

ED 021 618

PS 001 066

By- Adkins, Dorothy C. And Others

PRELIMINARY EVALUATION OF A LANGUAGE CURRICULUM FOR PRESCHOOL CHILDREN. FINAL REPORT.

Hawaii Univ., Honolulu. Educational Research and Development Center.

Spons Agency- Office of Economic Opportunity, Washington, D.C.

Report No- OEO-4219

Pub Date 67

Note- 23p.

EDRS Price MF- \$0.25 HC- \$1.00

Descriptors- BASIC SKILLS, \*CURRICULUM DEVELOPMENT, CURRICULUM PLANNING, \*CURRICULUM RESEARCH, \*LANGUAGE ABILITY, LANGUAGE DEVELOPMENT, LANGUAGE ENRICHMENT, \*LANGUAGE INSTRUCTION, \*LANGUAGE PROGRAMS, LANGUAGE RESEARCH, LANGUAGE SKILLS, LANGUAGE TESTS, \*PRESCHOOL CHILDREN, PRESCHOOL PROGRAMS

Identifiers- Hawaii, Head Start

During the summer of 1967, a 1-week training program was conducted at the University of Hawaii in which seven teachers and six aides were instructed in the use of the Bereiter-Engelmann teaching strategy for language programs. The Bereiter-Engelmann approach involves concentration upon the development of language skills and facility with basic sentence usage. The teachers who participated in the summer training program were each then assigned a summer Head Start class. The 49 experimental pupils were administered an experimental language curriculum. A control group of 20 children received a more extensive but less intensive Bereiter-Engelmann language curriculum. A group of analog tasks was developed to complement the basic learning tasks of the experimental program. The analog tasks, as distinguished from the basic tasks, involved only nonverbal responses. The pupils of both conditions were administered the School Readiness Tasks as posttests. The results showed that on most tasks there was no significant difference between the performance of the experimental and control groups. (WD)

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE  
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE  
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS  
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION  
POSITION OR POLICY.

University of Hawaii

Head Start Evaluation and Research Center

Dorothy C. Adkins, Director

Preliminary Evaluation of a Language Curriculum for Preschool Children

Dorothy C. Adkins and Ian E. Reid, Co-Investigators

Marion Gross Kagan, Coordinator and Director of Teacher Training

Doris C. Crowell, Curriculum Consultant

Vidya Bhushan, Statistical Consultant

Theo M. Cade, Assistant in Research

Final Report

Contract No. OEO 4219

The research reported herein was performed pursuant to a contract with the Office of Economic Opportunity, Executive Office of the President, Washington, D. C. 20506. The opinions expressed herein are those of the authors and should not be construed as representing the opinions or policy of any agency of the United States Government.

Education Research and Development Center

David G. Ryans, Director

College of Education

University of Hawaii

Honolulu, Hawaii

ED021618

PS 001036

### Abstract

The present project provided for the training of teachers and aides for a summer Head Start program which employed a modified version of the language curriculum developed by Bereiter and Engelmann. A preliminary assessment of the curriculum was also completed in respect to the level of achievement realized by the students, using the School Readiness Tasks by Barbara Bateman as the dependent variable, in respect to absolute norms and in comparison with other Head Start pupils that had received a different adaptation of the language curriculum.

The curriculum was also examined for evidence of an internal, hierarchical structure among the specific skills that comprise the objective of the curriculum, and was further analyzed for correlations with a series of analogue tasks that were specifically developed for the purpose of assessing at least some of the same cognitive characteristics that were thought to be fostered by the curriculum.

## General Introduction

The University of Hawaii Head Start Evaluation and Research Center has continued to explore the Bereiter-Engelmann teaching strategy. During 1966-67, Center personnel developed detailed curricular materials that were based upon the Bereiter-Engelmann curriculum but intended for use under local conditions, and tried out and revised the materials in preliminary form. During the provisional trials, frequent sessions with the teachers were conducted for the purpose of gaining feedback concerning the adequacy of the materials and procedures. Many class sessions were video taped.

In a broad sense, a primary purpose of the present project was to provide continuity for further work in 1968-69 with the curriculum that had been evolving during 1966-67. In particular, if the materials were to be used at the beginning of regular Head Start programs in 1967-68, supervised training of teachers who were likely to be employed in 1967-68 and a limited tryout of the first sections of a manual seemed essential. For this reason, a special training program was conducted in the summer of 1967 for six summer Head Start teachers and six aides. Following the training, they used the beginning sections of the manual with summer classes.

Consideration of the Bereiter-Engelmann strategy led to some additional questions which could be explored in a preliminary way at the same time the new manual materials were being used with summer Head Start classes.

### The 1967 Summer Training Sessions on a Language Curriculum

#### Selection of teachers and aides.

The local Head Start Evaluation and Research Center had a number of applications from teachers for Head Start teaching positions. From this pool of applicants, seven teachers who were highly rated by their peers were selected for training during the summer project. Six aides, unskilled members of the community, were hired under the provisions of CAP organizations.

