that she often reinforced inattentive behavior. If all of the children were not attending, she would proceed with the task. The children in the group might find it more difficult to work on the task than to look around the room. The looking around the room is relatively reinforcing in this situation, and the teacher is reinforcing such behavior. It was noted that the off-task behavior problems of this teacher tended to escalate (at least during the first two months of instruction). Generally, the children tend to settle down after this period, but they don't respond as well in the instructional setting as the children who work with the good teacher.

The relatively not-so-good teacher tended to interrupt her performance. She tended to reinforce after every task. (Such reinforcement was often indiscriminate, including even those children who were not attending or responding as well as the good teacher would have required). The fact that the not-so-good teacher tended to interrupt herself seemed to handicap her when she wanted to produce a quick series of responses from the children. She had trouble managing this kind of task, because there was normally a substantial time lag between the presentation of tasks.

4. The good teacher seemed to use very crisp signals, and she tended to treat her signals as if they had very precise meanings. Perhaps the best way to illustrate the difference is to compare the teaching of two teachers on the task, "say-it-fast," in which a word is presented to the children in slow motion, and the teacher then says, "say it fast."

The not-so-good teacher presented her initial tasks this way. "Listen...Motor (pause) cycle, say it fast." Some of the children responded, "motor (pause) cycle." The teacher said to the group, "Good. Motor-cycle." Several months later, the children in this group were having trouble with "say-it-fast" tasks.

The good teacher presented her initial tasks this way. "Listen... Motor (pause) cycle (pause, pause)..." During the pause some children started to say, "motorcycle." The teacher held out her hand and returned to the beginning of the task." Motor (pause) cycle (pause, pause)..." Again, the children interrupted her. Quietly she said, "You've got to wait." Again, she returned to the beginning of the task. "This is tough. See who can wait. Motor (pause) cycle (pause, pause) <u>say it fast</u>." Most of the children responded by saying the word fast. Others, however, said, "motor (pause) cycle." The teacher smiled at them and said, "say it <u>fast</u>." Again, the children said, "Motor (pause) cycle." And again, the teacher said, "say it <u>fast</u>." Before the children could respond again, the teacher said, "I can do it. Watch me. Motor (pause) cycle (pause) say it fast...Motorcycle." With a smile, she continued. "I did it. I said it fast. Your turn. Listen. "Motor (pause - cycle (pause, pause, pause) say it fast." The children responded, "motorcycle." The teacher almost yelled, "That's saying it fast. Here's another word. Listen..." Within two weeks, this teacher had all of the children in her group performing consistently on "say-it-fast" tasks, including very difficult tasks.

The example above illustrates the teacher's passion for perfect performance from the children. It also illustrates her crisp use of signals, and her treatment of signals as just that -- signals with a very precise meaning. Note that the teacher set the task up so that there was a pause before she presented the signal, "say it fast." This meant that the children had to wait for the signal. They could not perform in the task by chance or by memorizing the cadence. The pause before the presentation of the signal varied in length. The children were therefore "hooked" on the signal. The teacher made sure that they were hooked. She also made sure that the children understood the meaning of the signal. After the children made a mistake, she repeated the signal, in much the same way one would repeat the signal, "Get out of the street," or "pull the cord." Before providing the correction, she demonstrated to the children that they were to produce a very specific response to this signal. She did not "shape" the children. She did not accent "sort of" responses. She accepted only true "say-it-fast" responses. When the children produced these, she reinforced them vigorously.

5. The final major difference between the good teacher and the not-so-good teacher was that the good teacher paced her presentation so that she presented examples quickly. When the children were required to produce complex responses, she did not rush them. However, when she was introducing a new concept, she structured the task so that the children could initially respond using only yes-no responses. She then presented examples as quickly as possible. For example, her introduction of the concept "over" took only a few minutes, but within the few minutes, she presented over fifty examples.

She started out with an eraser. "What's this?...Watch the eraser. Is the eraser over?..." She moved the eraser over the table. "Yes. Is it over?" She moved the eraser higher over the table. "Yes. Is it over?" She moved it over another part of the table. "Is it over?" She moved it on the table. "No. Is it over? No." Keeping it on the table, she slid it to another part of the table. "Is it over?..." Back to the middle of the table. "Is it over?..." Above the table. "Is it over?..." Higher. "Is it over?" On the table. "Is it over?..." And so forth.

Note that she answered the first few questions. After these, she never led the children, never prompted them -- simply asked the questions and waited for their response. They did not make a mistake throughout her presentation, either on the group work or when she called on them individually.

The not-so-good teacher tended not to pace the instances as well. During the time that the good teacher presented over fifty examples, the not-as-good teacher presented only four. She required five statements from the children. She kept the examples more static. The good teacher almost provided children with a "picture" of what "over" meant. The not-so-good teacher did not paint so good a picture, primarily because she did not pace her presentation so that the children were exposed to as many examples of the concept.

SUMMARY

The present investigation dealt with "trouble-shooting" techniques for improving the Bereiter-Engelmann curricula. The investigation was mostly informal. The procedure was to identify curricula and teaching problems as they arose and to work out possible remedies.

The results of the investigation were qualitative in terms of the techniques for effective program construction that derived from the investigation. That these changes were at least partially successful was reflected in the improved performance of the children on achievement and IQ tests. During part of the grant period, the language curriculum was studied in great detail. This study was conducted in connection with the project dealing with the achievement component of the Stanford-Binet (See Table 3). During this part of the investigation, a more detailed check was kept on the teachers' behaviors and on the performance of the children, as measured by check lists of all of the skills taught in the program. Since the teachers were responsible for bringing every child to criterion performance on as many skills as possible, the study afforded the investigators an opportunity to note the effects of curricula changes and teacher-behavior variables on the performance of the children. Although these variables were not quantified, they were observed. The investigators felt that many interesting leads derived from qualitative studies.

From the studies, the investigators distilled procedures for effective program revisions and noted teacher behavior that tends to lead to better achievement performance of the children. The investigators feel that more carefully controlled experiments are needed to give more accurate data on the influence of the various variables noted in the present investigation. They feel, further, that such experimentation would be more fruitful, since it would carry implications both for effective skill programming and for teacher training.

The Dual Kindergarten

Ernest D. Washington
and Jean Osborn

INTRODUCTION

The Dual Kindergarten was a pilot study for a public school program. Children who exhibited special educational needs would spend one-half day in a traditional kindergarten and one-half day in a special class which would be directed toward giving them extra preparation needed for successful achievement in the first grade.

The pilot study was undertaken with the cooperation of Unit 4 School District, Champaign, Illinois which at the time was studying an integration plan for its community. The integration plan included busing children from economically poor neighborhoods into schools throughout the community. The dual kindergarten classes, which would give selected children a full day of kindergarten, had two objectives as a public school program.

1. The integration of socio-economic and ethnic groups in the public schools.

2. Equalization of the readiness of children in all groups for participation in the educational process which would begin in the first grade.

The Unit 4 School District was particularly interested in the Dual Kindergarten program as a means of reaching the six to twelve percent of the district's children (depending upon the subject matter area and the grade level) who were more than one grade level below their peers on standardized achievement tests. Even with a comprehensive elementary education program which met the needs of most of the children in the community, 80 to 100 children entered the Champaign Schools each year who began achieving below grade level and continued that pattern throughout their school experiences (Champaign Community Unit School District No. 4, 1968).

The two-class all-day kindergarten program manifested several advantages over a single-class all-day program: by spending one-half-day in a community kindergarten the children selected for the dual-kindergarten program would establish social relations with the diverse kinds of children who would be their classmates in years to come. In these classes they would take part in the learning experiences and readiness activities commonly associated with kindergarten. In the other part of the day, the children would be assigned to special classes according to their needs. Which children would be eligible for the classes, and what type of classes they would attend would be determined by diagnostic testing and teacher evaluation of the performance of the children in their regular kindergarten classes. It was felt that most children who would be eligible for the program would need classes which emphasized language, concept development, and pre-academic and pre-reading skills, but that other classes that focused on social adjustment and behavior modification or on the remediation of some special handicap would also be established as a part of the program.

The second half-day program would be made available for all children in the community, not just the children from the neighborhoods that were being bused into new schools. The classes would be held in a regular school building. Children who did not demonstrate the need for special classes would continue to attend kindergarten for half-day sessions only.

OBJECTIVES

The University of Illinois pilot study was for a class in the development of language, reading and arithmetic skills.

The specific goals of the pilot study were:

1. To investigate the feasibility of combining a two-year compensatory education program into a one-year program.

2. To compare the gains made in academic and intellectual growth of two groups of children, one group attending a half-day public school kindergarten session, the other attending a half-day public school kindergarten session plus a special class at the University of Illinois.

Work that had been done the previous three years by Bereiter and Engelmann (1966) at the University of Illinois indicated that a highly organized and structured direct teaching program in which groups of children entered at age four and continued in for two years resulted in large gains in intellectual capacity as measured by the Stanford-Binet and in first and second grade scores in reading and arithmetic on standardized achievement tests. The classes, which were held daily for two and a half hours, consisted of carefully programmed instructional units in reading, arithmetic and language which were carried on in small groups, and in larger group activities which involved related and reinforcing activities in art, music, games and seatwork projects.

The feasibility of compressing this two-year program into one year was to be examined as a part of the pilot study.

To measure the effect on educationally disadvantaged children of a special class which emphasized academic skills, twenty children were selected from summer Head Start classes. All twenty attended kindergarten in the public school every morning, ten of these children were placed in the special class at the University which met every afternoon.

The following were the objectives for each of the three subject areas:

A. Language: The language course was designed to teach each child, directly and systematically, the language of the public school. The following description of the language course comprises the minimum objectives for all of the children in the language program. At the end of the year, each child was to be able to handle both the concepts and the language of the following:

1. Statement of what an object is:

a. First order statements

Identification statements.	This is a cup.
Not statements.	This is not a cup.
Plural statements.	These are cups.

b. Second order statements

Classification statements.	This dish is a cup.
Shape statements.	This cup is round.
Color statements.	This cup is red.
Pattern statements.	This cup is striped.
Polar statements.	This cup is big.
	This cup is little.
	This cup is made of plastic.

c. Instructional statements

Prepositions.	This cup is on the table.
Conjunctions.	The cup and the spoon are on the table.
Other function words.	All the cups are white.
only, one	Some of the cups are pink.

Same-different.	These cups are the same because they are all made of plastic. These cups are different because some are white and some are pink.

2. Statements of what an object does:

Use of <u>does</u> and <u>can</u>.	Does a cup hold milk? Can a cup hold milk?
Use of verb tenses.	Is there milk in the cup? Was there milk in the cup?
Use of pronouns.	Is he drinking from the cup? Is it in his hand?
Situations of function and use.	Do we drink from a cup? Do we wear a cup?

3. Statements of what an object has:

Part-whole relationships.	A cup has a handle and body.

The preceding is the format by which a larger number of objects, instructional words, and related descriptive concepts are taught. It was anticipated that by the end of the dual-kindergarten period, the children would enter first grade with a language competence that would enable them to learn in a middle-class public school setting.

B. <u>Reading</u>: The objective for the reading program can be expressed rather simply -- to teach children to read (decode) words.

However, this objective implies a number of sub-objectives that had to be met.

1. To identify letter symbols (identifying them as <u>sounds</u>) using the short vowels, long vowels, and the consonant sounds.

2. To sequence action events in time.

3. To read a symbolic representation of action events, ordered on an arrow that proceeds from right to left.

4. To identify words that are presented "slowly." (Tell me what it is and you can see the picture: motor----cycle. What is it?)

5. To say words that are presented at a normal speaking rate slowly. "Listen: me. Say it slow..."

6. To slide continuous sounds, without pausing between sounds (The child reads the symbols f a s m a s fffffaaaaassssssmmmmm, with no pause).

7. To read words that do not involve stop sounds. (The child reads the word <u>fan</u> first by reading the symbols without pausing---fffaaannn--and then saying it fast, identifying what he has said as a word. What word is it? Fan.)

8. To handle words that begin with a stop sound (can).

9. To read irregularly spelled words. (These are introduced only after the child has cracked the code and can handle regularly spelled words.)

10. To read groups of words.

11. To identify upper- and lower-case letters by their conventional names.

12. To recite the alphabet.

13. To write the letters presented in the reading program.

C. <u>Arithmetic</u>: There was assumed to be a range of individual differences within any group assigned to the Dual Kindergarten. Some children will proceed more rapidly than others. Some will know more initially than others. Therefore, two sets of objectives are implied -- one for the slower children and one for those who proceed more rapidly.

The objectives for the slower children:

1. To identify the numerals 1-10.

2. To be able to handle verbal problems of the following form: If you had four raisins and I gave you one more raisin, how many raisins would you have?

3. To solve written problems of the form:

   3 + 2 =

   3 - 2 =

4. To handle problems of changing such statements as

   4 + 6 =

   so they are true. This objective is extremely important for disadvantaged children. Often they fail to learn what the equal sign means (namely that what is on one side is the same numerically as what is on the other).

5. To work problems of the form:

   5 + 3 =

   using finger operations.

The objectives for the faster children included all of the objectives for the slower children and in addition:

1. To identify the numerals 11-100.

2. To learn basic multiplication operations (using a multiplication chart).

3. To group a fixed number of objects into two piles and state the addition facts that derive from the grouping.

4. To handle 0 and 1 problems that are presented verbally.

5. To handle problems of the form:

   3 + a = 5

   3 - a = 1

METHOD

Subjects

Table 1 shows some of the characteristics of the 20 children who participated in the study.

Treatment

The children in this study were randomly assigned to the experimental dual kindergarten program or the control kindergarten class.

Table 1

Group Composition

Program	N	Binet CA	Mean Entering Binet I.Q.	Intelligence Strata Mean High	N	Med	N	Low	N	Race - Sex Black M	F	White M	F
Experimental Dual Kindergarten	10	61.8	86.2		0	93.3	3	83.3	7	4	4	0	2
Control Dual Kindergarten	9	61.2	89.4	100	1	94.7	3	84.3	5	2	4	3	0

The experimental dual kindergarten group attended a regular public school kindergarten class in the morning and an academically oriented kindergarten class in the afternoon at the University of Illinois. The control kindergarten class attended the same kindergarten classes in the morning as the experimental dual kindergarten group, but did not attend any kindergarten program in the afternoon.

The ten children selected were placed in a class of fifteen five-year-old children who had attended the experimental class as four-year-olds. These fifteen second-year children attended only the afternoon class. Being placed in a class of experimental children had several advantages: the experienced children helped the new children become acquainted with the routine of the school; the experienced children were familiar with the subject matter; and the combined classes enlarged the grouping possibilities. All of the children were grouped for subject matter areas. After the first few weeks, it was possible to combine some of the slower children from the first- and second-year classes into one group. The twenty-five children were taught by three experienced subject-matter teachers, plus one teacher who handled activities other than subject matter areas. Twelve student teachers also participated in the program under the direction of the experienced teachers. The children attended the class from September to early June and followed a normal public school calendar.

Testing Procedure

The battery of tests used consisted of the Stanford-Binet, Illinois Test of Psycholinguistic Ability (ITPA) and the California Achievement Test Lower Primary. The Stanford-Binet and ITPA were used as pretest and posttest measures and the California Achievement Tests were added to the battery as posttest measures. Certified school psychologists administered the Stanford-Binet and ITPA while a specially trained teacher administered the achievement test.

Result

The Stanford-Binet, ITPA and the California Achievement Test were used as posttest measures to evaluate the progress of the groups. The groups differed somewhat on initial scores and the initial Stanford-Binet scores were used as an initial coveriate to statistically adjust for the initial differences between the groups.

A comparison of the experimental and control dual kindergarten groups on the California Achievement test is shown on Table 2. The P values from the table indicate that there are substantial differences in achievement between the two groups. Because the titular descriptions of the various achievement areas are not very

Table 2

Unadjusted and Adjusted Mean Grade Placement Scores on California Achievement Test

	Experimental Dual Kindergarten		Control Dual Kindergarten		Adjusted Mean Difference	P*
Reading Vocabulary	1.19	1.20	.88	.87	.33	.01
Reading Comprehension	1.21	1.24	.45	.42	.82	.01
Arithmetic Reasoning	1.24	1.24	.93	.92	.32	.005
Arithmetic Fundamentals	.99	1.00	.15	.14	.86	.005
Spelling	.75	.79	.00	-.04	.79	.01

P* based on analysis of covariance of posttest scores using initial Stanford-Binet IQ as the single covariate (d.f. = 1.16).

informative, a brief description of the various subtests seems in order, along with the report of the accomplishments of the two groups.

The adjusted mean reading vocabulary scores for the experimental and control dual kindergarten groups were 1.20 and .88. This data would seem to indicate that the groups differed by approximately three months in reading vocabulary. This is somewhat misleading; the reading vocabulary consists of four separate subtests and the groups differed markedly on several of the subtests.

Reading vocabulary consists of the following subtests: word form, word recognition, meaning of opposites and picture association. Word form consists of twenty-five items in which two words are separated by a dotted line and the child is required to write "S" or "D". Word recognition is made of twenty sets of three words and the child's task is to underline the word pronounced by the examiner. Meaning of opposites contains fifteen boxes in which are a response word and two distractors. The child matches the key word with its opposite. The picture association subtest contains fifteen items and requires the child to identify objects or interpret pictures.

The raw scores on Table 3 show that the groups differed hardly at all on the word form subtests but differed consistantly in word recognition, meaning of opposites and picture association. Performance on the word form subtest, which involves simply saying whether the forms of words are the same or different, is little related to other subtests which involve making use of the meanings of words. Not one of the subjects in the control group gave a correct answer to the items on the word recognition and meaning of opposites subtest. One of the children in the control group made one correct response to the picture association subtest.

The reading comprehension test requires a written response to a missing letter or the reading and understanding of a few sentences or a brief paragraph. There was a difference of over eight months in grade placement scores on this test between the two groups. The experimental group averaged 2.9 correct responses while the control group's average was .66.

These data from the comprehension test complement the data from the spelling test which should be viewed as a test of skill in the use of phonics. The spelling task begins with the words: as, it, can, cat and fox. As Table 2 indicates there is a significant difference in performance between the groups in spelling. None of the children in the control group was able to sound out a single word while the experimental group spelled an average of 1.4 words.

