DOCUMENT RESUME

ED 027 076 PS 001 555

By-Nimnicht, Glen; And Others

Research on the New Nursery School. Part I, A Summary of the Evaluation of the Experimental Program for Deprived Children at the New Nursery School Using Some Experimental Measures. Interim Report.

Colorado State College, Greeley.
Pub Date Dec 67
Note-46p.
EDRS Price MF-$0.25 HC-$2.40
Descriptors-*Autoinstructional Programs, Concept Formation, Disadvantaged Environment, *Enrichment Programs, Language Skills, Nursery Schools, *Preschool Children, *Preschool Evaluation, Preschool Programs, Preschool Tests, Problem Solving, Program Effectiveness, Self Concept, *Spanish Americans
Identifiers-Autotelic Responsive Environment, Children's Categories Test, Cincinnati Autonomy Test Battery, Metropolitan Reading Readiness Test, PPVT, Preschool Inventory

The New Nursery School (NNS) program was set up to help 3- and 4-year-old, Spanish-surnamed, environmentally deprived children. The objectives set were (1) to improve self-image, (2) to increase perceptual acuity, (3) to improve language ability, and (4) to improve problem-solving and concept-formation skills. The school is organized as an autotelic responsive environment which the children attend for 3 hours a day. The program has been operating for 3 years, with the number of children participating each year being 30, 30, and 50. Evaluations have been based on pretests and posttests of NNS children, on comparisons with middle class children who also use the school, and on comparison in kindergarten and first grade with children from similar backgrounds. Tests have been selected or developed to measure program effects. Due to the small sample sizes, the results are quite tentative. But the following conclusions are supported for graduates of the program: (1) the school seems to improve their self-image and (2) their language and perceptual development is ahead of what would be expected had they not had the program. One negative datum is that first grade teachers do not see any difference between NNS and other deprived children which suggest that the program effects may wash out by the middle of the first grade. (DR)

INTERIM REPORT: RESEARCH ON THE NEW NURSERY SCHOOL

PART I: *A Summary of the Evaluation of the Experimental Program for Deprived Children at the New Nursery School Using Some Experimental Measures*

December, 1967

Glen Nimnicht
John Meier
Oralie McAfee

COLORADO STATE COLLEGE
Greeley, Colorado

ACKNOWLEDGEMENTS

We would like to acknowledge the assistance of our staff members at the New Nursery School at Colorado State College in Greeley, Colorado. Specifically, the contributions of Gerald Brudenell, Nellie Denning, Ann FitzGibbon, Berdie Lee Rogers, and Ronald Warner are appreciated.

Glen Nimnicht
John Meier
Oralie McAfee

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (CONTINUED)

PROGRESS REPORT ON EVALUATION AND ON THE USE OF EXPERIMENTAL MEASURING DEVICES AT THE NEW NURSERY SCHOOL

INTRODUCTION

The particular concern at the New Nursery School has been for three- and four-year-old environmentally deprived Spanish-surnamed children.[1] In addition to environmental deprivation, these children have a different culture and language. Our basic assumption is that if we can demonstrate the effectiveness of a carefully designed nursery school program with these children, a similar program will benefit other environmentally deprived children. The effectiveness is to be measured by how well the children do in school when compared to children from non-deprived background and to deprived children who had not been exposed to our program. During the first two years all of the children (30 per year) had Spanish surnames; beginning the third year the group was expanded to forty-five children, some of whom were deprived anglo children.

OBJECTIVES

The objectives of the school are:

1. To develop a positive self-image;
2. To increase sensory and perceptual acuity;
3. To improve language ability; and
4. To improve problem-solving and concept formation abilities.

We chose these four objectives because the studies and research indicated that environmentally deprived children had not developed in these areas to the extent that one would expect from observation of other children. This lack of development logically seems to be related to deficiencies in their environment.

THE APPROACH

The entire school is organized as an *autotelic responsive environment.* An *autotelic* activity is one done for its own sake

---

[1] In the main, these children are Spanish and Indian. The designation Spanish-American or Mexican-American is used interchangeably by the general population, but the individuals involved make distinctions according to their family's origin. Some families came from Spain, settled in the Southwest when it was still under Mexican rule, and intermarried with the Indians; others settled in Mexico, and then moved to the United States. Regardless of origin, some individuals prefer to be called "Spanish-American" and some prefer "Mexican-American." To avoid offending any of these people and to simplify writing, we refer to them all as Spanish-surnamed.

rather than for obtaining rewards or avoiding punishment that have no inherent connection with the activity itself.