#### Content of instruction.

The major content for the summer training of teachers and aides focused upon a formal presentation of the material presented by Bereiter and Engelmann (1966), but was supplemented by other relevant articles. The following topics were covered during the training:

1. The specific objectives of the summer project.
2. The organizational problems in the administration of the summer program.

3. The language problems of the disadvantaged child.
4. The instructional language program as developed by Bereiter and Engelmann and its implementation in a normal nursery school setting.
5. The use of contingency management, as developed by Lloyd Homme (1966), and reinforcement schedules.
6. Paper work: The coding forms which had been developed to determine individual progression through the sequence of grammatical forms.
7. Implementation of the parent education program, and discussion of its aims and scope.
8. The utilization of aides in the classroom, incorporating the concept of the aide as a teaching person.

#### Training of teachers and aides.

The teachers met daily for a week and on Tuesday and Thursday the aides and teachers met together. In the discussions which followed the presentation of the program and methodology, emphasis was placed on the concept of the aide as a teaching person who would reinforce the sentence patterns introduced by the teacher. Video tapes were helpful during these sessions in providing concrete episodes to illustrate program development and child behavior.

#### Follow-up sessions.

Teachers and aides met once weekly with separate discussion leaders to discuss:

1. Problems relating to child behaviors.
2. Problems relating to the vertical progression of the children individually and in groups.
3. Questions relating to the part that the aides could play in the teaching process.
4. The use of volunteer mothers.
5. The success or failure of various types of teaching techniques and lesson planning.
6. Problems related to the use of reinforcement only in conjunction with the language curriculum.
7. Problems encountered in shifting from material to token rewards.

### Outcomes.

On the basis of teacher discussion, it was possible to incorporate many suggestions into the manual which was in the process of development during the summer and which was to be the basis for an experimental treatment in 1967-68. Methods of coding were systematized and ideas for a good teacher observation form were developed. It was decided on the basis of the summer experience that aides should not supplant the teacher in the language sessions but that the regular teacher should be responsible for this aspect of the program. On the basis of teacher and aide comments, it was decided to provide materials in the manual for two other semi-structured activity groupings to be carried on during the teaching hour.

Part of the 1966-67 research program consisted of the development of a series of T.V. tapes in order to obtain a record of the experimental procedure and content. During the summer training session for teachers and aides, the tapes were used in order to demonstrate the teaching procedure and to show deviant behavior and how it could be handled during the structural language sessions. Since the tapes proved to be useful as a training device, analysis of all the tapes was begun during the summer with a view to producing one single 45-minute track for training new teachers entering the program in the fall.

The following guidelines were proposed:

1. The purpose of the tape would be to demonstrate this instructional approach to a preschool language curriculum.
2. It would include examples of the major steps in the sequence of the curriculum from the beginning program through the advanced program.
3. Emphasis would be placed on the need to eliminate extraneous stimuli.
4. It would emphasize the use of the systematic reward schedule, only with the language curriculum.
5. It would include examples of variations in the physical set-up of the room, of teaching strategies, and of dialogue and tandem teaching.
6. Teaching segments which illustrated the need to include tasks of varying level within the individual language lesson would be noted.

It was also suggested that in order to stimulate critical thinking and discussion about the instructional approach, segments should be included to illustrate problem areas, including those related to individual versus group participation and disciplinary problems.

## Problems Related to the Evaluation of the Language Curriculum

### Background.

Bereiter (1967) has suggested that much of the failure of education has been due to the unwitting adoption of a model that leads to a statement of broad curricular objectives. The problems in assessing attainment of broad objectives, however, lack sufficient definition to permit solution. On the other hand, problems in measuring attainment of very narrowly defined objectives tend to be more specific and limited, and consequently more amenable to solution. It is argued that if global objectives are partitioned, the problems of education will be resolved within the context of the limited objectives, leading to greater educational gains.

At times such an approach has been refined to the point where the behavior elicited in the training situation is, in fact, the objective of instruction. This has been the case, for example, in some linear programming. In these instances, presumably, the behavior can be systematically reinforced in the stimulus context in which the behavior is desired, with the result that behavior in the future will be more "appropriate" to the situation.

As educational objectives become more specific, they become increasingly congruent with educational activities. The instructional activities tend to be "to the point," and it is possible to determine whether or not success has been achieved. With specific behavioral objectives, it is a relatively simple matter of observation to determine whether or not they have been realized. For objectives couched in terms of general characteristics, a more complicated problem of determining criteria of success confronts the investigator.

This general trend to specificity in curricula has carried over into compensatory education (e.g., Bereiter and Engelmann, 1966). In this case, however, more is required than a relatively simple refinement of general objectives. It is necessary first to define the objectives.

The culturally deprived are deficient in nearly every area of ability that is susceptible to objective assessment. They have lacked the type of intellectual environment that is characteristic of children from the middle and upper social echelons. It would be possible to gear a program of compensatory education to any one of these many differences, and indeed programs have been directed to many of them. But at this point it is productive to recall Gulliksen's (1950) distinction between intrinsic and extrinsic correlates of success in education. Since our primary concern is with programs that will enable the educationally disadvantaged to compete better in an academic situation, it is necessary to focus attention upon those specific characteristics or deficits that are intrinsically related to academic achievement.