The difference between the two groups on the arithmetic reasoning test was highly significant at the .005 level. This subtest

Table 3

Means and Standard Deviation of the Raw Scores of the Four subtests of Reading Vocabulary				
	Experimental		Control	
	M.	S.D.	M.	S.D.
Word Form	13.30	4.9	11.11	3.33
Word Recognition	9.50	6.2	0.00	.00
Meaning of Opposite	1.30	.17	0.00	.00
Picture Association	5.90	2.33	.11	.33

also is composed of a meanings and a problems section. The meanings subtests involve numerically identifying objects, reading number words and comprehension of size, sequences of numbers, value of coins and ability to tell time. The problems section involves simple arithmetic problems in which the answer is found by manipulating the objects on the page. The raw scores indicate that the two groups did not do equally well on the two subtests. The E.D.K. group had an average of 13.9 correct responses on the meaning subtest of the arithmetic reasoning test while the C.D.K. group had a mean of 7.2 correct responses. Neither of the two groups performed well on the problems section of the arithmetic reasoning test. The E.D.K. averaged 1.1 correct responses while the C.D.K. group had no correct responses. It would seem that most of the differences between the two groups on arithmetic reasoning can be attributed to the meanings subtest.

On the Arithmetic Fundamentals Test the differences between the two groups were substantial and significant at the .005 level. The E.D.K. group had an adjusted mean score of 1.00 while the C.D.K. group had an adjusted mean score of .14. This subtest consists of sections on addion and subtraction. The E.D.K. group had a raw score mean of 8.3 in addition and a raw score of 3.7 in subtraction while the C.D.K. group had a raw score of .11 in addition and .00 in subtraction. These data seem to indicate that the E.D.K. group learned a considerable amount about beginning addition and subtraction but that C.D.K. group learned very few facts in this area.

The scores reported in Table 4 are I.Q. scores for the Stanford-Binet and Language Age scores for the I.T.P.A. The two groups differ significantly from each other only on the auditory vocal association subtest. This task appears to be a test of verbal analogies, however, an acceptable response from the child requires only that the child respond with an "obviously" related answer. This subtest then seems to measure ability to understand extended sentences.

CONCLUSION

The consistently higher scores of the experimental children on the California Achievement Test indicate that the academically oriented program taught those children many of the skills essential to doing arithmetic and reading. Such skills and concepts would hopefully be of immediate benefit to the educationally disadvantaged child when he arrives in first grade.

There was no significant difference in the gain on the Binet and the ITPA between the two groups, although both groups made good gains on both tests. This seems an indication that the children in all the classes received strong instruction and language

Table 4

UNADJUSTED AND ADJUSTED MEAN SCORES FROM STANFORD-BINET AND I.T.P.A.

	Exp. Dual Kindergarten		Control Dual Kindergarten		Adjusted Mean Difference	P*
	Unadjusted	Adjusted	Unadjusted	Adjusted		
Stanford-Binet	99.50	100.59	97.44	96.22	4.37	N.S.
I.T.P.A.						
Auditory-Vocal Automatic	57.60	59.10	52.44	50.78	8.32	N.S.
Visual Decoding	72.10	71.32	68.11	68.97	1.35	N.S.
Motor Encoding	61.10	61.00	61.77	61.89	.89	N.S.
Auditory-Vocal Association	69.40	71.07	60.78	58.92	12.15	.05
Visual Motor Sequencing	65.60	66.34	59.00	58.18	8.16	N.S.
Vocal Encoding	60.70	61.71	74.89	73.76	-12.05	N.S.
Auditory Vocal Sequential	64.10	65.27	67.44	66.14	- .87	N.S.
Visual Motor Association	70.50	71.32	72.89	71.99	- .67	N.S.
Auditory Decoding	65.00	65.48	66.66	66.12	- .64	N.S.

P* is based on analysis of covariance of posttest scores using initial Stanford-Binet scores as the single covariate (d.f. = 1.16) and concept development. The gains made by the experimental children were less than had been experienced in the first year of instruction by groups in previous years. Several explanations for this are possible: the children were five instead of four when they entered the program, the long school day for five-year-old children, the presence of a large number of student teachers. The experienced teachers all felt that the work required of them in training the student teachers, and in making time available for the student teachers to work with the children, compromised their efficiency in their work with the children. They felt that the children did not get the intense, concentrated kind of teaching that previous groups of children had experienced. Finally, the compressing of a two-year program into a one-year program could be an important reason for less gain on these tests. The investigators did not feel that enough evidence was available at the end of the year to truly assess the feasibility of this acceleration. Some of the children were able to progress at a rapid rate through the early stages of the program, and a few of them were able to join the middle group of the second-year children. It was not possible, however, to get the lower children through the entire two-year curriculum.

As a program for the public schools, the Dual Kindergarten seems a plan worthy of further investigation. It is strongly recommended that if such a program were adopted by a school system, that there be strong articulation between kindergarten and first grade so that the skills and concepts acquired in kindergarten could be continued systematically and without interuption in the first grade. Such articulation would be particularly beneficial to slow children who needed a longer period of time to be taken through the special curriculum. It would seem, however, that a strong and continuous kindergarten and grade school program for educationally disadvantaged children would be the best possible way to maximize and strengthen the early gains derived from intense and direct instruction in academic skills in the kindergarten year.

SUMMARY

From a group of twenty children attending morning kindergarten in the public schools, ten were placed in an experimental class every afternoon held at the University of Illinois. The ten control children attended only the public kindergarten. The study was a pilot investigation of a proposed Dual-Kindergarten program for the Champaign Unit 4 Public School District. The curriculum for the experimental class was developed from a two-year language, reading and arithmetic program which had been developed for preschool and kindergarten age children by Carl Bereiter, Siegfried Engelmann and their co-workers at the University of Illinois. The feasibility of compressing the two-year program into a one-year program for older children was a part of the investigation. The program ran

from September to June. A regular school schedule was followed. The twenty children were selected from summer Head Start classes.

The Stanford-Binet and the Illinois Test of Psycholinguistic Abilities were used as pretests and posttests. The California Achievement Test, Lower Primary, was used as a posttest measure of reading and arithmetic skills. Substantial differences between the two groups of children on the California Achievement Tests indicated that the experimental group learned considerably more than did the control group about word recognition, meaning of opposites, picture association, reading comprehension and spelling. In arithmetic skills the experimental group performed significantly higher, at the .005 level, on arithmetic reasoning and fundamentals. Both groups made gains on the Stanford-Binet and the ITPA. The difference between the gains was not significant, except for the Auditory Vocal Association subtest of the ITPA, in which the experimental group scored significantly higher.

Follow-Up Data on the Achievement of Disadvantaged Children Who Participated in an Academically-Oriented Preschool Program

Ernest D. Washington and Helen B. Bereiter

INTRODUCTION

This report summarizes data collected on the achievement of children two years after they participated in a preschool devised to teach academic content to the disadvantaged. This effort, initiated by Carl Bereiter and Siegfried Engelmann, may best be described as an attempt to extend instructional technology to the area of preschool education.

In recent years, there has been a vast expansion of research into the needs and characteristics of disadvantaged children; this study was a part of that expansion. It differed from most, however, in that it sets as its goals the teaching of specific academic content rather than a general cultural enrichment. The assumption underlying this effort was that disadvantaged children lack many of the prerequisite skills for academic learning and that a carefully sequenced curriculum could fill in the gaps and unsnarl these impediments to efficient learning. With this objective in mind, the experiment had two major purposes: to detect and eliminate deficiences in learning, and to determine how much academic material could readily be learned by disadvantaged preschool children. Bereiter and Engelmann decided to center their curriculum around language, reading and arithmetic -- not because they were certain that these areas were the key to eliminating cultural deprivation, but because they knew that these areas were ones in which disadvantaged children often encountered problems in schools.

C. Bereiter and S. Engelmann (1968) described the evolution of the preschool, sketched the curriculum and have presented some data on the achievement of these children during their two years in the preschool. The task at hand, therefore, is to describe the accomplishments of these children in primary school, to discuss some of the problems they encountered there, and to provide follow-up data from the Stanford-Binet and ITPA.

METHOD

It should be remembered that this initial group was somewhat unique and was chosen on a different basis from subsequent experimental groups. Teachers in schools of a predominantly lower-class

Negro neighborhood in Champaign-Urbana, Illinois, were asked to list students experiencing difficulty in learning who came from families in which "cultural deprivation" was a factor and who had a four-year-old child in the family. On the basis of these lists, showing the children's family history and current circumstances, staff members of the preschool initially selected those four-year-olds who would be most likely to encounter difficulties in school. They then contacted the parents of these children and explained the nature of the preschool program. In only one case did the parents elect not to send their child to the preschool.

As a result of the method of selection of the experimental group and the absence of a contrast or control group, the usual inferential statistics do not apply and test norms seem to be the most appropriate standard with which to compare the current accomplishments of this group. In addition to the standardized tests, a questionnaire was available on which the second-grade teacher rated the children on their academic achievement, work habits and social development. Also available was a comprehensive case study follow-up of these children by H. Bereiter (1968). Material from each of these sources is used in describing the accomplishments and adjustment of the children.

Follow-up data are available for only twelve of the fourteen children who were reported on by C. Bereiter (1968). One child moved from the community; another is in a behavior modification class and arrangements could not be made for his testing.

RESULTS AND DISCUSSION

Table 1 shows the achievement test data for the California Achievement Tests (lower primary). These data rather than data from Intelligence and Language Tests will be the focal point of this report because it is achievement that is the goal of the preschool and of the public schools.

In looking at these data, then, scores should be interpreted with an average grade placement of 2.7 as a reference point. A glance at the total grade placement scores shows that five children in the two-year group are achieving at or above their grade level; three children are somewhat below grade placement but functioning at a second grade level, and four children are functioning at a first grade level. It should be noted that Child D was in the first grade during the past year while the others were in second grade.

In addition to the achievement test results, data are available at four intervals on the Stanford-Binet and Illinois Test of Psycholinguistic Abilities. Mental Age scores for the Binet (Table 2) and ITPA Language Age scores (Tables 3-12) are presented to give some indication of the progress made by the children while

Table 1

California Achievement Tests (Lower Primary)

Subject	Reading Vocab.	Reading Comp.	Total Reading	Total Arith.	Total Lang.	Total Batt.
A	3.0	4.0	3.3	3.0	2.8	3.0
B	1.3	1.2	1.3	1.2	1.8	1.4
C	1.3	1.5	1.4	1.6	1.7	1.6
D	1.6	1.6	1.6	1.8	2.0	1.8
E	3.3	2.8	3.3	3.1	3.2	3.2
F	2.3	2.3	2.3	2.0	2.3	2.2
G	3.3	4.0	3.6	2.4	3.6	3.3
H	2.2	2.1	2.2	1.8	2.0	2.0
I	2.4	2.5	2.5	2.0	2.0	2.1
J	2.6	2.3	2.5	2.5	2.9	2.7
K	2.3	2.1	2.3	3.0	3.4	2.9
L	1.8	1.9	1.8	1.9	1.8	1.8

Table 2

Stanford-Binet Mental Age

Subject	CA	1964	1965	1966	1968
A	4-0	5-0	4-7	6-8	8-8
B	4-8	4-5	5-0	6-0	6-6
C	4-2	4-1	5-1	6-6	7-0
D	3-11	5-0	4-7	5-4	6-2
E	4-11	4-7	5-5	5-8	7-8
F	4-11	5-3	5-11	6-11	8-6
G	4-10	4-9	5-4	6-2	9-2
H	4-2	4-1	4-10	5-7	7-4
I	4-1	4-5	5-4	6-2	8-2
J	4-6	4-7	5-9	6-8	8-6
K	4-8	4-4	5-2	6-2	8-2
L	4-10	4-6	5-3	6-2	7-6

Table 3

ITPA Auditory-Vocal Automatic

Subject	CA	1964	1965	1966	1968
A	4-0	4-3	6-1	6-1	6-6
B	4-8	2-4	3-10	3-10	5.0
C	4-2	3-6	4-3	4-3	6-1
D	3-11	2-4	3-1	3-6	4-7
E	4-11	4-7	6-1	5-4	8-0
F	4-11	2-4	4-3	5-0	8-0
G	4-10	4-3	4-7	5-9	8-0
H	4-2	3-6	6-10	4-3	6-6
I	4-1	3-10	5-0	5-0	8-0
J	4-6	4-3	5-9	6-10	7-3
K	4-8	3-10	5-0	7-3	9-1
L	4-10	4-3	4-7	6-1	6-1

Table 4

ITPA Visual Decoding

Subject	CA	1964	1965	1966	1968
A	4-0	4-5	6-8	5-10	8-9
B	4-8	4-1	3-4	4-9	6-3
C	4-2	4-1	4-5	5-10	6-3
D	3-11	4-1	5-6	6-3	5-1
E	4-11	2-8	4-9	4-1	6-8
F	4-11	3-4	6-3	8-9	8-9
G	4-10	4-5	4-5	4-9	7-3
H	4-2	3-8	4-9	6-8	6-3
I	4-1	2-8	5-6	5-6	8-9
J	4-6	4-5	6-3	5-10	7-3
K	4-8	4-1	5-2	6-8	7-1
L	4-10	2-8	2-8	5-6	7-3

Table 5

ITPA Motor Encoding

Subject	CA	1964	1965	1966	1968
A	3-11	5-5	4-2	6-4	6-10
B	4-8	2-11	4-2	3-2	5-5
C	4-2	2-11	2-11	5-10	5-5
D	3-11	2-9	4-2	4-2	5-0
E	4-11	3-6	4-2	5-5	5-10
F	4-11	5-0	5-5	6-10	8-8
G	4-10	2-6	2-6	2-6	5-10
H	4-2	3-2	4-2	3-10	6-4
I	4-1	3-6	4-7	5-0	7-4
J	4-6	3-2	5-5	4-2	7-4
K	4-8	2-11	3-10	6-4	8-8
L	4-10	3-2	4-2	6-10	8-8

Table 6

ITPA Auditory-Vocal Association

Subject	CA	1964	1965	1966	1968
A	4-0	3-8	4-11	6-1	8-3
B	4-8	2-8	4-5	5-3	5-10
C	4-2	3-1	4-5	5-3	6-10
D	3-11	2-10	3-6	4-8	6-6
E	4-11	4-8	5-10	6-10	9-0
F	4-11	4-8	5-3	6-10	7-3
G	4-10	4-8	5-6	6-6	7-8
H	4-2	3-8	4-11	5-6	6-10
I	4-1	3-11	4-8	6-10	7-3
J	4-6	4-2	4-5	7-3	7-8
K	4-8	4-11	4-2	5-3	8-3
L	4-10	3-11	4-8	5-10	6-6

Table 7

ITPA Visual-Motor Sequencing

Subject	CA	1964	1965	1966	1968
A	4-0	3-11	3-0	4-10	5-8
B	4-8	4-7	4-4	6-4	6-4
C	4-2	4-7	5-8	5-1	5-4
D	3-11	2-7	4-2	5-4	5-4
E	4-11	5-1	5-1	5-1	5-8
F	4-11	4-2	4-10	5-8	6-9
G	4-10	4-10	5-4	4-10	8-5
H	4-8	3-8	3-4	5-4	9-0
I	4-1	4-10	3-4	5-1	6-4
J	4-6	3-11	5-8	5-4	5-4
K	4-8	4-7	2-7	3-11	8-5
L	4-10	3-0	3-8	4-4	5-8

Table 8

ITPA Vocal Encoding

Subject	CA	1964	1965	1966	1968
A	4-0	5-4	8-11	5-8	8-11
B	4-8	3-2	5-1	6-4	4-9
C	4-2	3-2	5-4	5-1	5-1
D	3-11	2-3	6-4	6-7	4-5
E	4-11	5-1	6-0	6-7	8-11
F	4-11	4-5	5-1	6-11	8-11
G	4-10	3-10	3-6	5-1	6-7
H	4-2	2-3	6-0	4-9	5-4
I	4-1	3-2	5-4	8-11	6-7
J	4-6	2-3	5-4	6-4	6-11
K	4-8	3-10	5-1	6-11	7-4
L	4-10	2-7	5-4	4-9	7-4

Table 9

ITPA Auditory-Vocal Sequencing

Subject	CA	1964	1965	1966	1968
A	4-0	5-11	8-6	8-6	8-6
B	4-8	3-3	2-11	5-11	5-1
C	4-2	6-7	6-7	8-6	8-6
D	3-11	3-1	3-7	5-4	5-11
E	4-11	4-7	5-11	6-3	7-4
F	4-11	6-3	8-6	8-6	8-6
G	4-10	3-9	3-7	5-4	6-3
H	4-2	3-5	4-7	5-1	6-3
I	4-1	5-4	8-6	8-6	8-6
J	4-6	6-7	5-11	7-4	8-6
K	4-8	4-7	4-7	5-1	5-11
L	4-10	4-2	4-10	4-10	5-11

Table 10

ITPA Visual-Motor Association

Subject	CA	1964	1965	1966	1968
A	4-0	4-4	6-6	5-5	8-7
B	4-8	2-11	6-1	7-10	7-6
C	4-2	3-8	4-8	4-8	6-1
D	3-11	2-11	7-2	6-1	6-1
E	4-11	4-4	5-9	5-1	7-6
F	4-11	4-4	5-9	5-1	5-9
G	4-10	2-11	6-6	5-5	8-7
H	4-2	2-11	7-6	7-6	7-6
I	4-1	5-1	4-0	5-9	8-3
J	4-6	4-0	5-1	7-6	8-7
K	4-8	4-4	4-0	4-4	8-7
L	4-10	4-4	8-3	4-8	7-6

Table 11

ITPA Auditory Decoding

Subject	CA	1964	1965	1966	1968
A	4-0	5-2	6-2	7-11	8-1
B	4-8	4-5	4-7	5-5	5-1
C	4-2	4-1	5-2	6-9	4-5
D	3-11	4-5	3-10	5-0	8-1
E	4-11	4-1	5-8	4-7	7-1
F	4-11	5-0	4-7	7-11	7-1
G	4-10	4-5	3-10	4-5	7-6
H	4-2	5-0	6-9	5-2	5-5
I	4-1	4-7	4-9	6-9	5-0
J	4-6	4-7	4-3	7-1	8-1
K	4-8	3-10	3-10	6-5	8-1
L	4-10	4-9	4-9	4-3	8-1

Table 12

ITPA Total

Subject	CA	1964	1965	1966	1968
A	4-0	4-9	5-11	6-6	8-3
B	4-8	3-5	4-3	5-6	5-9
C	4-2	4-0	4-11	5-9	5-9
D	3-11	3-1	4-6	5-3	5-1
E	4-11	4-4	5-6	5-5	7-3
F	4-11	4-6	5-5	6-8	8-1
G	4-10	4-0	4-3	4-11	7-4
H	4-2	3-7	5-4	5-4	6-8
I	4-1	4-3	5-1	6-7	7-9
J	4-6	4-3	5-3	6-5	7-9
K	4-8	4-2	4-3	5-8	8-5
L	4-10	3-7	4-9	5-1	6-1

in the preschool and in the public schools. In examining these data, it should be remembered that the first complete testing of these children occurred after four-to-five weeks of schooling, rather than at the very beginning of the preschool experience. It is during these initial few weeks that the children generally become less concerned about their new surroundings, learn to attend to the teacher and generally get on with the business of becoming students rather than observers. Weikart (1967) includes a characterization of the adjustment of children to a preschool in a very interesting fashion: Stage One -- Silence; Stage Two -- "Des god damn peaches am burning"; and Stage Three -- "These god damn peaches are burning." By this cryptic description, he shows that the child is at first anxious, he then begins to talk and, finally, he begins to learn appropriate classroom behavior. The first two stages are traversed very quickly, and it is here that most preschool programs make their gains and their impact.