A responsive environment satisfies the following conditions:

1. It permits the learner to explore freely;
2. It informs the learner immediately about the consequences of his actions;
3. It is self-pacing, i.e., events happen within the environment at a rate determined by the learner;
4. It permits the learner to make full use of his capacity for discovering relations of various kinds; and
5. Its structure is such that the learner is likely to make a series of interconnected discoveries about the physical, cultural or social world (Moore, 1963, p. 2).

By insisting that all activities are *autotelic*, we create a situation where we know the child is doing something because he wants to and not because an *adult* is applying pressure. This means that in observing the child's behavior in the classroom we can assume the child makes choices and carries out certain activities that are not pressed upon him by an adult. Thus, we can study curriculum development, and the relationship between maturation and learning without fear of pushing the child beyond his capacity.

The notion of a *responsive environment* is equally important. We control what the child will do by the choices we make about what to include in or exclude from the learning environment. Once the child enters the classroom, he is free to explore. He can spend as much time on any activity as he likes; no one will ask him to stop one activity to begin another. This has some interesting consequences. For example, the concept of "attention span" must be modified. These children do have a short attention span if they are required to do what the adult wants to do when the adult wants to do it. But when the children are allowed to choose their own activities, this no longer holds. Many children have been read to for an hour and a half. One child painted 25 pictures without stopping. Another spent the whole three hours, except for time out for refreshments, playing a game which required him to recognize and match pictures. Some children will spend over half of their time, particularly at the beginning of the year, playing with the blocks. But as the year progresses, their activities become more varied and they spend some time in the reading corner, the listening corner or the manipulative toy area. There are group activities such as singing and story telling, but no child is required to take part. At the beginning of the year several (five or six of the fifteen) will choose not to join the group, but day by day they scoot closer until they also join in the activities. After that it is a rare occasion when a child chooses not to come.

The notion that the environment informs the learner immediately about the consequences of his actions determines the kind of equipment that is used, the way it is used, and the behavior of the teacher and her assistants. The learner is informed either by the self-correcting toys, machines, other children, or the teacher. Most of the manipulative toys are self-correcting. The nesting and stacking toys go together or stack in only one way; the puzzles are the same. Concentric circles, squares, or rectangles must fit inside each

other to complete the pattern, and so forth. The Bell and Howell *Language Master* is an example of a machine that tells the child about the consequences of his actions. The *Language Master* records and plays back sound recordings on two channels on magnetic tape located across the bottom of cards that vary in size from 8-1/2" by 11" down to 3" by 6". One can write or draw on the card so that a child sees and hears something at the same time. The child can operate the machine without assistance, and he is free to play with it. For example, the colors can be painted on cards and "This color is red," and so forth, can be recorded on one of the sound tracks. The child can then run the cards through the machine to find out the name of the color on that card.

The teacher and her assistants are another source for the child to use in finding out the consequences of his acts. The important thing for the teacher and her assistants to remember is that they are a part of the responsive environment and therefore they respond to the child as he spontaneously encounters and manipulates his surroundings--*they do not teach; they facilitate children's learning*. This statement will become evident as we elaborate upon the specific approaches we use to obtain the objectives of the school. In general we do the following:

1. Discourage adult initiated conversation but encourage child initiated conversation;
2. Never ask a child if he wants to be read to but always read to him when he asks to be read to;
3. Avoid asking a child to give up one activity to do something else; and
4. Never insist that any child come to a group activity.

Most of a child's three hours in school is spent in self-directed activities such as painting, working puzzles, looking at books, dressing up, building with blocks, and a host of other activities. About fifteen minutes a day are devoted to group activities such as singing, listening to a story, or participating in a planned lesson.

Once each school day a booth assistant asks a child if he would like to play with the typewriter. If he says, "Yes" the assistant takes him to one of two booths equipped with an electric typewriter. The child is allowed to play with the typewriter for as long as twenty minutes. The child begins in the booth by simply playing with the typewriter. The assistant answers his questions and names the symbols he strikes, such as "x," "a," "y," "comma," "space," and "return." The child will move from this first phase to finding and striking a letter that is shown to him; he will move on to typing words and eventually to dictating stories to the booth assistant who transcribes the stories. Finally, he will transcribe his own stories.[1]

LIMITATIONS IN EVALUATING THE EFFECTIVENESS OF THE PROGRAM

There are many factors that limit the correct evaluation of the program. Some general problem of assessment must plague any research project dealing with young children. Various measures of intelligence

---

[1] A separate report on the typing booths accompanies this report.

as predictive devices are not highly reliable with a typical population of young children; and when they are used with a group of deprived children with language problems and cultural differences, the reliability obviously decreases. Some children cannot be tested. If we assume that those children could not achieve a basal score on the test and assign a low score, the changes in mean scores for the total group from the pre-test to the post-test will be spuriously high because we can predict that at least some of this group that was not tested will make the greatest gains. We have chosen to be conservative by including the children in the report who were not pre-tested so that they will remain in the longitudinal study, but not to make any assumptions about their IQ scores at the time they entered the nursery school.