One of the salient characteristics of culturally disadvantaged children is their severe deficit in language skills. Since they are also retarded in general educational development, these two characteristics are at least related. Since much of the curriculum is presented through language and since the deficit in language skills antedates formal educational activities, it seems reasonable that remedial efforts should be directed early toward overcoming the language deficiency.

The concept of language skills in the present context lacks precise definition and could cover a variety of activities such as word naming, word rhyming, or oral reading. Bereiter and Engelmann focus more upon the characteristics of sentences and emphasize their use in oral expression. They contend that in language training the child learns concepts and rules governing the manipulation of concepts; once the formal characteristics of language are acquired the child has an important tool that will foster the acquisition of new knowledge and lead to the solution of logical problems.

#### Specific problems.

The objectives of the Bereiter-Engelmann language curriculum are specifically defined in that efforts are concentrated upon development of language skills and facility with basic sentence usage. But they are at the same time general in that specific behaviors fostered in the instructional setting are not the ultimate goals, but are exercises for the development of skills that will have their important applications in situations outside of the instructional setting. The child who learns the logic of language that makes inference possible should be more intelligent, at least in a limited sense, as a result of instruction.

To the extent that a curriculum seeks to foster specific behaviors, assessment lies primarily in a structured observation of behavior in order to determine whether or not the desired behavior occurs in a specified situation. An evaluation of the utility of the curriculum ideally would involve a determination of the extent to which the behaviors could be observed in comparable children who had not had the specific training. It would be hoped, naturally, that children who had received the training would exhibit particular skills to a greater degree than children not having had the benefit of the training.

A secondary problem related to specific attainment within the present curriculum is the extent to which it has a hierarchical structure. The cognitive skills cultivated by the curriculum have not been regarded as independent. On the contrary, in many instances the acquisition of one skill has been considered prerequisite to the attainment of another. The skills have been ordered in terms of increasing complexity and difficulty as this is revealed by logical analysis, but so far little is known concerning the interdependence of the skills.

To the extent that the attainment of one skill is prerequisite to that of another, the statistical characteristic of the performance of the subjects should be that an item measuring a higher level of achievement



will be passed only if all items at a lower level have been passed. Thus passing a higher item implies passing all lower items. Given variance among the items, assuming a more or less constant amount of training per child, the means of the items should be ordered in accord with the ordering of the items dictated by the vertical structure; that is, the means should descend in magnitude. Certain expectations regarding the correlations among the items also will be taken up later.

With more general objectives, which are also applicable with the Bereiter-Engelmann curriculum, additional methods for evaluation are needed. The tendency in this case has been to rely upon indirect assessment of the desired characteristics through the use of standardized tests.

Accumulating evidence indicates that changes do occur in test performance on various tests of intelligence and cognitive functioning as a result of participation in special educational programs, but the interpretation is extremely difficult. Three well-known factors which mitigate against a straightforward interpretation of increases in test performance are the effects of regression, retesting, and maturation.

These problems to a degree can be circumvented with appropriate statistical control, although few studies reflect the necessary randomization of subjects or testing conditions. A further problem in the interpretation of apparent gains is far more resistant to appropriate control. This problem arises from the possibility of cultivating subject characteristics that encourage better test performance but contribute neither to a heightened level in the characteristic the test is intended to measure nor to better performance on the sorts of tasks that the test is intended to predict. Such characteristics are extrinsic to the validity of the test. Of many characteristics that may contribute to performance on a test, only one may correlate highly with performance on criterion tasks. While theoretically the square root of the reliability of a test sets the upper limit to its validity, customarily reliabilities greatly exceed validities. Much of the difference could indeed be due to the assessment of characteristics not intrinsically related to criterion performance.

The development of rapport during the first months of schooling, for instance, may be reflected in better rapport and increased scores on an intelligence scale. Educational experiences contributing to task persistence might very well be reflected in a testing situation. A program that elicits a greater degree of pupil activity, both motor and verbal, may result in more relevant behavior in a testing situation, without actually contributing to the relative strength of specific characteristics that the test is intended to measure. But it is not so important to catalog as to illustrate the possibility of developing some form of "test-wiseness" that is extrinsic to the validity of a test.

While gains on standardized tests may be taken as support for the claim that curriculum contributes to the development of general skills, they

certainly fall short of an adequate demonstration. Ideally, then, the assessment of general characteristics would involve specific behavioral observation in a situation so structured that the desired characteristics would be unambiguously manifested in the subject's performance.

The foregoing consideration of problems related to the type of curriculum being developed, and for which teachers and aides were to be prepared, suggested the following questions to be explored, at least in a preliminary way, using data from the summer classes:

1. Do children trained by procedures dictated by the Bereiter-Engelmann approach actually attain the characteristics ostensibly cultivated by the curriculum? In the present case, the curriculum minimally attempts to develop a general facility in the use of language. Do the children attain even a specific ability in the use of language?
2. Do the techniques employed with the present curriculum constitute an efficient means toward the attainment of the objectives of the curriculum?
3. To what extent is there experimental confirmation of a hierarchical structure in the curriculum?
4. Does the curriculum result in the development of generalizable skills? An important objective of the instruction is the development of logical skills that are presumably fostered by a mastery of the language. Can situations be developed that would provide for the demonstration of general skills?