When one looks at these data, it is quite noticeable that some of the subtests of the ITPA show a great deal of variability. The tables which show the least variability and which are the most reliable are the Mental Age scores of the Binet and the Total score of the ITPA. It also happens that these scores are the most difficult to interpret. On both of these tests, the children show a fairly consistent shift from below average to above average during their preschool years.

These data on Mental Age and ITPA Total also indicate something about the differential effect of the performance of the children in the preschool and in the public schools. While in the preschool, all of the children showed gains in the Binet and ITPA, although there was some indication that the lower-achieving students were doing less well. With the change from preschool, those children who have not been doing well in the preschool began to make even less progress on the Binet and ITPA. The data from the subtests of the ITPA mainly corroborate the data from the ITPA, with the added disadvantage of a ceiling effect. On several of the subtests, one or more of the children have reached their ceiling, and little additional information can be gained from giving all of the subtests when the children are re-evaluated.

Since the children are so few in number, it would not be inappropriate to make a few comments about each one. This may help give greater insight into their behavior in the preschool and the public school and also point out the effect the children's homes may have had on their behavior in the school setting. While it is true that the case study approach can be somewhat misleading because it ignores normative kinds of data and may concentrate on idiosyncratic behavior, this is not necessarily a flaw in the case of disadvantaged children. For it is just such stuff with which the school must deal. To ignore these data is to ignore the realities of the circumstances in which the child lives.

Four of the children who participated in the initial academic preschool are functioning well below grade level; and it is important to examine their current status and try to learn something that will be of help in teaching other children. It isn't enough to say that these children are not successful; one must look to find the shortcomings of an approach and then search for ways to remedy the inadequacies in the program.

First we ought to consider Child D, for it now appears that she is a full year younger than the other children who participated in the preschool. The mother was not available when D entered the preschool, and D's birthdate was recorded incorrectly. Her progress was very slow during her first year, but she began to move at a faster rate during her second year. She functioned somewhat below average during her year in first grade. Because she was younger than her classmates, her teacher and her mother decided that D should repeat first grade even though she was functioning at a higher level than some of the children who went on to second grade. When considering her grade placement score of 1.8, therefore, it should be remembered that while somewhat below expectation for someone repeating first grade, it is appropriate for someone of her chronological age.

L lives with mother, father, and four siblings. The father is a seasonal truck driver, and the mother works in a hospital laundry. Their home is a depressing structure, among the worst in the community. During his years in the preschool, L's performance was very disappointing. He learned things slowly and could not seem to transfer his learning to new situations. He applied himself diligently and constantly sought positive attention and feedback, but he seemed to learn very little. His educational performance did not change during first grade. He learned very little and was eventually placed in the lowest group in his class. During first grade, his behavior became more and more disruptive, and he was placed in a behavior problem class during his second year. With the judicious use of positive reinforcement, he is no longer a disruptive influence in the classroom. His teacher now reports that he is adjusting well in class and that his educational performance is improving.

B and C, cousins, are both performing at the first-grade level -- a performance, we should note, that is in line with the expectations of their families. Neither of these children made very much progress during the preschool, but for seemingly opposite reasons. C learned quickly but seemed to retain very little from day to day, while B learned with agonizing slowness but seemed to retain what she did learn. Both come from the same large extended family which exists in extreme poverty. B is being reared by her grandmother who is less and less able to provide supervision and direction. Her mother has never been available to provide information about the family, and the grandmother could, or would, provide very little such information. B did very poorly in first grade and did not get along well with her teacher or her peers. Her teachers feel that she is uninterested in school, and her inappropriate behavior and dress have alienated her from her peers.

C lives with his mother, father, and four siblings, but much of his supervision comes from his grandmother. His mother has told staff members that he is "dumb" and she does not seem to believe that C can learn very much. While in the preschool, his IQ changed from 95 to 117; but this change in score seemed unrelated to his learning in the preschool, for he learned very little. When he entered the public schools, his teachers quickly found out that he knew little and seemed to make the quick judgment that he could not learn.

What, then, can we surmise about the reasons for the failure of B, C, and L? In looking at their classroom behavior, we find that they have not been successful for varied reasons; but we can say generally that they knew little when they came, the learned very little, and, probably most important, they seemed to have little interest in learning. It is obvious that very little can be concluded from such a small sample, but if we look at the other groups of children who participated in the Bereiter-Engelmann program, we find almost always that 20 to 25% of the children do not seem to profit from the preschool experience. Weikart (1967) also reports that about the same percentage of children in his experimental program appear to derive no benefit from the experience.

One has only to look at these children to understand why there is so little success in this group. These are the children who are deeply immersed in poverty. These children come from multiple-problem homes. Their families are on the bottom of the socio-economic scale and neither they nor their families see much chance of improving their lot in life. These children and their families require more than a preschool experience to help them out of the depths of their disadvantage. It is unrealistic to hope that any preschool program, or anything the public schools can do, can have much impact on their lives without help on a massive scale.

As contrast to the children described above, let us now give a brief account of the children who have been functioning nearer to their grade level, or above it.

Child A comes from a family which is intact; his father is a construction-laborer and his mother works as a waitress. The child has four older siblings who have consistently performed very poorly in school, but his achievement tests place him above grade level. His scores are also well above average on the Stanford-Binet. At the end of the preschool, he ranked at the class median in achievement; and the few problems he has had during preschool and in public school have had more to do with motivation than with "ability to learn." While it is difficult to separate motivational considerations from learning, it is possible to surmise that the child probably benefits most from gaining a greater perception of his own competence. His brothers and sisters had always performed poorly in school, and his parents had no reason to believe he would do any better. But when he began to return from the preschool and demonstrate his new knowledge at home, both he and his family appeared to gain new insights into his abilities. The belief that he could do well at school began to reinforce his achievements.

E is now performing well above grade level, but she began preschool as a hyperactive student who had difficulty adjusting to academic instruction. Under more "traditional" circumstances she might have been labeled in some fashion and placed in a special class; instead, a serious attempt was made to teach her academic materials and appropriate behavior. During her second year in the preschool, her parents were divorced. It is possible that tensions created by this situation at home were reflected in her erratic performance at school. After a while, the defiant behavior at school all but disappeared and the child responded well to classroom instruction. E is now participating in a class for promising disadvantaged children.

Child G, whose father is a cook and whose mother occasionally works as a waitress, learned easily and well during her two years in the preschool and is now performing well above grade level. When she entered an accelerated middle-class white school, she encountered many of the problems common to disadvantaged children. Although she was ahead of her peers in knowledge of academic subject matter at the beginning of first grade, she lacked the breadth of experience her peers possessed. During first grade she was ahead of the class for some time, but gradually her classmates began to catch up. She is now functioning at about an average level in the class. It is highly likely, however, that she would not have been able to remain in an accelerated class if she had not possessed some academic skills when she entered the public schools. She would have lagged behind from the beginning and never had a chance to catch up.

In contrast to her first grade performance in the middle track in her school, Child H is now functioning below her grade level in the second grade. A member of a large extended family, she saw very little of her mother during her preschool years. After her mother's remarriage just before H entered first grade, H went to live with her mother, step-father, and two younger siblings in an attractive bungalow in a lower-middle class neighborhood. However, H was assigned to attend a school located in a different neighborhood, one which was considered very anti-Negro. She encountered great difficulty adjusting to this school, and for second grade, her mother placed her in a Catholic school. Her teacher there reports that her reading skill is that of an average second grader, but that she reads very slowly. Her slow reading rate often reflects itself in poor reading scores.

Child I has recently completed second grade and is functioning somewhat below his grade level. Both parents are employed full-time in rather menial jobs and have little time to devote to their four children. Though the child is not functioning to expectations, he is the star pupil in his home and often helps his older siblings with their homework. He was not highly motivated to achieve during the preschool; and teachers in the public schools have noted that he has the ability to learn but rarely has much interest in learning. An important factor in this lack of motivation can doubtless be found in the attitude of his parents: they assume their children will probably never finish high school and have expressed this opinion both to child I and to his older siblings.

J is completing second grade in a class for intellectually promising lower-class children. She learned a great deal during her years in the preschool and was held in high esteem by her classmates for two unrelated reasons: she was smart and she was light-skinned. Because of her fair skin and preferential treatment at home, she fastened on to the notion that she was white. For obvious reasons, this notion caused her some problems in a racially-mixed elementary school. While she was in first grade, her father's long illness and subsequent death affected her interest in attending school. Her year in second grade, however, seems to have been more productive, and she currently enjoys school more than she has in the past.

K was completing second grade at the time of testing and ranked in the top third of the children, well above grade placement. His achievement scores in the preschool had also placed him in the top third of the class. During his time in the preschool, he lived with his mother and seven siblings. His father had not been living in the home and, as a matter of fact, was killed during his first year in the preschool by a close relative of two other children in the class. His father's desertion and death did not seem to affect K's behavior or learning in class, but very likely did affect that of his older brothers and sisters. K and his younger brother, who also attended the academic preschool, are doing quite well in school; the older children are all having difficulties. Ironically, it was discovered only late in his first grade year that his progress has probably been impeded by a severe hearing loss.

SUMMARY

This report provides some follow-up data on the achievement of the initial group participating in an academically oriented preschool. The children for this experiment were chosen because

they came from homes where their siblings were doing poorly in school and where cultural deprivation seemed to be a factor in their failure. They received instruction in reading, arithmetic and language for one hour per day over a span of two years and then entered the public schools. Data for this report covers the children two years after they participated in the program.

A goal of the preschool was to bring the children up to their age norms on school-relevant language and intellectual abilities. At the end of the preschool, the children had accomplished this goal, as indicated by their scores on the Stanford-Binet and the Illinois Test of Psycholinguistic Abilities. They also made considerable headway in arithmetic and reading.

Two years later, twelve of the original fifteen children were available for follow-up. Five of these children were functioning above grade level, three were somewhat below grade level, and four were well below grade level on the California Achievement Tests (lower primary). Data from re-administrations of the Stanford-Binet and the Illinois Test of Psycholinguistic Abilities showed that children in the lower-achieving groups had not maintained their gains on those measures as well as the high achievers.

Achievement Components of Stanford-Binet Performance

Ernest D. Washington, Siegfried Engelmann, and Carl Bereiter

PROBLEM

Strodtbeck (1964) has introduced a useful term for discussing the school-relevant aspects of social-class differences. He speaks of the "hidden curriculum" of the middle-class home, which provides children with those unspecified learnings that constitute adequate cognitive and behavioral preparedness for school. What is this hidden curriculum and what are the crucial parts of it that disadvantaged children miss? So far the question has been answered mainly by conjecture and the conjectures have tended to fasten upon the grossest and most obvious differences between middle- and lower-class childhood experience.

The hidden curriculum is, of course, a hypothetical construction, and so it would be futile to argue about what it really comprises. The most that could be hoped for in the way of definition is a set of specifications that 1) are in accord with the facts of childhood experience and behavior, 2) have some demonstrable relevance to subsequent academic performance, 3) are expressed precisely enough to permit objective evaluation, and 4) can rather directly be translated into pedagogical procedures or plans. Most of what is done in the name of "stimulation" or "enrichment" in early compensatory education can be viewed as an attempt to implement a hypothetical, hidden curriculum. But this underlying curriculum, to the extent that it is described at all, is specified in ways that fail on all or most of the above criteria.

In light of the hopes currently invested in preschool education, it would seem mandatory to explore more systematic ways of formulating the content of the implicit curriculum that the preschool purports to teach. In our work to date we have attempted to do this through the analysis of key bodies of subject-matter--expository language, reading, and arithmetic. The strategy has been that of working backward from more advanced curricula. Through this analysis and the ensuing curriculum experimentation, we have, it is believed, identified some of the critical preschool learnings relevant to these particular areas of achievement. We now propose a different approach to identifying the content of this hidden curriculum, which promises to encompass learnings of more general utility for academic achievement. It amounts to working upward from test content rather than backward from more advanced curriculum requirements. It should thus add another dimension to the work done previously.

The central idea behind the new approach is that the Stanford-Binet may be considered as an achievement test for this hidden curriculum. Ordinarily the Stanford-Binet, along with other tests of general intelligence, is considered as a measure of basic capacity for intellectual attainment. However, the Stanford-Binet purports to measure this basic capacity by testing the child's achievement in a number of areas not directly touched by the school curriculum. It was assumed that what the child just "picked up" from incidental experience would provide a surer index of his basic intellectual capacity than what he had been taught in school (since the latter might be influenced by the kind and quality of instruction he had received). As many recent critics have pointed out, this assumption only holds if children's out-of-school environments are faily similar. But turning this point around, we may say that for children of different cultural backgrounds, who are presumably equal in basic mental equipment, differences in Stanford-Binet scores reflect the differential success of cultural environments in promoting those learnings that undergird academic achievement.

If, therefore, we wish to identify those elements of the out-of-school curriculum that are significant for school success, the content of the Binet items may provide us with some valuable guidance. We cannot presume that the authors who have taken part in development of the Stanford-Binet had the outlines of a "hidden curriculum" in mind in constructing the items. But the items that have gained a place in the Binet had to pass certain empirical tests that afford some reason to believe that they may validly reflect the content of such a curriculum nevertheless. They had to show a sharp age progression, indicating that the achievements were obtained at about the same time by most children. In addition, they had to have predictive validity for later academic performance, which does not necessarily mean that the achievements tested are themselves instrumental for later achievement, but it makes the hypothesis tenable. Other intelligence tests meet the same empirical criteria, but they do not contain the great diversity of substantive content that the Binet does, and, therefore, they are less promising as a guide to the identification of relevant content in the hidden curriculum.

The problem to which the present study is addressed is that of deriving a curriculum from a test. It is much like the problem one would face if he were shown the final examination for a high school course and was asked to deduce from it the course content. If the test satisfactorily sampled what was actually taught, the job would not be impossible and one could expect different people to come up with similar, though not identical, approximations to the actual curriculum. If the test were good but the curriculum was shoddy or not followed by the teacher, the inferred curriculum might bear little resemblance to the actual one; but one would then be inclined to say that the inferred curriculum is the one that should have been followed if the test indeed reflects the hoped-for outcomes. That is the situation with deriving a curriculum from the Stanford-Binet: it does not represent a validatable guess as to what actually goes on in the middle-class home; it represents an ideal construction of what curriculum ought to be followed, in some fashion or other, in home or at school, if the Stanford-Binet indeed reflects the hoped-for outcomes of early learning experience.

In using a test as the basis for curriculum planning, it is important to distinguish curricula generated from content specifications and curricula generated from item types. Much current remedial work on learning disabilities uses curricula generated from item types. If a child exhibits inferior performance on a certain kind of item, say verbal analogies, he is given practice on a variety of verbal analogy tasks. Such practice often results in improved test scores, although it is always questionable whether such training will generalize beyond performance on the particular item type used in training. The assumption behind this kind of training appears to be that it will alter basic organismic characteristics, thus resulting in a change in school performance. Thus, supposedly, what the child gets out of training on verbal analogy tasks is not simply learning relevant to performance on these tasks, but some change in central nervous system functioning that will increase his capacity for other kinds of learning and performance. The similarity of this assumption to that of faculty psychology should be obvious. When applied to compensatory education for disadvantaged children, this assumption carries the implication that lower-class children differ from middle-class children in organismic characteristics and not merely in what they have learned. This hypothesis is tenable, although there is no evidence in its favor.

If a compensatory curriculum is to be based on hypothesized learning deficits rather than hypothesized organismic deficits however, this requires that test items be analyzed into learnable content components and not treated as integral units of learning. The hidden curriculum of the home does not teach or fail to teach verbal analogies as such. Indeed, the verbal analogy task, like most other test tasks, was deliberately chosen as one on which children would not have received specific training. Such specific training would simply destroy the predictive validity of the item, while teaching a skill that was selected because of its peripheral and hence very likely useless nature. We must assume, instead, that what the home environment typically teaches are certain bodies of information, relational concepts, skills, and attentional and procedural habits, which help the child execute verbal analogies tasks as well as a number of school tasks. The job for curriculum development, therefore, is to identify these underlying bodies of content and teach them rather than treat the test items as if they represented the ultimate tasks for which children were being prepared.

OBJECTIVES

This study began with three related objectives:

1. To carry out a task analysis of the achievement components of the Stanford-Binet items in the three-to-six year old range.

2. To construct and implement, through direct instruction techniques, a curriculum based on content categories identified through the above analysis.

3. To evaluate and revise the original analysis on the basis of comparison of test item performance and achievements in the curriculum.