Another general problem is the testing process itself. The variability of the tester does affect the test results with young children. The first two years we had one tester administer all of the Stanford-Binets, but the testing period extended over three months. Other researchers, who are working with young children, have reported significant change in IQ scores over the first two months' period of time; this means that some of our children probably had made significant gains before they took the pre-test. The third year we employed several testers, but there was an observable difference in the rapport the different testers had with the children. For example, some testers had difficulty in getting the children to go with them to the testing booth while other testers did not have the same problem. We cannot estimate the effect this has upon the test scores.

We are attempting to overcome some of these problems by experimenting with new tests that are easier to administer and provide more precise information about the children. We are training the teaching assistants and booth assistants who are working with the children to administer these tests to minimize examiner effect. Since these individuals are in daily contact with the children, they do not have to take time to establish rapport; and since they are regular staff members, the testing can be completed in a reasonable time.

In addition to these general problems, we also have problems which are specific to our project. The first two years the primary focus of our activities was the development of procedures and curriculum that seemed promising according to both a review of relevant research and to a study of other attempts at early intervention. Our basic approach was to start with what we considered a sound approach and to modify it as we analyzed what we were doing and trying to accomplish. This has two implications. First, it is impossible to say precisely what any one component of the program contributed to the achievement of our objectives. We have now developed a rationale for including each component of the program, but our research design for the first three years does not enable us to empirically validate our rationale for including each component. The second problem is that the program has been in a rapid state of change as we test new approaches and eliminate others. To some extent this will always remain a problem, but by the beginning of the third year the school had settled into a basic routine with clearly defined procedures and objectives. The developmental work continues, but the rate of change is much slower, so that each succeeding year

will be more nearly a replication of the previous year; we will thus be able to examine each component more precisely.

Another problem in analyzing our data is the fact that we have three- and four-year-old children in the school. This means that the first year's data must be analyzed for two separate groups because at this age a year obviously makes a difference in the child's development. Beginning the second year, the analysis of data must take three groups into account: the three-year-olds, the four-year-olds who are in their second year, and the four-year-olds who are in their first year. This reduces the size of each group to the point that statistical analysis of data is difficult. This will be less of a problem as the study continues and the number of children involved increases.

Recognizing all of these limitations, we are reporting on our results at this time because we are obviously making choices in developing curriculum and procedures based upon existing data; we have a responsibility to provide the best estimate of our program that is possible at this time. This report will serve as a reference as the project progresses, and provide baseline data for future analysis.

BASIS FOR EVALUATION

We have the following information:

1. Pre- and post-test scores over a three-year period for the New Nursery School children (NNS) using the Stanford-Binet and the Peabody Picture Vocabulary Test.

2. Starting the second year we started a school, the Responsive Environment Nursery School (REN School), for children whose parents could afford to cover the cost of the school by paying tuition; therefore, these children come from middle-class homes. These children are offered the same kind of program using the same procedures. We have collected some of the same data on this second experimental group as we have on the NNS group.

3. When the children who attended the New Nursery School enter kindergarten, we selected a comparable group of children from the same kind of background. While our selection was based on sociological data, the group of children who did not go to the school cannot be considered as a control group because we attempted to select the most deprived children to attend the school, and the population we are drawing from is not large enough to provide a comparable group who are equally deprived. As a result, the second group comes from deprived homes, but their families are in somewhat better circumstances. Some of these children attended an eight-week Head Start program. When these children enter kindergarten, we test them using a Stanford-Binet and collect data from the schools on both the NNS group and the other group. To date we have kindergarten and first grade teachers' estimates of the children's success in school for the groups

that entered kindergarten in September, 1965, and an assessment of their self-concept in the first grade. We also have Metropolitan Reading Readiness Test scores administered at the end of the kindergarten year for the two groups who entered kindergarten in September, 1966.