The following sections will be addressed to these specific areas of concern.

### Procedures

Subjects. The teachers who were trained under the provisions of the present study were assigned to six summer Head Start classes in four geographic areas on Oahu, and the pupils within these classes were used as the subjects for the present study. For the purpose of providing some meaningful comparisons, these subjects were divided into two groups on the basis of their geographic location, group 1 comprising those pupils from Pearl City, Waimanalo, Palolo, and Honolulu (N=28), and group 2 consisting of those from the Kalihi area (N=21). All were taught the experimental language curriculum that has been under development at the University of Hawaii Head Start Evaluation and Research Center.

It was originally intended that additional pupils from among the remaining Head Start classes would be selected to provide an informal control condition for the purpose of assessing the utility of the curriculum under study. It was subsequently learned, however, that the state OEO office had provided all of the summer Head Start teachers with a curric-

ulum guide that, although not based exclusively upon the Bereiter-Engelmann curriculum, was substantially influenced by this approach to language development. The main characteristic that differentiated the curriculum provided for use with the remaining classes was its scope; the pupils were given training in virtually all areas of the Bereiter-Engelmann curriculum that were subsequently tested, although the intensity of the training on that portion covered in the experimental classes was not commensurate with that provided by the experimental curriculum.

Because of general differences in ability that occur among children within the various geographic areas, only children from one area, Kalihi, where there was a sufficiently large population to provide subjects for both conditions, were included in the present study for the basis of providing the sort of informal assessment of the experimental curriculum that could result from comparisons of achievement among the pupils. It should be emphasized that the children used for comparison purposes, group 3 (N=20), although selected from the same geographic area as the children in group 2, were not randomly assigned to this condition. Any comparisons made are necessarily inconclusive, and are intended to be suggestive only.

Analogue Task Development. An important objective of the present investigation was the development of some independent means for assessing outcomes of the language curriculum under development. An approach that seemed worth while was to attempt to develop a series of tasks that would require the same cognitive skills as those thought to be developed by the language curriculum, but tasks that could be administered non-verbally.

Among those tasks used in experimental learning studies, the procedures associated with discrimination learning seemed most suitable, for they permit a motor response rather than a verbal response and the discriminations can require fairly sophisticated concepts. An apparatus was built along the lines of the Wisconsin Test Apparatus, with two compartments over which stimulus cards can be placed. The stimulus dimensions were restricted, initially, to variations in color for the elementary problems, and color and shape for the more complex problems. With these conditions, it was possible to develop a series of tasks that appeared to be analogous to many of those in the curriculum:

1. Black positive, other colors negative. This task requires a consistent response to a particular stimulus and consequently is similar to an identity statement. It does in fact require that the subject identify one particular stimulus within the set of stimuli employed.
2. Black negative. This is a simple reversal of the above task; it is analogous to a negative statement to the extent that "not black" defines the concept that would be reflected in criterion performance.

3. Same tasks as above, but with one irrelevant dimension, shape, added.
4. Simple conjunction. Either of two attributes, one color and one shape, is positive.
5. Conjunction. Two specific attributes, one color in combination with one shape, constitute a positive instance.
6. Disjunction. Either one of two attributes is positive, but not in combination.

Following their initial development, the tasks were tried out with a number of the children in order to test their suitability for the sample under study. The children were told only that they were going to play a game, that each time the apparatus was revealed they were to pick up one of the cards and if they found a piece of candy they could keep it. The most significant thing to emerge from these preliminary trials was that the children were extremely variable in their performance. Some children very quickly reached a criterion level of performance on the first task and very readily made the reversal required by the second. Other children, however, continued fifty trials or more without reaching criterion on the first task.

The major question related to these findings was whether the children who failed to evidence learning of the first problem failed because of the difficulty of the concept that was presented (which was a simple identity) or whether their failure was simply an artifact due to the experimental procedure.

In order to test the first hypothesis, that the variability was due to differences in the ability of the children to learn the concept, additional subjects were employed for the same tasks, with the exception that in this case the rule was specified verbally, e.g., "The candy is always under the black one." Given the verbal rule, the children reflected very little variability in performance, coming through all trials pretty much error free. This suggested strongly that the variability of the students in the first series had been due to something other than the cognitive skill thought to be relevant; for the subjects, provided the rule, could apply it unerringly.

In reexamining the procedure used with the first group of children, it was observed that some children, in a real sense, had reflected a set for the type of task that was presented; that is, they behaved as if they expected to find some rule of correspondence between the stimulus and the occurrence of the candy. For those children, anything less than perfect performance would provide information concerning the inadequacy of their rule, which would subsequently be modified and retested. Other children, without such a set, were reinforced on a 50% schedule for participating in the game, and, as a matter of fact, since they were

continued for many more trials, received more candy during the duration of this experiment than did those who were successful in the discrimination.