As has been mentioned previously, the Stanford-Binet was chosen because of its diversity of content. It was assumed that this range of content would lead via task analysis to a curriculum which encompassed a wide range of verbal skills. As the analysis proceeded, it became clear that this was indeed the case. Moreover, it was also apparent that many of the achievement components of the Stanford-Binet could be taught in a very parsimonious manner. The discussion here will focus on the verbal items of the Binet because as one inspects the changes in item content from year three through six, the required responses change from predominantly nonverbal to verbal. The verbal items have the highest validity for predicting later school achievement and this is reasonable because the schools are institutions based on oral and written verbal materials.

The task analysis of the vocabulary subtest provides an example of the manner in which the task analysis began and the final direction it took in being incorporated into the curriculum. Concrete nouns from the Dolch (1936) list of "The First Thousand Words for Children's Reading" were chosen to teach vocabulary. This list was chosen because it avoided many of the problems involved in sampling from dictionaries and a useful vocabulary could be taught without the implication that the vocabulary of the Stanford-Binet was being expressly taught.

When one looks to what are considered correct responses for the vocabulary subtest at the six-year level, nine of the first ten words are concrete nouns and the requirements for a correct response is usually that an attribute of the noun be given. Looking over the first fifteen words, the following attributes were found to generate questions whose answers were frequently sufficient for a plus response.

1. What are its physical dimensions?
2. Where is it found?
3. What are its uses or what purposes does it serve?
4. What is it made of?
5. What are its parts or of what things is it a part?
6. What are its special sensory or personal characteristics?

As the analysis proceeded, the teaching of the above attributes served two purposes. It aided in teaching knowledge which was applicable and useful for responding to other subtests. In addition, these questions implied other dimensions which should be taught in addition to those above. Thus, if use or purpose was taught, then this task might include comprehension as it occurs at year IV in the Binet. The teaching of "Where is it found?" also inferred the teaching of locations, "What is found in this location?" Thus the teaching of these attributes with the analysis of the other items led eventually to a more general list of attributes or concepts.

The final list of attributes was expanded to include the following: size, color, shape, part of, action, location, use, material, number and order. Teaching each of these concepts requires a knowledge of certain terms and grammar. After these basics were taught, each of the concepts was used in teaching similarities, differences and absurdities. We thus had a twelve by three matrix in which many different kinds of things could be taught in breadth and reinforced in both the similarity-difference and incongruity format (Table 1).

A second and very important characteristic of the content implied by the matrix is that the content can be taught at various levels of difficulty. Concepts and Attributes could thus be taught at the various levels in the same-different and incongruity format. Consider the various levels of difficulty at which Materials could be taught.

A. Same-Different.

1. a. Which one is the same--as this one?
   b. Which ones are the same?
2. Yes-no. Are these two the same?
3. Description. How are these two the same? Different? (Used with pictures)
4. Materials from memory. What kinds of materials are used in _____?
5. Description from memory. Like number three but without pictures.
6. Which materials could be used for _____? Which ones could not be used for _____?

Table 1

Similarities, Differences, and Incongruities

This is a basic knowledge program built entirely around judgments of sameness and difference and detection of incongruities and absurdities. Same-different judgments and incongruity judgments run in parallel, covering the same set of attributes, as illustrated in the following list:

Attribute	Language	Same-Different	Incongruity
		Find the ones that are/are not	What's wrong in this picture?
A. size	terms; big, little, short, tall, etc. grammar: second-order statement patterns, multiple polars: later comparatives -- this pencil is wider than this pencil	the same size	Disproportionately large or small (shoe too small for foot)
B. color	color names grammar: second-order statement patterns, multiple polars, comparatives, superlatives	the same color	Unnatural color (pink tree)
C. shape	names of shapes grammar: second-order statement patterns, multiple polars	the same shape (pumpkin and apple)	Wrong shape or mis-shapen (square wheels, bent pencil)
D. part of	basic statement: A is a part of B	part of the same thing (pedal and handle-bars)	Missing or wrong part

(Table 1 continued)

Attribute	Language	Same-Different	Incongruity
E. action	tense, person, number adverbs: fast, slow, hard, soft, etc.	doing the same thing (at more advanced level, find ones that do or don't do same thing, action not shown)	Impossible or inappropriate action (dog reading book)
F. location	terms: place names, prepositions. grammar: second-order statements with prepositions	found in the same place (stove and fridge vs. gas pump)	Out of place (ship in forest)
G. use	terms: examples of uses of various objects. grammar: second-order statements, multiple polars	used the same (spoon and fork vs. umbrella)	Inappropriate use (eating a book)
H. material	names of materials grammar: second-order statements, multiple polars	made of the same thing (plastic pen and comb)	Impossible or impractical material (brick airplane)
I. number	terms: rate counting, counting of objects, number identification -- first, second, third, etc.	"Find the ones that have the same number of ___s."	Wrong number (six fingers or not enough cookies for each child)
J. order	examples of sequences	In the same order (large to small, or arbitrary order of given figures)	Wrong order (wrong sequence in action or growth series)
K. genus	learning of class names, example: truck, car are included in class of vehicles	in the same class (animals furniture, etc. -- will overlap with other attributes)	Foreign element in class (bird among mammals)

7. Compound material task: Find the one that is located in the same place.
8. Compound verbal identification tasks: I'm thinking of something that's the same size as A and is made out of the same material as B.

B. Incongruity. (Materials)

1. Point to the material that does not belong.
2. Yes-no. Is there anything wrong with this boat? Yes.
3. Explanation: What's wrong with _____?
4. Which material does not belong?
5. Description from memory. What are _____ made of?

Language usage, same-different and incongruities do not exhaust the formats for teaching the attributes and concepts. Bereiter, Case, and Anderson (1968) have suggested four other promising formats for teaching these concepts. The first they call knowledge, that is teaching the facts and principles which go beyond what the child already knows so that through guided cues the child can learn to extrapolate his knowledge. The second is productive thinking, that is, teaching the child to use concepts to solve problems. The third is operations or nonverbal tasks which involve getting or using information related to concepts. Finally, questioning is a format which teaches the child to ask questions about the concepts being taught.

The concepts and attributes are not mutually exclusive and certainly the various formats for presenting the concepts are not finalized. It remains to be seen whether these formats are more fruitful than others or if some combination of two formats is sufficient. Little effort was made to use the first three additional formats suggested by Bereiter although some attempt was made to teach the children self-questioning. That is, how does one go about teaching disadvantaged children to ask a question and to use the answer to formulate yet another question to solve a problem. Some beginning was made in this problem with the game which the children played analogous to twenty-questions in which the teacher responded only with yes-no.

At a low level of difficulty, the game began with the teacher placing two figures on the board and saying, "I am thinking of one of these figures." At the beginning the children often will claim that they know which of the figures the teacher is thinking about. These children are told that they can not know what figure the teacher is thinking about unless she gives some clue. It is possible to dramatize the point by showing the child two closed hands and asking the child, "Which hand has the raisin?" In the beginning neither hand has a raisin and the child is rewarded for saying "I don't know."

Later figures which differ on two dimensions are presented on the blackboard and the child is taught that if one dimension is eliminated, the correct response is the remaining dimention. This general approach is extended to several dimensions with the child learning to eliminate alternatives. The terminal task of interest is verbally presenting the child with a class name such as animals and having the child figure out which animal the teacher is thinking about.

METHOD

Subjects

Subjects were 20 children selected according to the same socioeconomic criteria as in previous studies in this project. Table 2 shows some of the characteristics of the children who participated in this the "Binet" study and in two other approaches to preschool education in the larger research program at the University of Illinois.

Treatment

The preschool ran for two hours a day, five days a week for the academic school year. Eighty minutes were devoted to instruction while forty minutes were used for supporting activities. During the first four months of the program, the children remained with one teacher for instructional purposes; during the remaining three months the children went from one class to another in much the same manner as children in the upper grades.

Testing Procedure

The Stanford-Binet (S-B), Wechsler Preschool Intelligence Scale (WPPSI), and Illinois Test of Psycholinguistic Abilities (ITPA) comprise the battery used to evaluate the progress of the children. The testing schedule used is presented in Table 3.

Qualified school psychologists administered the WPPSI, S-B and ITPA while undergraduate assistants administered the curriculum test. The S-B was given four times so that the effectiveness of the curriculum could be repeatedly assessed. The Wechsler provided an independent check of the effectiveness of the curriculum to provide learning with some generality. The Achievement Test (Appendix C) was devised by the staff to assess the content of the curriculum.

Table 2

Initial Group Composition

Program	N	Mean Binet CA	Mean Binet IQ	Intelligence Strata Means						Race - Sex			
										Black		White	
				High	N	Middle	N	Low	N	M	F	M	F
Binet	20	48.9	92.8	105.0	4	95.3	10	80.6	6	6	8	4	2
Traditional	27	52.4	94.1	107.6	8	93.2	12	83.0	7	10	7	5	5
Bereiter (Direct Verbal)	29	50.7	93.2	108.5	8	94.1	9	82.4	12	9	9	4	7

Table 3

Testing Schedule

Instruments	Time
$WPPSI_1$, $S\text{-}B_1$, $ITPA_1$	0-3 months before school
$S\text{-}B_2$	2 months after the start of school
$S\text{-}B_3$	5 months after the start of school
Curriculum Achievement Test	8 months after the start of school
$WPPSI_2$, $S\text{-}B_4$, $ITPA_2$	at end of school

RESULTS

Table 4 indicates a mean gain of 12.9 on the Stanford-Binet over the year. Almost half of this gain (5.9) was achieved within the first two months of the program. During the following three months the children gained an additional 4 points. It is of some interest to note that the children gained, on the average, only 2.8 points during the final three months of the program.

The time periods are of some interest. The data indicates that the 6 points gained in the first three months represent the total gain of most preschool programs. The data also corroborate the findings of Kohlberg (1968) who noted that children in most structured preschool programs tend to gain from 12 to 16 points during the first year.

Table 4

Mean Stanford-Binet Scores for the Binet Group at the Four Intervals

Months of Intervention	0	2	5	8
Mean	92.8	98.8	103.0	105.8
S.D	9.67	10.77	9.21	10.50

Table 5 gives the scores on the Verbal, Non-Verbal and Total scores for the WPPSI. Table 5 shows that the children gained 15.4 points on the Verbal, 7.0 on the Non-Verbal and 12.6 on the Total score. These data indicate clearly the emphasis in the preschool program upon verbal skills and the relative de-emphasis on non-verbal skills. The total score is simply a combining of the verbal and non-verbal scores and is not nearly as informative as looking at the verbal and non-verbal scores.

Table 5

WPPSI, Pre- and Post-Test Scores

	Pre-Test	Post-Test
Verbal Mean	82.5	97.9
Non-Verbal Mean	85.3	92.4
Total Mean	82.2	94.8

It should also be noted that the Binet and WPPSI data are similar in terms of total gains but differ markedly in terms of level. The final Binet IQ is 105.8 while the final WPPSI total IQ is 94.8. The ten point difference between the WPPSI and the Binet is in line with other data obtained in other studies when both of these tests were used. Disadvantaged youngsters generally score about ten points lower on the Wechsler Intelligence Scale for Children than on the Binet.

Table 6 shows an analysis of Binet and WPPSI IQ gains by IQ strata or level. The data from the Stanford-Binet indicate that the high group, those children with initial IQ's over 100, had a mean gain of 3.7; those with IQ's between 90 and 99 had a gain of 14.7, and those children with IQ's below 89 had a mean gain of 16.2. These data suggest that the program was more effective for children with IQ's below 100, and that children with IQ's over 100 tended not to benefit as much from the program.

The WPPSI data present quite another picture. The WPPSI was included in the analysis to give an independent measure of the effectiveness of the preschool effort. The WPPSI verbal data indicate that all three strata benefited substantially from the program. The WPPSI non-verbal data suggest that limited gains were made by each of the strata. The total WPPSI scores indicate little difference in gains by strata.

A paired t test was used to measure the differences on the WPPSI between the pre- and post-tests (Table 7). On the verbal score this difference was significant beyond the .001 level. The difference between pre- and post-testing on the non-verbal WPPSI was smaller and significant at only the .05 level. The total score reflects the large gains in verbal skills and again the difference between the pre- and post-test was significant beyond the .001 level.

PREDICTION OF STANFORD-BINET PERFORMANCE FROM COURSE ACHIEVEMENT

To the extent that (a) the original conception of the Stanford-Binet as an achievement test is valid, (b) achievement components of Stanford-Binet performance were accurately identified, (c) the experimental curriculum embodied these components, and (d) the Achievement Test given at the end of the experimental program accurately indicated the children's degree of mastery of these separate components, then it should be possible to predict from Achievement Test results which Stanford-Binet items individual children would pass and fail. Accordingly, as a check on whether or not the above conditions jointly obtained, a small study was conducted on prediction of S-B item scores from Achievement Test item scores.

Table 6

IQ Gains by Intelligence Strata

Initial IQ Strata	Stanford-Binet	WPPSI Verbal	WPPSI Non-Verbal	WPPSI Total
High	3.7	14.5	5.5	10.0
Medium	14.7	17.1	7.5	13.8
Low	16.2	13.2	7.3	11.8

Table 7

Paired t Tests on Verbal, Non-Verbal and Total Scores of the WPPSI

	Pre-Test IQ	Post-Test IQ	Difference	Paired Standard Error	t	Level of Significance
Verbal	82.5	97.9	15.4	2.22	6.91	.001
Non-Verbal	85.3	92.3	7.0	2.45	2.86	.05
Total	82.2	94.8	12.6	1.91	6.59	.001

One of the investigators (Bereiter), who was not present during the conduct of the experiment and thus was not acquainted with the subjects, but who was acquainted with the curriculum, constructed prediction formulas for predicting success or failure for separate items on the Stanford-Binet, over age levels IV through VII, employing item response data from the Achievement Test. Obtained scores on the Achievement Test were consulted in devising the prediction formulas, but not obtained S-B scores. In other words, the predictions were carried out in ignorance of the individual S-B scores. The formulas, instead of making a dichotomous prediction of pass or fail, assigned probabilities of success, ranging from .00 to 1.00. The following are the prediction formulas used:

Item IV (1), Picture Vocabulary: Statements[1] -- .2 for each item correct in excess of 5.

Item IV (2), Objects from Memory: No prediction.

Item IV (3), Opposite Analogies: Polars -- .3 for each correct.

Item IV (4), Picture Identification: Objects -- .2 for each object for which at least one "why" question was correctly answered.

Item IV (5), Discrimination of Forms: Shapes -- .2 for each correct.

Item IV (6), Comprehension II: Instrumental Acts -- .2 for each correct.

Item IV-6 (1), Aesthetic Comparisons: No prediction.

Item IV-6 (2), Opposite Analogies I: Polars -- .2 for each correct.

Item IV-6 (3), Picture Similarities and Differences: Same and Different -- 1.0 if both parts of any item correct; .6 if at least one same and one different correct, but not on same item; .3 if at least one correct; .0 if none correct.

---

[1]This and corresponding terms in subsequent items refer to the section of the Achievement Test to which the prediction formula applies. Thus, to estimate the probability that a given subject will pass S-B Item IV (1), find out how many items he got right on the "Statements" section of the Achievement Test and give him .2 point for every one right over five. There are 10 items on the "Statements" section.

If a subject got 7 right, that would be two more than five. Counting .2 point for each of these gives an estimated probability of .4 for passing S-B Item IV (1).

Item IV-6 (4), Materials: Lower of the probabilities obtained by (a) Materials -- .7 for items 2 and 5 correct, .1 for each additional; .3 for 2 or 5 correct, .1 for each additional; .0 for neither 2 nor 5 correct; (b) Objects: .5 if correct material given for house; .2 for each additional object for which material correct; .0 if incorrect material for house.

Item IV-6 (5), Three Commissions: Function words -- .3 for each.

Item IV-6 (6), Comprehension III: Instrumental Acts, items 6-10 -- .3 for each correct.

Item V (1), Picture Completion: Objects: .2 for each object for which at least two correct parts mentioned.

Item V (2), Folding Triangle: No prediction.

Item V (3), Definitions: Categories -- .4 for each correct.

Item V (4), Copying Square: No prediction.

Item V (5), Picture Similarities and Differences: Same and Different -- .5 for each item with both parts correct.

Item V (6), Patience: rectangles: No prediction.

Item VI (1), Vocabulary: Categories -- .3 for each correct.

Item VI (2), Differences: Same and Different -- .2 for each correct difference.

Item VI (3), Mutilated Pictures: Objects -- .8 minus .4 for each Absurd Question missed.

Item VI (4), Number Concepts: Counting -- .3 for each correct.

Item VI (5), Opposite Analogies II: Polars -- .1 for each correct.

Item VI (6), Maze: No prediction.

Item VII (1), Picture Absurdities: Objects -- .1 for each Absurd Question correct plus .1 for each Object with at least one correct response for each part.

Item VII (2), Similarities: Same and Different -- .2 for each correct similarity.

Item VII (3), Copying Diamond: No prediction.

Item VII (4), Comprehension IV: Instrumental Acts, items, 6-10 -- .2 for each correct.

Item VII (5), Opposite Analogies III: Polars and Categories -- .1 for each correct.

Item VII (6), Five Digits: Memory for Unrelated Words and items 3 and 4 -- .5 for each correct.

It will be noted that no predictions were made for 7 of the items. These are performance items for which no clearly relevant achievement data were available. The construction of these prediction formulas was necessarily carried out intuitively. The relevance of the Achievement Items to the S-B Items is usually obvious, but the particular weights assigned to Achievement Items reflect complex and possibly idiosyncratic judgments. Once the formulas were set down, however, the making of individual predictions from them was a perfectly objective procedure. Accordingly, for each of the 20 subjects, probabilities of success were calculated for each of 23 S-B items.

If the 23 probabilities for a given subject are totaled, they give a prediction of the total number of items out of the 23 that he will get right. The mean predicted score obtained in this way was 16.08 correct, with a standard deviation of 3.44. The actual mean number correct was 13.90, with a standard deviation of 3.46. Thus, there was a mean over-prediction of 2.18 items. This discrepancy is highly significant ($t = 4.45$, d.f. $= 19$), suggesting a consistent over-prediction. Indeed, scores were over-predicted for 17 subjects and under-predicted for only 3. The product-moment correlation between predicted and obtained total scores for the 20 subjects was .80.