On the assumption that the application of a rule in a specific situation implies the presence of the particular cognitive skills prerequisite to the attainment of the rule, several of the analogue tasks were subjected to further test under the verbal procedure, that is, with the rule stated. It was found that there was very little variance in the performance of the subjects under these conditions--so little, as a matter of fact, that the first set of tasks seemed inappropriate for the purpose intended; very nearly all of the subjects manifested a fairly high degree of facility with the concepts of identity, negation, conjunction, and disjunction. The striking characteristic that served to differentiate among the members of the sample, on the other hand, was a set toward the formation of concepts or rules of correspondence.

The level of proficiency of the subjects, at least as manifested in the types of tasks that were used, generally exceeded that level of skill that was being trained in the summer curriculum, but not necessarily the cognitive skills developed by the curriculum generally. Additional tasks were developed, therefore, that sampled more broadly the cognitive skills that are ostensibly cultivated by the complete language curriculum.

The series of tasks subsequently developed can be divided into three sets: one requiring the identification of five colors, three shapes, and two sizes; a second requiring a response based upon these characteristics; and a larger set in which two characteristics are correlated in a display while only one characteristic is revealed in a test situation. In these items, the correlation of characteristics should provide the basis for an inference concerning the "missing" characteristic. In anticipation of further attempts to devise a technique for evaluating the curriculum, each of the tasks was constructed in such a fashion that it could be used in concept learning studies. Again the first concern was with the level of difficulty of the tasks (the ability to discriminate among the members of the sample under study) and the correspondence of the tasks to those abilities fostered by the curriculum.

The analogue tasks were subsequently administered at the same time as the School Readiness Tasks in order to get additional information concerning the characteristics of the tasks.

Testing. Much of the time during the actual conduct of the summer Head Start classes was given over to the development of standard criteria for the assessment of the pupil's level of achievement.

The School Readiness Tasks (SRT) by Barbara Bateman<sup>1</sup>, since it was specifically designed for this purpose, was adopted as a criterion of achievement for the curriculum. Following the summer project, the pupils from groups 2 and 3 attended common classes, thus receiving a common post treatment, and were drawn from these classes for the purpose of testing both with the SRT and the analogue tasks. Considerable delay was realized in making necessary arrangements for the testing to take place; consequently the final testing did not take place until about two months following the beginning of the regular kindergarten class.

Analysis. The SRT tests were scored according to the criteria that had been developed by the University of Hawaii Head Start Evaluation and Research Center, and scores for both "emerging" and "adequate" levels of performance were recorded for each subject with the provision that if the subject was judged adequate on any task he was also credited at the emerging level of achievement. The sum of these two scores, then, produced a three-point rating scale of the students' level of achievement on each Bateman task. The scoring of the performance of the subjects on the analogue tasks proceeded in a binary fashion, the subject receiving a score of one for each problem responded to correctly. Intercorrelations among all of the items as well as sub-scales and total scores were computed, as well as means and standard deviations for each item, sub-scale, and total score.

### Results

The results of the investigation are presented in four sections corresponding to the four problems that were presented in the preceding sections.

Student Achievement. The experimental curriculum presumably varies in two dimensions, horizontally in the sense that multiple exemplars are presented for each particular linguistic form, and vertically, indicating that various linguistic forms (basic sentence structures) are presented in the order of increasing complexity. The SRT test involves the presentation of materials in conjunction with questions that require answers in the grammatical forms that comprise the curriculum. A criterion of adequacy is applied to the responses, and the subject is rated on a three-point scale, 2 for adequate, 1 for emerging, and 0 for inadequate, in respect to each of the particular skills. If all students pass the criteria for success at some particular level, the mean for the task is 2.00.

Summary statistics for the testing with the Bateman are provided in Table 1. Only the first 6 tasks, through prepositions, were included in

---

<sup>1</sup>Experimental Edition, unpublished mimeographed form. University of Illinois.

Table 1

Means and Variances for Two Experimental Groups (1 and 2) and One Control Group (3)  
Receiving Variations of the Language Curriculum

Tasks	Group 1 (Other; N=28)		Group 2 (Kalihl; N=21)		Group 3 (Kalihl; N=20)	
	$\bar{X}$	s <sup>2</sup>	$\bar{X}$	s <sup>2</sup>	$\bar{X}$	s <sup>2</sup>
1a. Identity statements, positive	2.00	.00	2.00	.00	2.00	.00
1b. Identity statements, negative	2.00	.00	2.00	.00	2.00	.00
2a. Descriptive statements, positive	2.00	.00	1.96	.05	2.00	.00
2b. Descriptive statements, negative	2.00	.00	1.96	.05	2.00	.00
3. Polar opposites	1.11	.17	.81	.34	1.10	.19
4. Prepositions	.89	.38	.38	.24	.10	.30
5a. Positive categories (inclusive)	1.39	.60	1.19	.44	1.55	.55
5b. Negative categories (exclusive)	1.11	.70	.76	.75	.90	.59
6. Inference, positive	.38	.29	.09	.09	.60	.64
7. Inference, negative	.90	.49	.67	.51	1.30	.41
8a. Conjunction, simple (and)	1.86	.19	1.76	.28	1.85	.36
8b. Disjunction, simple (or)	.57	.74	.48	.63	.55	.65
8c. Conjunction (complex)	1.50	.32	1.13	.53	1.75	.29
9. Color naming	1.50	.61	1.29	.78	1.80	.26
10. Counting	1.56	.24	1.10	.37	1.50	.25
11. Number concepts	1.54	.61	1.29	.78	1.90	.10
Total	32.32	14.97	19.10	17.69	22.90	11.49

the summer curriculum for groups 1 and 2, so assessment for this portion should be restricted to these skills. All of the pupils in groups 1 and 2 performed to criteria on the first two tasks (1a and 1b). They were uniformly successful in their use of first-order identity statements, e.g., "This is a \_\_\_\_\_." and "This is not a \_\_\_\_\_." With the exception of one pupil, all also were adequate in their use of second-order descriptive statements (2a and 2b) of the form "This \_\_\_\_\_ is \_\_\_\_\_." and "This \_\_\_\_\_ is not \_\_\_\_\_." The criterion for adequacy for the fifth task requires a polar deduction for each of four polar opposites, big-little (or small), down-up, long-short, and fat-thin (or skinny), e.g., "This box is not big. It is \_\_\_\_\_." Only 12% of the subjects responded correctly to all of these items, although 86% responded correctly to at least 2 of the 4 items. This is somewhat surprising in light of the fact that training was ostensibly given on all of these pairs. It may be that since for most classes this was the terminal level of instruction during the summer program, training at this level was not complete.

Comparison with Control Subjects. As previously mentioned, the term "control" is used only in a casual sense in the present study. The students could not be randomly assigned to classes, and there was in fact no control exercised over the curriculum offered to the children in the other classes. The teachers who taught the control classes had been provided with a version of the Bereiter curriculum that was condensed but that had wider coverage. Further, the control subjects (group 3) were confined to one geographic area (Kalihi), whereas the remaining subjects were drawn from four different areas, only those in group 2 coming from the same area. These qualifications are important in light of the fact that a comparison between experimental and control subjects shows that over all tasks, including those that were not specifically included in the summer curriculum, the control subjects performed significantly higher ( $p < .05$ ) than the experimental subjects, collectively, and higher than group 2 specifically ( $p < .01$ ).

Comparison of the achievement of the three groups on the specific tasks on which training was given shows that all groups performed at essentially the same level on the first four of the tasks.

Training on these tasks absorbed the major portion of the effort of the teachers in the experimental classes and relatively less effort within the remaining classes. Consequently, assuming an initially equivalent level of skill among the pupils in groups 2 and 3, it may be that the level of achievement necessary to reach criteria on the Bateman test could have been achieved with greater efficiency with the experimental curriculum. It is possible that the test itself fails to assess the scope of the educational attainment that has been realized with the curriculum, in which case a revision of the test would be indicated.

The differences between groups 2 and 3 on the remaining two items that deal with skills included to some degree in the summer Head Start project were statistically significant, group 3 doing appreciably better on the polar opposites task and group 2 exceeding group 3 in the use of



prepositions. To attempt an interpretation of the first difference would be hazardous, but the second can probably be related to the specificity of the particular skill assessed. Relational aspects were covered in both curricula, but only the experimental curriculum gave a high degree of consideration to the exact words required by the Bateman for the expression of relations. For instance, underneath was not acceptable in place of under. The superiority of the experimental groups would have been greater had the teachers progressed beyond this point during the summer curriculum.

The general superiority of group 3 over group 2 on the remaining items ( $p < .01$ ), those on which the experimental groups did not receive summer training, could be attributed to either the effect of their own curriculum, which did in fact embrace these topics, or to a general superiority of the children in the control group. With respect to the former hypothesis, it is difficult to conclude other than that the efficiency of the curriculum, as it was presented to the summer students, leaves much to be desired, for greater gains could have been realized across the curriculum generally as it was used for the control subjects. With respect to the latter hypothesis, the difference between the two groups in the use of prepositions would have been even greater with comparable groups. The data suggest that general ability is no substitute for specific training when the objectives of instruction are very specific.

Tests of Vertical Structure. The sequence in which the tasks comprising the curriculum are presented is dictated by logical hierarchical ordering, presumably proceeding from the relatively simple to the more complex. If in fact there is a latent organization that dictates such a hierarchy, there will be two restrictions on the statistical characteristics of the students' performance on the tasks. In the first place, the mean level of performance on the tasks should descend as the tasks ascend the hierarchy if there is any true variance among the students. The reason is that attainment on a lower task would be prerequisite to attainment at a higher level. While it would be possible to attain a lower level without attaining a higher level of functioning, it would be impossible to attain the higher level without having first attained the lower level. In this sense, then, the mean of a task at a lower level imposes the upper limit for a mean at a higher level. Any true variance among the subjects would be reflected in the variance of the means on the particular tasks, and since they can vary in only one direction, it would necessitate that the means descend at increasingly higher levels of the curriculum.