If the 20 probabilities for a given item are totaled, they give a prediction of the number of subjects who will pass the item. The predicted and actual numbers passing each item are shown in Table 8. Success of prediction ranged from perfect on item IV-6 (5), where all but one person was assigned a probability of 1.00 of passing the item and all but that one person did, to disastrous on item VI (1), where no one passed although the average assigned probability of passing was .92. In spite of the overall tendency to over-predict success, the formulas actually under-predicted the number passing on as many items as they over-predicted it, the largest errors, however, being ones of over-prediction.

It may be profitable to examine some of the largest errors for what light they may shed on the substance of the study. The failure of any children to pass the level VI vocabulary item, when almost all had been predicted to do so, was the most puzzling

Table 8

Results of Prediction of Success on Stanford-Binet Items from Performance on Achievement Test Items

Item	Predicted No. Passing	Actual No. Passing	Error of Prediction	Mean Probability Score: Subjects Passing Item	Mean Probability Score: Subjects Failing Item
IV (1) Pict. vocab.	19.0	18	1.0	.94	1.00
(3) Opp. anal. I	17.3	18	- .7	.90	.45
(4) Pict. ident.	17.8	19	-1.2	.92	.40
(5) Discr. forms	19.2	20	- .8	.96	-
(6) Compr. II	19.8	18	1.8	1.00	.90
IV-6(2) Opp. anal. I	14.2	18	-3.8	.73	.50
(3) Pict. s. & d.	16.3	19	-2.7	.81	.30
(4) Materials	8.7	16	-7.3	.54	.00
(5) 3 commls.	19.0	19	.0	1.00	.00
(6) Compr. III	16.3	19	-2.7	.81	.90
V (1) Pict. compl.	16.4	12	4.4	.88	.72
(3) Definitions	19.4	17	2.4	.98	.93
(5) Pict. s. & d.	11.0	16	-5.0	.66	.12
VI (1) Vocabulary	18.3	0	18.3	-	.92
(2) Differences	7.8	3	4.8	.67	.34
(3) Mut. pict.	11.6	15	-3.4	.59	.56
(4) Number Conc.	9.6	4	5.6	.90	.38
(5) Opp. anal. II	7.1	9	-1.9	.41	.31
VII (1) Pict. absurd. I	10.0	2	8.0	.75	.47
(2) Similarities	11.4	6	5.4	.83	.53
(4) Compr. IV	13.4	1	12.3	1.00	.65
(5) Opp. anal. II	14.9	5	9.9	.92	.69
(6) 5 digits	0.0	2	-2.0	.00	.00

result. Note that 17 children passed the Definitions item at level V, which is of the same type. This would suggest that the difficulty was lack of specific vocabulary rather than inability to give definitions. The Achievement Test provided no systematic inventory of vocabulary. The section on which predictions were based was Categories, in which the child is required to provide the class label, given a series of instances of the class. It was reasoned that this was a necessary, though certainly not sufficient constituent of S-B vocabulary test performance. But it would also appear that all children were able to perform this task to a certain degree, since every child got at least two of the five Category items correct. Be that as it may, it is plain that the program was not successful in building general vocabulary up to the point of the other components.

Large over-predictions are also found for the number of children passing absurdities, comprehension, and opposite analogies items at level VII. Similar types of items are passed with under-predicted frequencies at lower age levels, however, suggesting that the difficulty with prediction at level VII is that the subtlety and complexity of the items exceeds that of the related kinds of material dealt with in the program, so that mastery of these easier materials is no assurance of success.

In general, it may be said that prediction for items was not nearly so accurate as prediction for individuals. The product-moment correlation between predicted and actual numbers passing each item was .58, compared to .80 for the correlation between predicted and actual scores for subjects. The total error of prediction is, of course, the same in either case; but in the case of persons the variance of predicted and actual scores was virtually identical whereas in the case of items the variance of predicted item totals was only half that of the variance of actual item totals, indicating that the prediction formulas were more sensitive to individual differences in ability than to differences between items.

As a final test on the efficiency of prediction, the correlation was computed between total number of items correct out of the 23 S-B items under consideration and the total number of correct responses out of the 192 responses recorded for the Achievement Test. The obtained correlation was .79, almost exactly the same as that obtained from the formula-derived estimates. Thus, as far as predicting an individual's overall performance on the S-B is concerned, his overall undifferentiated performance on the Achievement Test is as good a predictor as the sum of the specific item-by-item predictions. This is even more tellingly demonstrated if the 192 achievement items are subdivided into the 124 which at some point or other entered into the item-by-item prediction formulas, and could thus be judged to be more relevant to Binet performance, and the 68 items which were not so used and could accordingly be judged less relevant. Scores on the "relevant" achievement items were found to correlate .77 with Binet performance and scores on the "less relevant" items were found to correlate .76.

CONCLUSIONS

If the present study has accomplished nothing else, it should at least help to silence those inevitable critics who sneer "Teaching for the Test" every time they hear a report of substantial IQ gains. Here for once, was a program which was avowedly devoted in toto to "teaching for the test" -- not in the trivial sense of drilling children on test-like items, but in the sense of attempting to provide training in the full range of conceptual content and skills which the test was believed to draw upon. Results indicated that:

1. The "Binet curriculum" was not more successful in raising Binet IQs than an academically-oriented program that made no direct attempt to teach Binet-related material and that was in largest part devoted to the teaching of reading and arithmetic skills which could be expected to have virtually no transfer to Binet items.

2. The "Binet curriculum" was not any more successful at raising IQ scores on the test toward which it was directed than it was in raising them on another test (the WPPSI), the content of which was unknown to teachers and curriculum writers.

3. Although achievement in the "Binet curriculum" proved to be highly predictive of post-test performance on the Stanford-Binet, the relationship seemed to be between overall performance on one and overall performance on the other rather than being a matter of specific connections between items of curriculum content and test items.

These results would suggest that "teaching for the test" is not a very adequate or meaningful way of accounting for the IQ gains obtained in other studies. It does not even seem to be a good way of accounting for the gains in this one. It is possible, of course, that some other way of generating a curriculum from the Binet test or some other way of teaching it would produce more positive results. Anyone who thinks so is welcome to try.

The results are much less informative on the question of what does account for large IQ gains. Motivational and test-wiseness effects could easily account for the 6-point gain obtained in the first two months of this study, as they may well account for gains of this magnitude in all other compensatory preschool studies (Zigler, 1968; Jensen, 1969). Gains continued, however, throughout the remaining six months. In light of the negative results concerning curriculum specificity, it seems reasonable for the present to entertain the possibility that these additional gains reflect the accelerated learning of basic thinking skills. It is also possible that these basic skills are taught equally well by concentrating upon academic subjects like reading and arithmetic rather than upon Binet-related material, as suggested by results with the academically-oriented preschool program. If this is true, then an academically-oriented program would be preferable because of its more direct contribution to scholastic achievement.

SUMMARY

A curriculum was devised by working backward from Stanford-Binet items to specification of a universe of content for which the Stanford-Binet could serve as a content-valid achievement test. It was reasoned that this curriculum should correspond in effect to the hypothetical "hidden curriculum" of the middle-class home. The curriculum was tested on 20 four-year-old disadvantaged children, selected according to the same criteria as other children in the current series of investigations. The program was conducted for eight months, two hours daily, with a teacher-pupil ration of one-to-five. The Stanford-Binet was administered four times during the course of the experiment, curriculum content and procedures being modified in the light of results. The Wechsler Preschool and Primary Scale of Intelligence (WPPSI) was administered at the beginning and end as a control measure for non-specific effects on IQ. The content of this test was not known to curriculum writers and teachers and pre-test scores were not made known to them either. An Achievement Test of 192 items was administered at the end of the program, testing the amount learned in the specific areas touched on in the "Binet curriculum" -- since these areas did not correspond to Binet items but rather to the organization of the curriculum.

Total IQ gain was 13 points on the Stanford-Binet -- no better than that achieved previously with the academically-oriented program which made no effort to teach Binet-related content. Gain on the WPPSI turned out to be of the same magnitude, thus indicating that the gains were in no wise test-specific. Prediction formulas were constructed for deriving from an individual's performance on relevant sections of the Achievement Test, estimates of the probability of his passing specific Stanford-Binet items. Predictions were made for 23 items in the age-level range of IV through VII. Actual and predicted numbers correct, for the 20 subjects in the study, correlated .80. Actual and predicted item difficulties for the 23 items, however, correlated only .58. It was furthermore found that total number of items correct on the Achievement Test correlated as well with Binet performance as did the formula-derived estimate, and performance on Achievement Test items judged most relevant to Binet performance yielded no better correlation with Binet performance than those not judged relevant to it.

These results were taken as indicating that there was not a close relationship between curriculum content and intelligence test performance, leaving open the possibility that what accounted for the non-trivial part of the IQ gain, in this as well as in the other studies in this series, might have been the accelerated acquisition of certain basic thinking skills.

A Model for the Interpretation of IQ Changes[1]

Carl Bereiter

PROBLEM

Changes in IQ associated with compensatory education programs have been variously interpreted as transitory effects on the expression of genetically determined intelligence (Jensen, 1969), as the result of motivational effects on test-taking behavior (Zigler, 1968), or as indices of achievement in a broad area of conceptual learning (see "Achievement Components of Stanford-Binet Performance" in this report). Very likely a number of factors are involved, some of which may eventually be isolated experimentally. In the meantime, one way to test possible interpretations is to attempt to derive from the interpretations formal models that can be tested against actual data on IQ changes before, during, and after educational treatment. The gross general finding that IQs tend to rise during special educational treatment and to decline toward their original levels after treatment ceases could be accounted for equally well by any number of hypotheses. But perhaps not all hypotheses can account equally well for the actual observed magnitudes of gains and losses.

OBJECTIVE

The objective of this study was to develop a simple mathematical model that could account for the IQ changes observed in the preschool experiments carried out in the present project, using parameters that were interpretable in terms of plausible or current hypotheses concerning the nature of IQ changes. It was hoped that this effort might provide suggestive evidence as to the validity of such hypotheses and furthermore provide a useful conceptual framework for the examination and evaluation of these and other experimental results.

METHOD

The method was to begin with the simplest and most parsimonious model that could account for the gross pattern of results and then modify it by the introduction of additional theoretically-derived assumptions until the model gave a good fit to actual data. The

[1] In the proposal this project was titled, "Analysis of the Nature of IQ Gains."

model would then be tested against available data from other studies in order to check its generalizability.

For simplicity, it was decided to develop the model only for error-free data, so that random fluctuations and regression to the population mean would not be taken into account. In practice, this should mean that the model ought to fit group mean scores but not individual scores, on the assumption that mean scores based on representative groups of reasonable size ought to be essentially free of measurement and sampling error. Since the immediate purpose of the model would be for the interpretation of group trends, this did not seem to be a serious limitation. It should be noted, however, that a model developed on the assumption of error-free data cannot be expected to fit group means where the groups have been selected on the basis of extreme scores on IQ test itself so that regression to the mean will figure as a major factor in IQ changes. Several major studies have used groups thus selected (Weikart, 1967; Hodges, McCandless, Spicker, 1967). Although the internal validity of these studies may not be impaired by the method of subject selection, their external validity necessarily suffers and the construction of a model that would permit comparison of their results with other results would present forbidding difficulties.

<u>Two types of model</u>. One interpretation of the IQ score is as an <u>index of relative brightness</u>, that may fluctuate in response to various influences but tends to hover about a fixed value for a given individual (Wechsler, 1944, p. 24). Following this interpretation, a model of IQ changes would take account of the duration, strength, and interaction of various factors hypothesized to have positive or negative effect on intellectual efficiency or test performance. To have any strength, such a model would have to make use of independent measures of the hypothesized factors and relate these to IQ score changes.

Another interpretation of the IQ score is as an index of rate of intellectual growth (Kuhlmann, 1939), based on a comparison of the level of intellectual growth actually achieved (mental age) with the amount normally expected (chronological age). According to this interpretation, IQ changes from one occasion to another are to be explained in terms of factors that increase or decrease the amount of intellectual growth taking place over the intervening period. Models based on such an interpretation are obviously applicable only to developing children. This kind of model has been used to good effect by Bloom (1964) in accounting for developmental trends toward stability in IQ score. As applied to longitudinal data it is an intrinsically stronger model than an "index of brightness" model, since it is based on cumulative and irreversible effects rather than fluctuating variables. It would thus seem <u>a priori</u> to be the preferred type of model to use experimentally in studying IQ changes induced in young children.

Simple form of the mental age increments model. Although the Stanford-Binet IQ, like most other IQ scores, now represents a standardized score of fixed mean and standard deviation, it is more appropriate in working with a mental age increments model to treat IQs in the old Stanford-Binet manner as a ratio of mental age to chronological age (Terman and Merrill, 1937). Numerically, the difference between standardized and ratio IQs is not large enough to be of much concern in the present context. Accordingly, an IQ of 90 for a child of 4, say, is taken to indicate that the child has a mental age (MA) of .90 x 4 or 3.60 and that on the average his yearly increment in MA has been .90 year. Assuming stable conditions, we expect the child to continue to gain .90 year of MA for each additional year of life, up to maturity, so that his IQ will remain at 90. Now if, for a certain period of time, conditions change (for instance, through his reciving special compensatory education) so that during that period he gains more than the usual .90 year of MA per year, his IQ will rise above 90. If conditions then revert to what was "normal" for the child, he might then, according to the most parsimonious assumptions, be expected to go back to his previous rate of intellectual growth, .90 year of MA per year. His IQ would then decline toward its original level of 90, although in principle it would never quite get back to it, because the extra-sized MA increments gained during special treatment would not be lost; they would merely count for progressively less when added in with all the other "normal" increments.

Table 1 presents hypothetical data illustrating the simple model of IQ changes described above, for a child with an IQ of 90 who, beginning at the age of 4, is given two years of compensatory education. "Compensatory" is here taken literally as education which alters conditions sufficiently to produce mental growth at the normal rate of one year of MA per year. This produces a rise in IQ to 93.3. Following cessation of special treatment, the yearly MA increment reduces to .9 and IQ declines, but very gradually, so that by the age of 15 it is still at 91.3.

Table 1

Hypothetical Data Illustrating IQ Changes Following a Simple MA Increments Model

CA	MA	IQ	ΔMA	
3	2.7	90.0	.9	
4	3.6	92.0	1.0	Period of Compensatory Treatment
5	4.6	92.0	1.0	
6	5.6	93.3	.9	
7	6.5	92.9	.9	
8	7.4	92.5	.9	
9	8.3	92.2	.9	
10	9.2	92.0	.9	
15	13.7	91.3	.9	

This simple model accounts for the gross pattern of rise and fall in IQ scores during and after special educational intervention. It also predicts that gains will be greater the younger the age of intervention, the lower the initial IQ, the more "enriched" the treatment, and the longer treatment is continued. However, the model clearly does not fit the actual magnitude of change. In the hypothetical example the gains are too small and the rate of decline is far too gradual to fit actual data. Indeed, an examination of mental age increments in the longitudinal data obtained in the present program of research immediately points up a fundamental failing of the simple incremental model. According to the simple model, the yearly mental age increment would never fall below what it had been before treatment, unless the children were to encounter conditions less favorable than those they would normally be expected to encounter. But in the present study, children who prior to treatment had been gaining on the average .96 year of MA per year were found, during the year immediately following cessation of special treatment, to gain considerably less than this -- .48 year in one experimental group and .78 in the other. Clearly, some additional assumptions are needed to account for such discrepancies.

Elaborated form of the mental age increments model. At a minimum, it would appear that a rational model of MA increments would have to include assumptions or estimates with respect to each of the following:

1. Expected IQ (IQ^E). For a given group of children in a stable social condition, it ought to be possible to arrive at an estimated IQ score that is normal for them, that tends to be maintained in the absence of notable changes for better or worse, and toward which individual IQ scores tend to regress.

2. Environmental effectiveness (E). This refers to the combined effectiveness of environmental influences working on the children to promote mental age gain. It refers to home as well as school influences, and refers only to effectiveness with respect to MA gain, not to other aspects of development. Moreover, it refers to effectiveness for the children in question and not to effectiveness in any universal sense. If we set $E = 1$ for those environmental conditions which maintain MA growth at the level indicated by IQ^E, then environmental conditions with $E>1$ would produce rising IQs, while conditions with $E<1$ would produce the opposite effects. Judging from data in the Coleman et al report (1966), for a variety of different student subpopulations in a variety of areas in America, $E = 1$. That is, there is little indication in any of the subsamples of upward or downward trends in standard score on mental tests. This does not necessarily mean, however, that all environments are effectively equal. It could merely be that the environment, including the school environment, is in all cases attuned to maintaining things as they are with

respect to intelligence, so that an environment of E = 1 for the children who are normally in it might have E greater or less than 1 for other children if they were transferred into it. Studies of school integration have in fact provided evidence that this is the case (U.S. Commission on Human Rights, 1967; Nichols, 1968).

3. Effect of IQ change (B). If, in response to environmental effects, IQ is increased or decreased from its expected level, IQ^E, it may be that this alteration in IQ will itself have some effect on future MA increments. Thus, if children of $IQ^E = 90$ undergo a 10 point increase in IQ, it might be expected that some or all of this increase would be reflected in increased ability to learn from the environment, increased potential for further growth, or what have you, so that in the succeeding year their expected gain in MA would be not .90 year but something more in the direction of the 1.00 year that would be predicted from their obtained IQ. The coefficient, B, may be regarded as a weighting factor to be applied to deviations from IQ^E, indicating their effect on subsequent MA increments. B = 0 would imply that IQ changes reflect only transitory effects on amount learned (Jensen, 1969) and have no effect on how much will be learned in the future, whereas B = 1 would imply full validity to IQ gains so that if IQs were raised to 100 children would be expected to function like average children regardless of their former or expected IQ levels. B, could of course, take values between zero and 1 and, theoretically, even greater than 1; thus it could be interpreted loosely as representing what proportion of observed IQ change "really counts."