Since the reliability of the full scale, based upon internal consistency (KR-20), was .71, indicating that in fact there is some true variance among the subjects, it is legitimate to examine the progression of the means.

It is necessary to disregard the first 5 tasks in this respect, for the training has had the effect of bringing all subjects to the same level.

The last three items were also eliminated, since they are regarded as more discrete skills.

Among the remaining nine items, 36 comparisons of means are possible, all of which should be consistent with the directional expectations developed above. Only 19 of the 36 confirm the hypothesis, a finding which does not appreciably differ from the expected frequency of 18 confirming comparisons assuming that the difficulty of the items varies independently of their location in the scale.

Since most of the discrepancies were associated with two specific items, 8a and 8c, these items were eliminated from further analysis. Among the remaining seven items, 15 of a possible 21 comparisons were confirming of the hypothesis. This yields a  $\chi^2$  of 3.85, significant at the .95 level of confidence, but this confidence is attenuated by the fact that the two items were deliberately eliminated because they failed to conform to expectation.

Given the required progression of the means among the remaining items, there is one additional characteristic required for vertical arrangement of the items. If attainment at a lower level is prerequisite to attainment at a higher level, then a judgment of adequacy at a higher level implies that a student is adequate at a lower level.

In terms of a four fold table, the frequency of subjects passing the "higher" item but failing the "lower" item will be zero and the ordinary formula for a phi coefficient reduces to:

$$\text{phi} = \sqrt{\frac{bc}{(b+d)(c+d)}} \quad \text{or} \quad \frac{\bar{X} - \bar{X}\bar{Y}}{\sqrt{(\bar{X} - \bar{X}^2)(\bar{Y} - \bar{Y}^2)}}, \text{ where } \bar{X} \text{ is } \leq \bar{Y}.$$

For the purpose of examining the intercorrelations actually obtained in comparison with those whose expectancy is developed above, the subjects were scored in a binary fashion, e.g., 1 for adequate and 0 for emerging or inadequate.

For those combinations of tasks where the mean for the "lower" task was greater than the mean for the "higher" task, a condition necessary for the inference that the attainment of a "lower" skill was prerequisite for the attainment of the "higher," the expected correlation was computed for comparison with the actual correlation. Table 2 shows the variables that were compared in this fashion, the expected correlations and the actual correlations.

There is little correspondence between the two columns. Perhaps exceptions to this generalization are the correlations involving combinations of 5a, 5b and 8b. Although these correlations are markedly

Table 2

Correlation Expected on the Basis of a Hierarchical Structure  
in Comparison to Those Actually Obtained

Variables	Expected	Obtained
3, 4	.80	.12
3, 6	.39	-.05
4, 6	.48	-.04
5a, 5b	.81	.56
5a, 6	.13	-.14
5a, 7	.30	.08
5a, 8b	.57	.28
5b, 6	.19	-.11
5b, 7	.62	.08
5b, 8b	.71	.40

N=49

attenuated, they are consistent in direction and general magnitude with the expectations developed above, and the correlations are significant ( $p < .05$ ). The actual criteria imposed in the determination of adequacy can markedly affect the progression of the means, and the present analysis is presented within the context of the present scoring criteria.

Analogue Tasks. The analysis of the analogue tasks proceeded along several lines. The tasks were separated into six sets, all of which, generally, were more similar among themselves than with the remaining tasks, and KR-20 estimates of reliability were obtained. The reliabilities in Table 3 are quite high, especially in light of the limited number of items associated with each sub-scale. There is also evidence of a fairly strong general factor reflected in the reliability of the total score (.86).

Evidence of the benefit of the experience with the first set of analogue tasks is found in the fact that of the 30 items in the present form, only 3 were subsequently discarded for lack of variance. The mean item difficulties indicate that in the main the items are discriminating quite well, and in this case there is a general decline of the means as the complexity of the reasoning is increased.

Examination of the tasks reveals that successful performance presupposes a knowledge of the names for the relevant stimulus dimensions, the color and shape in most instances. The variance in these basic abilities accounts for at least part of the variance on the more difficult items. But the proportion of children responding correctly on the succeeding items is considerably lower than would be anticipated simply on the basis of success in the basic skills, especially as these analyses are made in respect to the specific color and shape used in the more difficult task. It does appear, therefore, that an important characteristic is being tapped by the present series of tasks that may in fact serve for the assessment of some portions of the present curriculum, or any other curriculum that holds the development of logical abilities to be an important objective.

Turning directly to the correspondence of the analogue tasks with the curriculum, performance on the analogue tasks can be compared with the tasks comprising the Bateman test. Certainly it would be hoped that the performance of subjects on the two instruments would be correlated to some degree, thus affirming that "we are at least in the same ball park" with the analogue tasks. At the same time, the correlations should not be too high, for the Bateman should be measuring a selection of fairly specific abilities that would not be common to the analogue tasks. The over-all correlation of .52 between the two instruments confirms general expectations along these lines, but because of the global nature of the total scores, tends not to be particularly revealing. An alternative explanation of the correlation might be that both instruments are simply measuring a common general factor, "intelligence." While common-factor variance might contribute to the correlation, the reliabilities of the

Table 3

Analysis of Analogue Tasks with Means, Standard Deviations,  
Mean Item Difficulties, and Estimates of Reliability

Tasks	No. of Items	$\bar{X}$	SD	$\bar{P}$	KR-20
I Color Naming	5	3.85	1.46	.77	.75
II Shape Naming	3	2.63	.72	.88	.60
III Simple Response	6	5.43	1.01	.90	.46
IV Conditioned Response	3	1.76	.94	.58	.28
V Inference	8	3.88	2.68	.48	.85
VI Complex Inference	2	.78	.54	.39	--
T Total Scale	27	18.33	3.23	.68	.86

two scales would lead one to expect a correlation of .60 or higher. The actual correlation, then, supports the interpretation of specific factor variance, even though it is not significantly lower than the expected correlation.

In comparing performance on the analogue tasks with performance on the Bateman, it is possible to look at the relationship with two types of criteria employed with the Bateman, the proportion of subjects judged adequate in their responses, and the proportion judged emerging. The intercorrelations among items were first examined to see if they revealed any coherent order that would support the claim of legitimate relationships rather than fortuitous correspondence. In the main, this was confirmed. Comparisons involving the first four Bateman tasks were precluded by the lack of variance on these items. Using the emerging criterion on Task 3 (polar opposites) of the Bateman, there were significant correlations with the analogue tasks requiring negation (.29), polar negation (.33), and generally consistent but low correlations with the set of 8 items requiring inference (group V, the combined correlation being .24, which is significant only at the .10 level. Within the context of these data, the variable seems to be tapping several general characteristics when assessed at the emerging level. but no appreciable relationships were found when the other criteria were employed. Evidently, the difference between "emerging" and "adequate" rests upon the attainment of some fairly specific skills, and the criteria should be reexamined.

So few subjects passed the criteria for adequacy in the use of prepositions (Task 4) that there is a reasonable chance that the correlations obtained are simply fortuitous, but there was a marked correlation between this variable and the first inference task in the analogue series (.45 on the first trial, .44 on the second). The correlations of these variables also were significantly related to performance as judged by the criteria for emerging (.28 and .29 for the first and second trials, respectively). The correlations with the remaining inference items were positive but generally low.

Task 5a of the Bateman (positive categories) correlated significantly with the shape-naming analogue tasks in this case, suggesting that shape-naming is a matter of classifying the stimuli on the basis of shape. Negative categories, Task 5b, on the other hand, was significantly related to the conditional response items (which include negation) and to the simplest inference items among those on the analogue task. But these findings hold only with the scores obtained from the emerging criteria.

The means for the subjects on inference on the Bateman (Task 6) were extremely low, thus restricting the variance on these items. Even so, the only significant correlations were in association with two of the inference items on the analogue tasks. This finding suggests that there is some true covariance among the items, but this question should be

further investigation. A similar result occurred with Task 7, negative inference, but in this case the correlations were less consistent and were restricted to those involving the scores indicating adequate responding.

Few of the remaining Bateman tasks presented consistent patterns of correlation with the analogue tasks, with the marked exception of Task 9, color-naming, which was significantly correlated with most of the analogue tasks. This might be at first interpreted as evidence of a general factor that was evidenced in all of the tasks, but might more constructively be regarded as a reminder that nearly every analogue task included color as a relevant dimension. Competency with color was a general requirement for successful performance on the analogue tasks in order to eliminate that source of variance in performance on the tasks involving inference.

The correlations obtained provide some concurrent evaluation of the curriculum and indicate several tasks that might be used for further evaluation of various portions of it. They fail, however, to provide the unambiguous evaluation that has been consistently lacking in this area of investigation. Such evaluation might profitably be undertaken in a piece-meal fashion, studying only selected portions of a curriculum at a time. It is far too simple to be either too close to a curriculum, such that the evaluation task simply repeats the learning task, or too far away. The findings reported above may have served in a small degree to point the way.

## References

- Bereiter, C. Nineteenth century engineering as a model for educational research. Mimeographed paper presented at the American Educational Research Association Convention, February, 1967.
- Bereiter, C. & Engelmann, S. Teaching disadvantaged children in the Preschool. New Jersey: Prentice Hall, 1966.
- Gulliksen, H. Intrinsic validity. American Psychologist, 1950, 5, 511-517.
- Harter, Susan. Discrimination learning set in children as a function of IQ and MA. Journal of Experimental Child Psychology, 1965, 2(1), 31-43.
- Haygood, R. C. & Bourne, L. E. Jr. Attribute- and rule-learning aspects of conceptual behavior. Psychological Review, 1965, 72(B), 175-195.
- Homme, L. E. Contingency management. Supplement to the IRCD Bulletin, 1966, 2(4A). New York: Eric Information Retrieval Center on the Disadvantaged Project Beacon, Ferkauf Graduate School of Humanities and Social Sciences.
- Levinson, B. & Reese, H. W. Patterns of discrimination learning set in preschool children, fifth-graders, college freshmen, and the aged. Monographs of the Society for Research in Child Development, 1967, 32(7).