4. Credit drawn against future MA increments (C). If, during one year, a child gains more than his expected increment in MA, this could be taken to mean that he has mastered certain tasks, learned certain things, that he would not normally have learned until a later year. Having already gotten credit for them, he will not get any further credit for them in later years, and so it is possible that his MA increment in a later year will be less than it would otherwise have been. To take a concrete example, the child who at age four has been involved in a lot of work on counting may at age five pass the "Number Concepts" item on the Stanford-Binet and receive his two-months credit in MA for it, thus receiving a boost in IQ over his less-favored peers. But at age 7, when his peers are able to pass the item for the first time, the child in question would have to pass some additional item which his peers cannot in order to receive the same MA increment as they. There is nothing in his early mastery of counting that enables him to pass some other item at age 7, and so he has a lesser MA increment and loses the IQ advantage over his peers. The coefficient, C, may be taken as an index of the general extent to which acceleration or deceleration of learning has this self-correcting tendency. Its magnitude could depend on characteristics of the children, of the environment, or of the test. That is, with respect to the children there could be built-in restrictions

on what can be learned when, with respect to the environment there might be restrictions on what the children have an opportunity to learn when, and with respect to the test the relations among items at different levels of difficulty may influence the extent to which early mastery of items at one level is or is not conducive to early mastery of items at a higher level.

5. Initial effect (P). Jensen (1969), Zigler (1968) and Bereiter and Engelmann (1966) have all remarked on the appearance of a 6 to 8 point IQ gain for disadvantaged children after such brief exposure to preschool education that it could not possibly represent mental growth in any usual sense. In the study of "Achievement Components of Stanford-Binet Performance," included in this report, a 6-point gain was observed after two months of schooling. Without specifying the nature of this initial effect, we may nevertheless represent it in the model as a special increment occurring with the onset of schooling and independent of other factors.

The following is an algebraic model for predicting MA increments derived from the lines of reasoning set forth in the five points above:

$$(1)\quad \Delta_{MA_t} = E_t \left[\frac{IQ^E + B\ (IQ_{t-1} - IQ^E)}{100} \right] - C\ (MA_{t-1} - \frac{IQ^E}{100} \cdot CA_{t-1}) + P$$

For simplicity it is assumed that the elapsed time between t-1 and t (the intervals between which the MA increment occurs) is one year. MA and CA are expressed in years, not months. Adjustment for intervals other than one year involves merely multiplying the first two terms on the right by the actual time interval. Environmental effectiveness (E) is taken to act as a multiplier on growth expected from IQ score; B acts as a multiplier on the difference between obtained IQ on the previous testing and expected IQ; C acts as a multiplier on the difference between obtained MA on the previous occasion and the MA that would be expected on the basis of CA and expected IQ. P is an absolute increment in MA, presumably with a value of zero except for the first period of compensatory education. This model reduces to the simple model used to generate the data in Table 1 by setting B, C, and P all at zero and IQ^E at 100, thus rendering variations in MA increment entirely dependent on variations in E.

Applying this model to group mean scores requires the assumption that mean IQ is proportional to mean MA divided by mean CA. This assumption is generally invalid, of course, but it is valid in the case where all individuals have the same CA. Since this is approximately the case with the groups of subjects to be considered, the CA range being less than one year, the assumption is made for computational convenience.

FINDINGS

The main empirical concern of this study was with fitting the model to Stanford-Binet IQ scores of the three preschool groups on which first-grade follow-up data are available. (See Volume II of this report, "A Follow-Up of Three of the Five Preschool Interventions: Evaluations Over Three Years.") Since the model calls for the estimation of 5 parameters and there are only 4 data points to fit for each group, it would be a trivial exercise to fit the model to the data without the addition of some independent constraints. Since the groups were presumably randomly equivalent, it appears reasonable to require that IQ^E, B, C, and P be the same for all groups, on the assumption that they reflect subject and situational characteristics that are standard throughout. On the same assumption, it may be required that IQ^E, B, and C remain constant over occasions. Finally, we may derive an independent estimate of P, for the occasion when it is not zero, from the study, "Achievement Components of Stanford-Binet Performance," which, as noted previously, was in agreement with a number of other studies in finding an initial IQ gain of about 6 points early in treatment. Translated into MA increments for a CA of 5 and an IQ of 90, this gives $P = .3$. Given these requirements, the only factor that can vary from group to group is E, and that only for the periods during which groups are receiving special treatment (otherwise $E = 1.0$, following evidence presented in the preceding section).

Taking the initial IQ scores as given, there are 3 additional scores for each group to be estimated, or 9 in all. If IQ^E is either estimated from the initial scores or independently of the data, that leaves B, C, and five different values of E_t free for adjustment in fitting the model to the data. Thus there are just two fewer coefficients than there are data points to be fitted. What this amounts to in practice is that there is no problem in fitting the model to the treatment-produced IQ gains for all three groups and to the post-treatment losses for one of the groups, but it remains to be seen whether any coefficients that do this will fit the post-treatment losses of the other two groups as well.

As a first attempt at fitting the model to data, IQ^E was set at 96, the approximate pre-test mean of the combined groups. Trial values of B and C were set at .2 and .4 respectively. Values for E_t were set at F except during periods of experimental treatment. Values of E_t for experimental periods were determined analytically by solving equation (1) for them. This fitting led to over-prediction of the drop-off in IQ following termination of treatment, particularly for the Traditional group.

Alternatively, IQ^E was set at 100 and, after some trial and error, B and C were both set at .4. Results of this fitting are shown in Table 2. The maximum error of estimate is .10 year of MA and a corresponding 1.46 points of IQ, for the final test scores

Table 2

Actual and Predicted Mean MAs and IQs
for Three Preschool Groups

+	CA	MA	$IQ_{(R)}$	E	$\Delta\widehat{MA}$	$\widehat{MA}$	$\widehat{IQ}$	$\widehat{IQ}-IQ_{(R)}$
				Amelioration Group (N=24)				
0	4.34	4.22	97.24					
1	5.00	5.47	109.40	1.41	1.25	5.47	109.40	.00
2	6.02	6.49	107.81	1.15	1.02	6.49	107.81	.00
3	6.93	7.27	104.91	1.00	.77	7.26	104.76	-.15
				Direct Verbal Group (N=10)				
0	4.26	4.16	97.65					
1	4.88	5.36	109.84	1.42	1.20	5.36	109.84	.00
2	5.92	7.00	118.24	1.70	1.64	7.00	118.24	.00
3	6.83	7.48	109.52	1.00	.58	7.58	110.98	1.46
				Traditional Group (N=25)				
0	4.37	4.17	95.42					
1	5.03	5.16	102.58	1.00	1.00	5.17	102.78	.20
2	6.05	6.06	100.17	1.00	.97	6.14	101.49	1.32
3	6.95	7.02	101.01	1.00	.87	7.01	100.87	-.14

of the Direct Verbal group. It should be noted that the IQs in Table 2 are ratio IQs and thus do not agree precisely with the standard-score IQs cited in other sections of the report.

For validation, the model was also applied to several other sets of data. Data from the original Direct Verbal pilot group, which spanned four years instead of three spanned by the groups referred to above, could also be fitted accurately by the model, retaining the same values of B and C and with Effectiveness values of 1.3 and 1.4 for the first and second years of treatment. However, it was necessary to assign that group an Expected IQ of 94, instead of 100. This is perhaps not so surprising, considering that they were selected according to a criterion which included having siblings with school problems. The model could also be fitted to the data of Long (1966) by assigning an Effectiveness of 1.0 to the treatment, which is in accord with the failure to find significant treatment effects in comparison with a control group. Data supplied privately by Robert L. Spaulding from the Durham Educational Improvement Project were especially interesting because they provided twice-yearly test data, from testings in the fall and spring. A good fit could again be made to the data, using the same values of B and C, yielding high Effectiveness values for the winter periods and Effectiveness values of less than one for the summer periods when school was not in session. This, again, is reasonable, because the hypothetical Effectiveness of 1.0 is intended as an average over time in school and out, but in the case of semi-annual testing this average would not be expected to hold.

Finally, the model was used to extrapolate the trends indicated in Table 2 -- in other words, to answer the question of what may be expected to happen to the IQ scores of those groups in the future, assuming the model to be valid. According to the model, the Traditional group had already by the end of first grade returned to its expected IQ level and is therefore expected to show no further change of significance. The model predicts that the Amelioration group will decline to within a point of its expected level in three more years and the Direct Verbal group in five more years.

DISCUSSION

It would be foolhardy to claim that the model stands validated on the basis of the tests to which it has so far been subjected. However, if one accepts the plausibility of a mental age increments model, it would appear that the findings so far are enough to indicate that a simpler version of this model will not be adequate. Without some notion of an expected IQ level, there is no reason why IQs, once raised, should ever decline, considering the range of educational environments which seem capable of maintaining IQs at constant levels. Without some factor corresponding to the coefficient B in our model, representing a forward-reaching benefit from induced IQ gains, it would be necessary to posit an altogether unreasonable Effectiveness value for the second year of treatment in the Direct Verbal group (its derived value of 1.7 is already suspiciously high, if this is taken to mean that the program was 1.7 times as effective as a normal kindergarten in promoting mental age growth). Without some factor corresponding to coefficient C in our model, representing credit drawn against future MA gains, the incremental model would be incapable of accounting for the sharp post-treatment IQ losses that are commonly found, without positing some unlikely violent negative reaction produced in the children by their change from the compensatory to the normal school program.

Accordingly, taking the model as at least plausible if not necessarily accurate in the values it assigns to various coefficients, we may consider what contribution if any, it appears to make to the interpretation of IQ changes. If the Effectiveness coefficient can be interpreted literally, which it is presumed in the model that it should be, then structured programs like the two experimental programs involved in the current series of studies and the Durham EIP program are about half again as effective as a traditional enrichment program in promoting mental age growth, while traditional programs are not appreciably more effective than the normal school curriculum. IQ gains for the Traditional group, it will be noted in Table 2, are accounted for by the model by assigning the preschool treatment an Effectiveness of 1.00, the same as regular schooling. The gain is attributable to (1) the .3 year gain in MA due to factor P and (2) the fact that the children were initially scoring below their expected IQ value of 100.

The fact that an expected IQ (IQ^E) of 100 fit the data better than an IQ^E of 96, based on their pretreatment scores, is of considerable importance in interpretation of further follow-up results on the children in question. An IQ^E of 100 implies that their IQs would have risen to a mean of 100 under normal school treatment and that they will not fall back below 100. If IQs do decline below 100, of course, this will indicate a defect in the model. Should they level off at 100, one would normally be tempted to claim some credit for the preschool programs in producing a lasting improvement. According to the model, however, no such credit would be owing, since IQs would have been expected to level off at that point without special treatment.

It is more difficult to assign any interpretation to the obtained values of coefficients B and C. At a crude surmise, they would suggest that something less than half of experimentally induced IQ gain carries over into increased potential for MA growth and something less than half of experimentally increased MA increment is subsequently offset by decreased MA increments.

If the present model or some modification of it should be found to provide an adequate fit to longitudinal IQ data in various educational experiments, it would have important advantages in making comparisons of results. IQ^E would be a more significant basis for judging the comparability of groups than would pretest IQ alone. E would provide a basic index of program effectiveness, theoretically independent of pupil characteristics or time of intervention. The model applies, of course, only to the effectiveness of programs in promoting tested MA growth, and programs equally effective in this respect could differ considerably in other kinds of effectiveness.

SUMMARY

A mathematical model of IQ changes was developed and tested on data from the current series of preschool experiments as well as on data from other studies. IQ gains during treatment and losses following termination of treatment are treated as a function of expected IQ for the group, effectiveness of treatment, the extent to which induced IQ gains carry forward into increased potential for future growth, and the extent to which induced mental age increments are offset by subsequent reduced increments.

Verbal and Nonverbal Factors in Cultural Deprivation: Evidence from Children with Sensory Handicaps

Carl Bereiter

PROBLEM

This study is a follow-up to an earlier study of the same title, an unpublished report of which was included as an appendix to the project proposal. The original study proposed that deaf and blind children might constitute "natural experiments" in the relative importance of language and sensory experience for cognitive growth, since each group is severely deprived of one kind of experience but not the other. In addition to citing well-known evidence on the superior scholastic attainment of blind as compared to deaf children, the original study cited three investigations (Myklebust, 1960; Quigley & Frisina, 1963; Templin, 1950) which provided indirect evidence that the usual correlation between socio-economic status and IQ did not obtain among the deaf. This conclusion was inferred from the failure of these investigators to find mean IQ differences between residential and day school populations which differed substantially in SES of parents. If the IQ-SES correlation were severely attenuated in deaf children, it was reasoned, this would suggest that language was the principle medium through which SES-related differences in intelligence were transmitted.

More direct empirical evidence was obtained by correlating IQ with mid-parent educational level for 219 institutionalized deaf children. The obtained correlation of .23 was lower than that for hearing children reared apart from their natural parents, suggesting that the correlation could be accounted for entirely by heredity and reflected no SES influence on IQ scores. An alternate hypotheses, that the attenuated correlation might be due to a high incidence of brain damage unassociated either with heredity or SES, was rejected because the distribution of IQs in the sample fitted the normal IQ distribution almost perfectly.

Applying the same reasoning to the blind, it would be predicted that blind children, being even more dependent than normal children on language as a learning medium, should have IQ scores that are correlated to an exceptionally high degree with parent educational level. The present study was carried out to test this additional inference.

METHOD

Using data on file at two different residential schools for the blind and partially sighted, IQ scores and mid-parent educational levels were obtained for the five groups indicated in Table 1. "Discharged blind" refers to students in the Illinois residential school who had been discharged during the prededing five years. As had been the case earlier with data on the deaf, no effort was made to screen out individuals suspected of brain damage, since in almost all cases the sensory handicaps resulted from central impairment, and so to rule out the possibility of further organic impairment was a virtual impossibility. The only restriction was that for a given sample, all individuals should have IQ scores on the same test.

RESULTS

The last column in Table 1 reports the correlations between IQ of child and mid-parent educational level for the five visually handicapped groups as well as for the deaf group investigated previously. It will be observed that the correlations for the two institutionalized blind groups are almost identical to those for the institutionalized deaf group, while that for the discharged blind group is somewhat higher and those for the partially-sighted groups are lower. In fact, however, none of the correlations are statistically significantly different from the correlation obtained for the deaf (the largest difference giving a $\underline{t}$ of only 1.4).

IQs for the visually handicapped groups tend to be lower than those of the deaf group and also to have a greater standard deviation. As the column in Table 1 marked "% 70" indicates, the visually handicapped groups have exceptionally large percentages of children with IQs below 70, especially the partially-sighted groups, where over a fifth of the children have IQs below 70. These high percentages would suggest a high incidence of the kinds of organic or emotional impairments that would not be found correlated with parent intelligence or educational level.

CONCLUSION

The data on the blind clearly do not support the hypothesis that blind children's IQs will be highly correlated with parent SES. The fact that the correlations between child IQ and parent educational status come out to be about the same for blind and deaf children and in both cases substantially lower than those for normal children, casts doubt upon the deaf data and the interpretation that was earlier put upon them. The most warranted conclusion of this study would appear to be that children with sensory handicaps are not good natural experiments for the investigation of the relative importance of various experiential factors in mental growth because of the unknown likelihood of direct organic impairment of intellectual functioning.

Table 1

Comparative Data on Deaf, Blind, and Partially Sighted Children

		IQ			Mid-Parent Educ.		
Group	N	M	S.D.	% 70	M	S.D.	r
Institutionalized Deaf	219	99.0	15.5	3.7	11.6	2.5	.23
Institutionalized Blind							
Missouri	73	93.2	17.7	10.9	10.4	2.1	.24
Illinois	62	91.6	17.2	8.1	10.3	2.0	.27
Discharged Blind	36	96.4	20.6	8.3	10.7	2.5	.46
Partially Sighted							
Missouri	49	86.7	19.6	22.4	10.6	2.7	.15
Illinois	14	89.3	19.0	21.7	11.8	2.1	-.02

SUMMARY

A previous study had found a very low correlation between child IQ and parent educational level among deaf children, leading to the inference that language was the primary medium through which social-class related differences in intelligence were transmitted. The present study was conducted as a further check on this hypothesis, examining comparable correlations for visually-handicapped children. The expectation was that these children, being more than normally dependent on the language medium, would show unusually high correlations of IQ with parental education level. Correlations in five groups revealed no significant differences from the correlation obtained among the deaf. The appearance of high percentages of children with IQs below 70 suggested that organic intellectual impairment may have been responsible for the low correlations. It was concluded that for this reason, children with sensory handicaps were not suitable subjects for testing hypotheses concerning experiential factors in mental growth.

The Performance of Advantaged and Disadvantaged Preschool Children on Tests of Sound Pattern and Speech Sound Auditory Discrimination

Girvin E. Kirk[1] and Carl Bereiter

Deutsch (1964) has postulated that disadvantaged children raised in a crowded, noisy urban environment are not only deficient in their discrimination and recognition of speech sounds, but are relatively inattentive to other auditory stimuli as well. Furthermore, she suggests that this deficiency may be caused by an organic or functional defect in that part of the child's brain receiving sound signals. Raph (1965) in her review of sensory-motor research conducted with disadvantaged children concluded from her inspection of the Deutsch (1964) study that educators should consider non auditory means of instruction when teaching subject matter skills, such as reading.

It is surprising to find such acceptance given to the notion that disadvantaged children have a generalized defect in auditory discrimination when past auditory discrimination studies with young disadvantaged children have been limited to the use of auditory discrimination instruments composed chiefly of speech sound stimuli as opposed to non-speech sound pattern stimuli. Without comparative research on speech sound and non-speech sound pattern auditory discrimination, it is difficult to evaluate either the Deutsch or the Raph position or to come to any conclusion regarding the developmental pattern of disadvantaged children in acquiring specific auditory discrimination skills.

PROBLEM

Two issues are investigated in this study. The first issue is concerned with a general versus a specific auditory or discrimination defect: Do disadvantaged children have the same or more difficulty on a test with speech sound discrimination than on a test of non-speech sound pattern discrimination? The second question focuses on the pattern of speech sound auditory discrimination test performance: Do disadvantaged children have the same difficulty in discriminating speech sound contrasts when the speech sound change occurs in the final position of words than when it occurs in the beginning position?

[1]Research Assistant Professor, Department of Educational Psychology, University of Illinois

It was hypothesized:

1. That the advantaged group would show significantly superior performance (>.05) on the Speech Sound Discrimination Test compared to the Sound Pattern Discrimination Test. The reverse was predicted for the disadvantaged group.

2. The performance by the disadvantaged group on the Speech Sound Discrimination Test should reveal a significantly greater number of correct responses for the speech sound contrasts which occur in the beginning rather than in the final word position. No significant differences were expected for the advantaged group between the number of correct responses for speech sound contrasts occurring in the beginning and in final word positions.

PROCEDURES

Selection of the Sample

Fifteen disadvantaged Negro children were randomly selected from the class lists of two preschool control classes established for an Institute for Research on Exceptional Children (IREC) project at the University of Illinois. Fifteen advantaged white children were randomly selected from a class list of a preschool class established in another IREC research project. The mean chronological age of the advantaged and the disadvantaged group was 58 months with an age range of 54 months to 65 months.

Measuring Instruments

The Preschool Auditory Discrimination Test was constructed for this study and includes two subtests: a Sound Pattern Discrimination Test and a Speech Sound Discrimination Test (see Appendix D).

The Sound Pattern Discrimination Test was designed to measure the ability of the preschool child to perceive differences between nonverbal sound patterns where the task is to discriminate between sounds of the same basic character (e.g.) sounds of water pouring from the same or different containers. Sixty sound pair contrasts were employed in the test. Each sound pair contrast was administered individually to each subject by means of pretaped cards on the Bell and Howell Language Master Machine.

The Speech Sound Discrimination Test was designed to measure the preschool child's ability to discriminate speech sound contrasts which occur in the beginning and in the final word position (e.g., gruel tool and jewel vs. webs, weds, and wedge). Twenty of the word pair contrasts were the same while forty were different. Each word pair contrast was prerecorded on a Bell and Howell Language Master card.

Test Procedures

Both the Sound Pattern Discrimination Test and the Speech Sound Discrimination Test were administered individually to the advantaged and to the disadvantaged preschool children by the investigator who had had past clinical and research testing experience with auditory tests. The total battery required a maximum of forty minutes for administration and was divided into two equal segments. The Sound Pattern Discrimination Test was administered first, followed by the Speech Sound Discrimination Test.

The task presented to the child for each item was to identify which of three Language Master cards contained the same sound as a given model card. Early discrimination tasks were given as pre-training to insure that the child grasped the intent of the task. Cards were presented in pairs, the model card randomly preceded or followed by one of the comparison cards.

The directions for both tests were specifically devised for preschool children to avoid any requirement on the part of the child to make a vocal same or different statement in response to a test item. Instead, the child was allowed to give a "yes" or "no" test response by either saying "yes" or nodding his head to signify "yes" when he thought the sound pair contrasts were the same or to say "no" or shake his head to signify "no" when he thought the sound pair contrasts were not the same. This test procedure was devised for two reasons: to avoid the use of the concept "different" in the test directions and to give the disadvantaged child who could not or would not talk an equal opportunity to give a test response.

RESULTS AND DISCUSSION

The results of the data analysis of the fifteen advantaged and fifteen disadvantaged children are presented in the following sequence:

1. A chi square test for observed difference in the performance on the Sound Pattern Discrimination Test and the Speech Sound Discrimination Test.

2. The "t" test results on six tests of significance for various auditory discrimination tasks.

In a comparison of the advantaged with the disadvantaged preschool children on the differences on the Sound Pattern Discrimination Test and the Speech Sound Discrimination Test, the members of each group were tabulated in a 2x2 table according to whether the child did better on one or the other test. The chi square test of significance with correction for continuity was applied for each group.

When the chi square test was applied to these observed differences in number of superior sound pattern auditory discrimination or speech sound auditory discrimination performances, significant differences in the predicted direction were found in the advantaged and disadvantaged groups (p>.01); i.e., the advantaged children were found to do better on the Speech Sound Discrimination Test than on the Sound Pattern Discrimination Test and the reverse was true of the disadvantaged children. Therefore, the first hypothesis was supported.

An analysis of differences between the means of the two groups was conducted by a "t" test on the differences between the two groups on six tests of significance for six auditory discrimination tasks: (a) speech sound discrimination, (b) sound pattern discrimination, (c) beginning sound changes in the speech sound discrimination test, (d) final sound changes in the speech sound discrimination test, (e) sound pattern auditory discrimination minus speech sound discrimination performance and (f) beginning minus final speech sound changes in the Speech Sound Discrimination Test.

All of the differences between the two groups were significant (p>.05). It is noteworthy that when the "t" test was applied to the auditory discrimination tasks within the speech sound discrimination stimuli, i.e., beginning minus final speech sound changes, significantly greater differences were found (p>.001) for the disadvantaged than for the advantaged preschool children indicating a greater discrepancy for the disadvantaged children between their ability to perceive differences occurring in the beginning and in the final word position. Therefore, the second hypothesis that the differences within the speech sound discrimination tasks between the beginning speech sound changes and the final speech sound changes would be greater for the advantaged than for the disadvantaged preschool children was supported.

The disadvantaged children responded, correctly to 80 per cent of the speech sound contrasts involving beginning consonants, but only 37 per cent of those involving final consonants.

This study suggests that the auditory discrimination difficulties of disadvantaged Negro preschool children are largely confined to words and, among words, to those differing in the final consonant. It would seem more plausible and parsimonious to explain this difficulty on the basis of dialect characteristics rather than on the basis of defective intellectual functioning. This conclusion would favor remedial efforts dealing directly with language problems rather than ones involving generalized discrimination training.

SUMMARY

Fifteen disadvantaged Negro children of kindergarten age were compared with fifteen advantaged white children of the same age on their ability to discriminate speech and non-speech sounds. The test task called for children to identify which of three sound cards bore the same sound as a model card. Each choice card was paired separately with the model card and the child indicated whether it was or was not the matching one. Advantaged children were found to do better on discrimination of speech sounds than non-speech sounds whereas the reverse was true of disadvantaged children. Disadvantaged children had relatively much more difficulty discriminating final consonants than beginning consonants, as compared to advantaged children. These findings were interpreted as favoring an explanation in terms of Negro dialect characteristics, rather than in terms of a generalized auditory discrimination deficit, as Cynthia Deutsch has proposed.

Appendix A

Examples of Instruction in Part - Whole Relationships*

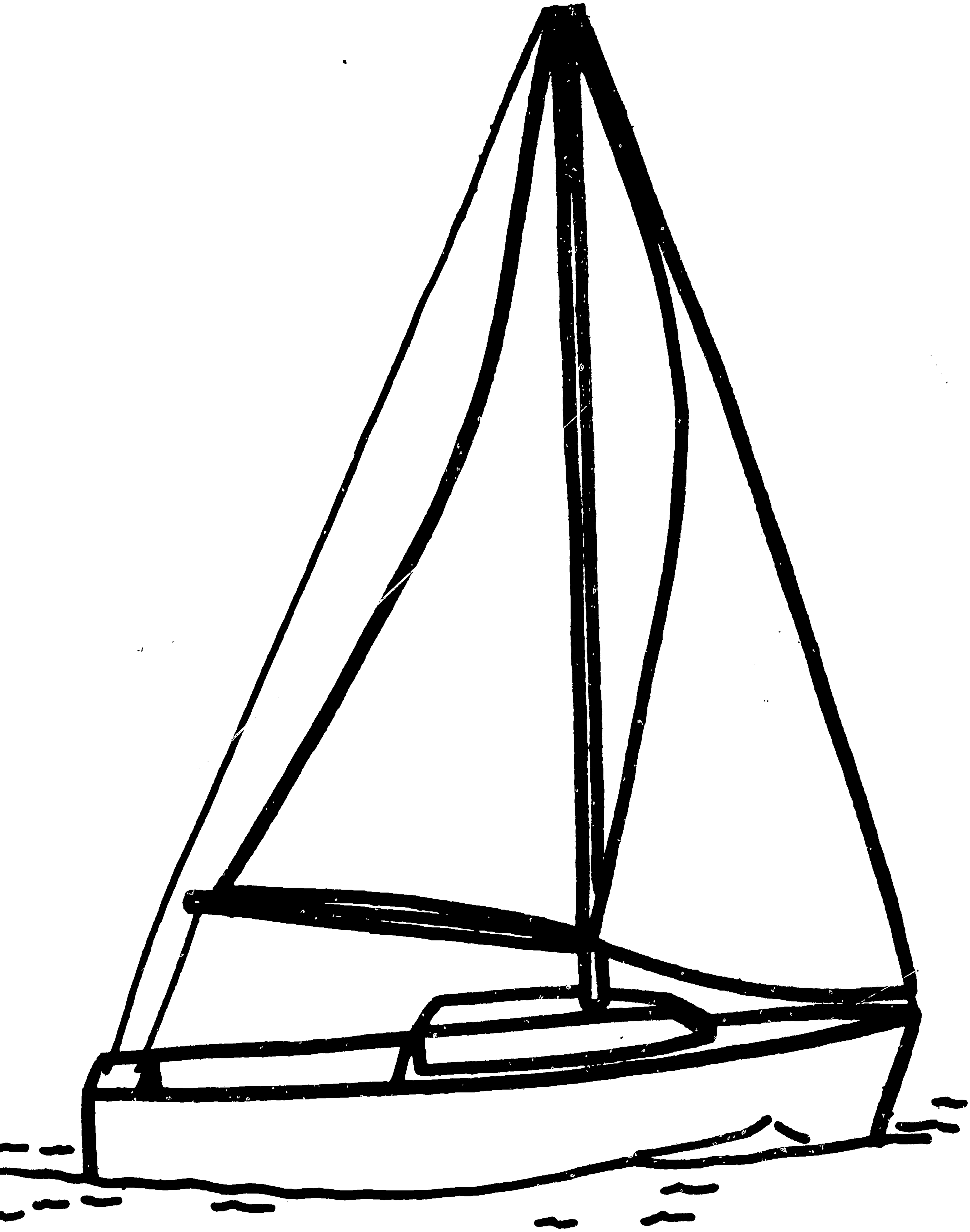
*From Engelmann, S., Osborn, J., and Lundeen, B., Learning Language: Part-Whole Relationships. Urbana, Illinois: University of Illinois Press, 1967.

object:	parts:
SAILBOAT	hull, sail, mast, rudder

questions:	answers:
1. What parts does a sailboat have?	A sailboat has a hull. A sailboat has a rudder. etc.
2. Where do we find a sailboat?	On the lake. On the river. On the ocean. On the water.
3. Why does a sailboat have a rudder? Why does a sailboat have a sail? etc.	To turn the sailboat. So the wind will make the boat move.
4. Is a sailboat a vehicle?	Yes.
5. What is a sailboat made of? What is a sail made of?	Wood. Metal. Cloth.

Absurd questions (e.g.):

6. Do we ride in sailboats on the street?

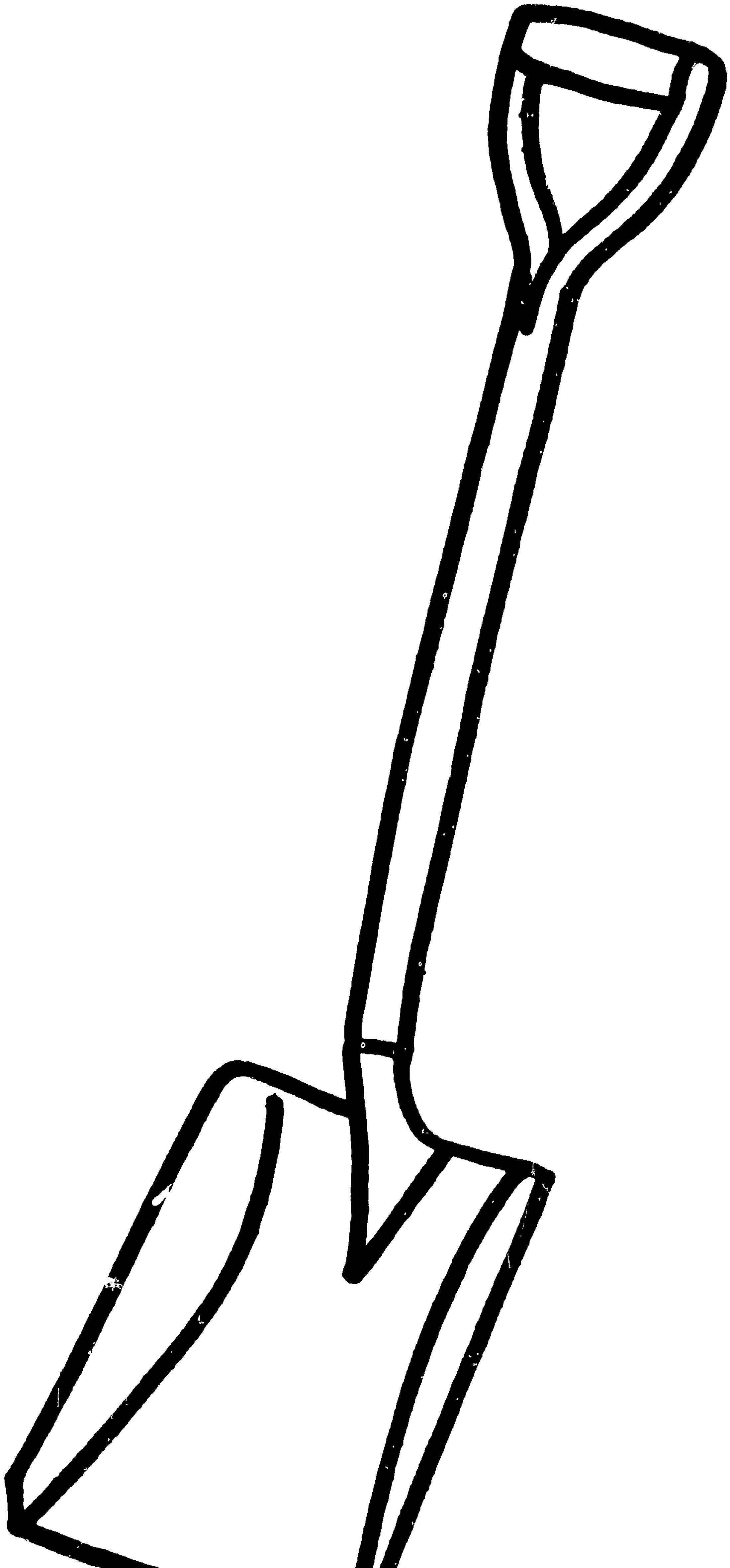


object:	parts:
SHOVEL	handle, blade or scoop

questions:	answers:
1. What parts does a shovel have?	A shovel has a handle. A shovel has a blade.
2. What do we do with shovels?	Dig.
3. Where do we find shovels?	In the yard Outside On the farm.
4. Why does a shovel have a handle? Why does a shovel have a scoop?	To hold it. To dig.
5. Is a shovel a tool?	Yes.
6. What is the scoop made of? What is the handle made of?	Metal. Wood.

Absurd questions (e.g.):

7. Do we dig into the floor with shovels?



Appendix B

Criterion Tests

1. Parts

In testing the names of the parts of an object, the tester first pointed to the picture of the object and asked "What is this?" The subject responded with the object's name. Pointing to each part of the object, the teacher then said, "An umbrella (for example) has a _________," and the subject completed the statement with the name of the part of the object being pointed to.

The child was given credit for each correct object and part name on the Parts Check List. Failure to produce the correct name was taken as an indication that the subject did not know the name.

2. Prepositions

The test of understanding of prepositions consisted of two parts. Simple objects were manipulated in both. In the first part of the test, the tester told the subject where to put a given object in relation to another, using the statement "Put the (object) on the (object)." The subject was given credit for correctly following these directions.

The second part of the test required the subject to describe the position of a given object in relation to another. The tester placed the object over (for example) another object and said "Tell me about the (object). Where is it?" The subject had to respond with "over the table" or "the car is over the table" in order to receive credit.

The prepositions tested were: in, on, next to, over, under, in front of, in back of, beside, and between.

Date__________ **Parts Check List** Group __________

Names												
1. Umbrella												
top												
handle												
2. Wagon												
body												
wheels												
handle												
3. Kite												
kite												
string												
tail												
4. Apple												
stem												
leaf												
apple												
5. Sailboat												
hull												
sail												
mast												

Date__________ **Parts Check List** **Group**__________

Names												
6. Shirt												
collar												
sleeves												
buttons												
pockets												
cuffs												
7. Shoe												
lace												
sole												
heel												
tongue												
top												
8. Coat												
collar												
zipper												
sleeves												
hood												
string												
9. Telephone												
receiver												
base												
dial												
cord												

Date ____________ **Parts Check List** Group ____________

Names												
10. Pot												
lid												
handle												
pot												
11. Pencil												
point												
eraser												
shaft												
12. Book												
cover												
pages												
13. Pop bottle												
bottle												
cap												
14. Crayon Box												
front												
back												
crayons												
15. Broom												
bristles												
handle												

Date ______________ **Parts Check List** **Group** ______________

Names												
16. Hammer												
handle												
head												
claw												
17. Toothbrush												
handle												
bristle												
18. Rake												
handle												
prongs												
19. Cat												
head												
body												
tail												
legs												
claws - paws												
ears												
whiskers												
20. Airplane												
windows												
wings												
tail												
body												
propeller												

Date ____________ Parts Check List Grade ____________

Names												
21. Chair												
back												
seat												
legs												
22. Table												
top												
legs												
23. Car												
fender												
lights												
grill												
steering wheel												
24. Tree												
leaves												
branch												
trunk												
roots												
25. Turtle												
shell												
legs												
tail												
head												
neck												

Date____________ Parts Check List Group ______________

Names												
26. Dog												
head												
legs												
paws												
tail												
body												
ears												
27. Freddie												
pants												
shirt												
shoes												
socks												
28. Sink												
faucets												
drain												
bowl												
basin												
29. Ice Cream Cone												
ice cream												
cone												

Date ____________ Parts Check List Group ______________

Names												
30. Flower												
stem												
petals												
leaves												
roots												
31. Pipe												
bowl												
stem												
mouth piece												
32. House												
window												
door												
roof												
walls												
porch												
stairs - steps												
chimney												
33. Cabinet												
drawer												
handle												
shelf												
door												
counter												

Date ______________ **Parts Check List** Group ______________

Names												
34. Stove												
oven												
burner												
knobs												
35. Refrigerator												
door												
handle												
body												
shelf												
freezer												
drawer												
36. Corn												
roots												
leaves												
stalk												
ears												
37. Bird												
wings												
tail												
head												
beak												
claws												

Date ____________ **Parts Check List** Group ______________

Names												
38. Celery												
stalk												
leaves												
39. Place - setting												
knife												
fork												
spoon												
plate												
glass												
napkin												
40. Horse												
head												
body												
legs												
hoofs												
tail												
neck												
(saddle and bridle)												

Prepositions Check List

	IN		ON		NEXT TO		OVER		UNDER		IN FRONT OF		IN BACK OF		BESIDE		BETWEEN		
Names of Children	Direction	Telling	Dir.	Tel.	Dir.	Tel.	Dir.	Tel.	Dir.	Tel.	Dir.	Tel.	Dir.	Tel.	Dir.	Tel.	Dir.	Tel.	

Appendix C

Achievement Test

Area I

1. Statements (Child is to repeat after the teacher)

Hold up an object and say, "This is an ________." Hand the object to the child and say, "Now you say it." (If necessary ask child to repeat statement - ask only once.)

Each child should have two chances to say the whole statement. In the space provided, place a plus or a minus to indicate whether the child passed or failed on that task.

Trials

Object

zebra ________
pencil ________
chalk ________
block ________
spoon ________

cup ________
eraser ________
car ________
book ________
ball ________

2. Plurals

If I have one apple we say, "This is an apple--if I have more than one apple we say, ______________."

Present two of the following objects and say, "Tell me _____." (If necessary, ask child to repeat statement--ask only once.)

Trials

Object

zebra ________
pencil ________
chalk ________
block ________
spoons ________

cup ________
eraser ________
car ________
book ________
balls ________

3. Prepositions (Place object between two erasers and say, "Where is the ________?")

1. Between

zebra ______ pencil ______ block ______

2. On (Place object on a book and say, "Where is the ____?")

chalk ______ cup ______ block ______

3. Beside (Place object side by side with the car and say, "Where is the __________?")

eraser ______ block ______ spoon ______

4. Inside (Place object in a cup and say, Where is the __________?")

pencil ______ car ______ ball ______

5. In back of (Place object in back of car. Specify: "This is the front of the car." (Point to front of car.)

chalk ______ cup ______ ball ______

4. Polars

Be sure to get the child's full attention before you start. Question may be repeated once.

1. "Listen. If you are not tall you are (short)?" __________

2. "Listen. If a dress is not new then it is (old)?" _______

3. "Listen. If a cloth is not wet then it is (dry)?" _______

4. "Listen. If a stick is not straight then it is (crooked)?" __________

5. "Listen. If a boy is not fat, he is (skinny)?" __________

5. Categories

"I'm talking about something." "What am I talking about?" (Tester names the examples.) May repeat list once.

1. (Toys) -- ball, doll, blocks, whistle, games, little wagon __________

2. (Food) -- apple, hamburger, juice, cracker, french fries __________

3. (Vehicles) -- train, car, bus, boat, bicycle, wagon, tractor __________

4. (Containers) -- sacks, boxes, cups, bags, purse, cartons __________

5. (Farm animals) -- cow, pig, duck, horse, sheep ________

6. <u>Function Words</u> (Use two blocks and three pencils.)

1. Say, "Hand me a block and a pencil." ___ ___ ___ ___.
   "What did you do?"

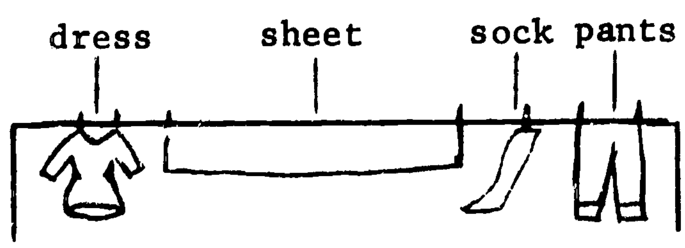
2. Say, "Hand me a block <u>or</u> a pencil." ___ ___ ___ ___.

3. Say, "Give me <u>some</u> of the pencils." ___ ___ ___ ___.

4. Say, "Give me <u>all</u> of the pencils." ___ ___ ___ ___.

5. Say, "Give me <u>both</u> of the blocks." ___ ___ ___ ___.

6. <u>Picture story on board</u>. Hanging the wash:



a) What did I hang first, second, third, fourth, last?

b) What did I hang before I hung the sock? The sheet?

c) What did I hang after I hung the sheet? The pants?

d) If I had hung the pants first, then the sock, then the sheet, then the dress. If the pants are first, what is second? _____ What is last? _____

7. <u>Same and Different</u>

Ask the child, "How are _____ and _____ the same and how are they different?" May be repeated once.

1. Horse and a cow. Same _______ Different _______
2. Red block and a green block. Same _____ Different _______
3. Tall man and short man. Same _____ Different _____
4. Wood and glass. Same _____ Different _____
5. Cup and box. Same _____ Different _____

Area II

1. Locations

   A. I see water in front of me and behind me. I see water to the right side and to the left side. I see water under me. I see a lot of water all around me. Where am I?_____

   B. I see pigs, sheep, cows, and a barn. Where am I?

   C. I see erasers, desks, chairs and a chalk board. Where am I?

   D. I see a giraffe and an elephant. Where am I?

   E. I see a sick man; men and women dressed in white. Where am I?

2. Memory for unrelated words

   Repeat each twice.

1.	boy, cup, draw	1. ________	2. ________
2.	horn, paper, fly, soft	1. ________	2. ________
3.	chain, bell, see, plant, face	1. ________	2. ________
4.	book, tree, shoe, map, dog, girl	1. ________	2. ________

Area III

1. Instrumental Acts

What do you do with your eyes? ________
ears? ________
legs? ________
nose? ________
teeth? ________

What do you do if you are cold inside the house?______________

What do you do if you want to cook a hot dog in a pan?________

What do you do if you want to find out if something is heavy?__

What do you do if you want to cross a river?_________________

What do you do if you feel sick?______________________________

2. Days of the Week

What are the days of the week?

What day comes before Thursday? What day comes after Saturday?

3. What are the months of the year?

What month comes after March? What month comes before September?

4. Part - Whole Relationship

Object:	Parts:
TABLE	top, legs

Questions:	Answers:
1. What parts does a table have?	A table has a top. ______ A table has legs. ______
2. Why do we have tables?	To eat on. ______ To study on. ______
3. Where do we use tables?	In the house. ______ In school. ______
4. Why does a table have a top? Why does a table have legs?	To put things on. ______ To hold the top. ______
5. Is a table a piece of furniture?	Yes. ______
6. What is a table made of?	Wood. ______ Plastic. ______ Metal. ______
Absurd questions (e.g.): 7. Do we sleep on tables?	______

Object:	Parts:
LAMP	Shade, light bulb, stand, cord, switch, plug, base

Questions	Answers:
1. What parts does a lamp have?	A lamp has a shade. A lamp has a light bulb, etc.
2. Why do we have lamps?	To give us light.
3. Where do we find lamps?	In a room. In a house. In a store.
4. Why does a lamp have a switch? Why does a lamp have a light bulb?	To turn the light on. To give us light.
5. What is a lamp made of?	Glass. Metal. Wood. Paper. Plastic.

Object:	Parts:
CORN	Roots, stalk, leaves, ears

Questions:	Answers:
1. What parts does corn have?	Corn has roots. Corn has stalks, etc.
2. What do we do with corn?	We eat corn.
3. Where does corn grow?	In the ground.
4. Do we eat the roots? Do we eat the ears?	No. Yes.
5. Is corn food?	Yes.

Absurd Questions (e.g.):

6. Does corn have eyes?

Object:	Parts:
SAW	Handle, blade, teeth

Questions:	Answers:
1. What parts does a saw have?	A saw has a handle. A saw has a blade, etc.
2. What do we do with saws?	Cut wood.
3. Where do we use saws?	In the house. Outside. Wherever they are building things or fixing things.
4. Why does a saw have a handle?	To hold it.
Why does a saw have teeth?	To cut better.
5. Is a saw a tool?	Yes.
6. What is a saw made of?	Wood. Metal.

Absurd questions (e.g.):

7. Do we cut paper with a saw?

Object:	Parts:
HOUSE	Window, door, chimney, porch , walls, roof, stairs (steps), shutter, garage, railing

Questions:	Answers:
1. What parts does a house have?	A house has a window. A house has a door, etc.
2. Why do we have houses?	To live in.
3. Where do we find houses?	On the streets. In the city. On farms.
4. Why does a house have a window?	To let the light in.
Why does a house have a door? etc.	So you can enter.
5. Is a house a building?	Yes.

6. What are houses made of? Brick.
Wood.
Stone.
Shingles.
Stucco.
Glass.

Absurd questions (e.g.):

7. Do houses take you places?

Area IV

1. Number

Recognition: 3____ 5____ 7____ 2____ 10____ 20____

Counting: Place a group of objects before the child and say, "Give me:"

2 __________ 7 __________ 10 __________

2. Sequences
Say to the child, "Do this!" Ask for demonstration not verbal response.

A. Pat your knee, clap your hand. 1 ___ 2 ___

B. Pat your knee, clap your hand, tap your head. 1 ___ 2 ___

C. Clap your hand, stamp your feet, pat your knee, tap your head. 1 ___ 2 ___

3. Shape
(The small colored pieces)
Circle___ square___ triangle___ rectangle___ trapezoid___ oval___

4. Colors
Use red, green, blue, yellow, black, white construction paper

Red___ green___ blue___ orange___ yellow___ black___ white___

5. Materials
What is this made of?

Plastic___ wood___ metal___ glass___ leather___

Appendix D

Sound Pattern and Speech Sound
Discrimination Tests

PRESCHOOL AUDITORY DISCRIMINATION TEST

Girvin E. Kirk and Carl Bereiter

Name of Child ______________________ Birthdate __________

School ______________________ Age (CA) __________

Grade ______________________ Test Date __________

Examiner ______________________ Sex (circle) M F

	Total Correct	Total Incorrect
Test 1. Sound Pattern Discrimination. . .	______	______
Test 2. Speech Sound Discrimination. . . .	______	______

PRESCHOOL AUDITORY DISCRIMINATION TEST

This auditory discrimination test was designed to measure the ability of preschool children to detect sound patterns that are the same and sound patterns that are different. Designed for testing children who are four and five years of age, the tests are contained in a two page booklet comprising two separate tests: Sound Pattern Discrimination and Speech Sound Discrimination. Each test consists of sixty sound-pair contrast items. The child is asked to judge whether each sound pair contrast in each test item is the same or different.

Description of the Subtests

Test 1. Sound Pattern Discrimination. This is a test of the preschool child's ability to discriminate between pairs of nonverbal sound discrimination stimuli. In each item, the child is to determine whether the pair of speech sound stimuli are the same or different.

Test 2. Speech Sound Discrimination. This is a test of the preschool child's ability to discriminate between pairs of speech sound discrimination stimuli. In each item, the child is to determine whether the pair of speech sound stimuli are the same or different.

General Administration Directions

1. The examiner should carefully examine a copy of each test, observe the directions and check the score sheet before administering each test.

2. Each child should be tested individually by the examiner in a quiet room taking care that the testing room is not situated next to a noisy hallway or on a street with passing cars.

3. Each test should be administered with the provided standard directions.

4. When administering the test, the examiner should seat the child in front of a small table, sit behind the table, face the child, and hold the test cards (on the same level of the child's eyes) above the testing machine on the table.

5. The directions to be read to the child are printed in large type. The instructions for the examiner are in small type. The directions are administered by the examiner exactly as they are printed.

6. The examiner may administer both tests in one test session or administer each test in two or more test sessions. This test procedure is employed to obtain information on the preschool child's ability to detect same and different sound patterns and not to measure his speed of response. For this reason the examiner gives ample opportunity for each child to respond to each test item.

7. The examiner should administer the Sound Pattern Discrimination Test before administering the Speech Sound Discrimination Test.

Directions for Test Administration

Test 1. Sound Pattern Discrimination. The examiner selects the Speech Pattern Discrimination Test cards. He holds up the two cards for the first test item sound pair contrast in front of the child and says: LISTEN. The examiner plays the first card with his left hand. Then he holds up the played card and says: WE WANT TO FIND THIS ONE. Holding up the card in his right hand the examiner says: IS THIS THE ONE? The examiner plays the left hand card first, followed by the right hand card and says: IS THIS THE SAME (holding up the right hand card) AS THIS ONE (holding up the left hand card)? The examiner notes the child's response and places a "Y" for yes or "N" for no on the score sheet in the appropriate item test box. Following the test the examiner notes the number of correct test item responses by noting the number of "Y" responses in the shaded boxes and the number of "N" responses in the non-shaded boxes. The examiner places the total number of correct items on the bottom of the score sheet.

Test 2. Speech Sound Discrimination. After selecting the Speech Sound Discrimination Test cards, the examiner follows the directions given in Test 1.

TEST 1. SOUND PATTERN DISCRIMINATION　　　SCORE SHEET

SOUND PAIR CONTRAST

	1.	2.	3.
1. Fog-Horn	■		
2. Bird-Sing			■
3. Truck-Start			■
4. Man-Shout	■		
5. Dog-Bark		■	
6. Hammer-Nail	■		
7. Door-Slam		■	
8. Man-Snore			■
9. Fire-Siren		■	
10. Man-Laugh	■		
11. Knock-Door		■	
12. Man-Hum	■		
13. Train-Pass			■
14. Man-Cough	■		
15. Train-Start		■	
16. Scissors-Cut			■
17. Man-Gargle		■	
18. Fingers-Snap			■
19. Water-Bottle			■
20. Man-Clap	■		

Number Correct______

Number Incorrect______

TEST 2. SPEECH SOUND DISCRIMINATION

SCORE SHEET

SOUND PAIR CONTRAST

				1.	2.	3.
1.	Sing	King	Ring			■
2.	Thin	Twin	Tin		■	
3.	Bun	Gun	Done		■	
4.	Webs	Weds	Wedge		■	
5.	Ricks	Rich	Ritz	■		
6.	Last	Lass	Laugh			■
7.	Sift	Rift	Shift	■		
8.	Barn	Barf	Bark	■		
9.	Lamb	Lamp	Land		■	
10.	Gruel	Tool	Jewel			■
11.	Cud	Cub	Cup		■	
12.	Tab	Tat	Tad			■
13.	Sue	Shoe	Chew			■
14.	Shock	Shod	Shop	■		
15.	Hear	Fear	Dear	■		
16.	Fate	Face	Faith	■		
17.	Folk	Choke	Joke		■	
18.	Back	Bag	Bad		■	
19.	Need	Mead	Read	■		
20.	Vat	That	Fat			■

Number Correct __________

Number Incorrect __________

REFERENCES

Bereiter, C., Case, R., & Anderson, V. Steps toward full intellectual functioning. J. Res. and Develop. in Educ., 1968, 1, 70-79.

Bereiter, C. & Engelmann, S. Teaching disadvantaged children in the preschool. Engelwood Cliffs, N. J.: Prentice-Hall Inc., 1966.

Bereiter, C. & Engelmann, S. An academically-oriented preschool for disadvantaged children: Results from the initial experimental group. In D. W. Brison and J. Hill (Eds.), Psychology and early childhood education. Toronto: Institute for Studies in Education, 1968. Pp. 17-36.

Bereiter, H. B. Academic preschool follow-up: Sixteen children through first grade. Unpublished report, Colonel Wolfe Preschool, University of Illinois, 1968.

Bloom, B. S. Stability and change in human characteristics. New York: Wiley, 1964.

Champaign Community Unit School District No. 4, Champaign, Illinois. Dual Kindergarten Operational Grant Proposal, Title III, Public Law 89-10. January 1, 1968.

Coleman, J. S. Equality of educational opportunity. Washington, D.C.: Government Printing Office, 1966.

Deutsch, C. P. Auditory discrimination and learning: social factors. Merrill-Palmer Quarterly, 1964, 10, 279-296.

Dolch, E. W. A basic sight vocabulary. Elementary School Journal, 1936, 36, 456-60.

Engelmann, S. Concepts and problem solving. In K. E. Berry & B. Bateman (Eds.), Conceptual learning (Dimensions in Early Learning Series). San Rafael, California: Dimensions Publishing Co., 1969.

Hodges, W. L., McCandless, B. R., & Spicker, H. H. The development and evaluation of a diagnostically based curriculum for preschool psycho-socially deprived children. VSOE Bureau of Research Project No. 5-0350, Final Report, December, 1967 (mimeo).

Jensen, A. R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.

Kohlberg, L. Early education: A cognitive-developmental view. Child Development, 1968, 39.

Kuhlmann, F. Tests of mental development. Educational Test Bureau, 1939.

Long, E. R., Jr. The effect of programmed instruction in special skills during the preschool period on later ability patterns and academic achievement. VSOE Cooperative Research Project No. 1521, 1966 (mimeo).

Myklebust, H. R. The psychology of deafness. New York: Grune and Stratton, 1969.

Nichols, R. C. Review of the U.S. Commission on Human Rights' Racial isolation in the public schools. American Education Research Journal, 1968, 5, 700-707.

Quigley, S. P. & Frishina, D. R. Institutionalization and psycho-educational development of deaf children. CEC Monogr., 1963, Series A, No. 3.

Raph, J. B. Language development in socially disadvantaged children. Review of Educational Research, 1965, 35, 389-400.

Strodtbeck, F. L. The hidden curriculum of the middle class home. In C. W. Hunnicutt (Ed.), Urban education and cultural deprivation. Syracuse, N. Y.: Syracuse University Press, 1964. Pp. 15-31.

Templin, M. The development of reasoning in children with normal and defective hearing. Inst. Child Welfare Monogr. Series, No. 24. Minneapolis: University Minnesota Press, 1950.

Terman, L. M. & Merrill, M. A. Measuring intelligence. Boston: Houghton, 1937.

U.S. Commission on Human Rights. Racial isolation in the public schools. Washington, D.C.: Government Printing Office, 1967.

Wechsler, D. The measurement of adult intelligence. (3rd ed.) New York: Williams and Wilkins, 1944.

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

Weikart, D. P. Preschool programs: Preliminary findings. Journal of Special Education, 1967, 1, 163-181.

Zigler, E. & Butterfield, E. C. Motivational aspects of changes in IQ test performance of culturally deprived nursery school children. Child Development, 1968, 39, 1-14.