4. For the REN and NNS children we have a daily report which includes information on (1) who gave the child an opportunity to type that day; (2) did he accept or reject the invitation; (3) how long the child stayed in the booth if he accepted the invitation; (4) how many strokes he made on the typewriter; and (5) a written report by the booth assistant on what the child did and said in the booths.

   Every fifteenth day the teacher and assistant in the room observe one particular child and fill out a report on the child's activities during that day. If any noteworthy event concerning a particular child occurs at any time, the teacher writes up an informal anecdotal report which goes in his file. Based upon this information and data about the child's background we have case studies for most of the children.

5. During the third year (1966-67) we have used a number of experimental tests with both the NNS and REN children. Each test will be discussed later in this report.

THE BASELINE DATA

In order to establish a baseline for the longitudinal study, we tested the children during the first two months they were in the school using the Peabody Picture Vocabulary Test (PPVT) and the Revised Stanford-Binet (RSB). They were retested on the PPVT at the end of the school year and on the RSB one full calendar year after they took the first test. The results are reported in Tables I, II, and III. Table IV shows the RSB scores for the New Nursery School children during the first two months in kindergarten and the scores of a comparable group who were tested at the same time.

CRITERIA FOR EVALUATION

As we stated in the first paragraph of this report, the purpose of the New Nursery School is to improve the children's chances of success in school; therefore, the criterion for measuring the success of the program would be the relative success of the children in school. We also define four basic objectives:

1. To develop a positive self-image (affective domain); and
2. To develop intellectual ability (cognitive domain) by:
   a. increased sensory and perceptual acuity;
   b. improved language ability; and
   c. improved problem-solving and concept formation abilities.

These four objectives can be considered intermediate criteria of success so we will discuss our evaluation and work with experimental tests in relationship to these objectives.

EVALUATING THE DEVELOPMENT OF A POSITIVE SELF-IMAGE--

THE BROWN-IDS SELF-CONCEPT REFERENTS TEST

To test the development of a positive self-image we experimented with a technique developed by Bert R. Brown at the Institute for Developmental Studies at New York University.[1] The procedure is to take a snapshot of the child, show him the picture, and ask him a series of questions about himself, such as ---

1. Is Jesse happy or sad?
2. Is Jesse clean or dirty?
3. Is Jesse good looking or ugly?
4. Does Jesse like to play with other kids or doesn't he like to play with other kids?

The child is asked the same questions using his mother as a referent (Does Jess's mother think he is happy or sad?), using the teacher as a referent, and finally his peers (the other kids) as referents.

We administered the test to 12 NNS children and 16 comparable children when both groups were in first grade (Jan. and Feb., 1967). We do not have complete data on all of the children because when we first started administering the test we thought the testing period might be too long; we did not use mother and peers as referents because our emphasis was on self-concept as it applied to school. But, after administering some tests, we decided we could use the other referents and they were included in the balance of the tests. Based upon an analysis of Brown's report we also reduced the number of questions from fourteen to eleven.

The NNS children had fewer negative responses than the comparable group. With self as a referent the NNS children had 17 per cent negative responses and the comparable group had 24 per cent; with teacher as a referent, the NNS children had 13 per cent negative responses and the other group had 19 per cent. Obviously, neither group had a high percentage of negative responses. Since Brown had reported significant differences between the self-concept of negro and white children but a small percentage of negative responses for both groups, we asked Dr. Barbara Mickey, a cultural anthropologist, to observe some of the test situations. Her report, plus notes made by the tester, lead us to conclude that the test probably is not discriminating as well as it should on the basis of their self-concept, because some of the children, even though they gave a positive response, probably had negative feelings. The notes indicate that some of the children were withdrawn, tense, and hesitant before they responded to the items. They answered with "I don't know . . . Happy I guess." This was in sharp contrast to some of the children with no negative responses who were talkative and smiling, and responded, "I'm happy" or "Clean," without pause or hesitation.

---

[1]Bert R. Brown, "Assessment of Self-Concept Among Four-Year-Old Negro and White Children," Mimeo report, Institute of Developmental Studies, 1966.

TABLE I

PPVT MEAN I.Q. SCORES AND STANDARD DEVIATIONS
FOR
NEW NURSERY SCHOOL CHILDREN WHO ATTENDED THE
SCHOOL DURING SCHOOL YEAR OF 1964-65-66 or 1966-67

GROUP	N	PRE-TEST AT BEGINNING OF FIRST YEAR. MEAN	S.D.	TEST AT END OF FIRST YEAR MEAN	S.D.	DIFF.
1964-65						
NNS 3	14	85.29	18.30	89.36	16.50	4.17
NNS 3*	-	-	-	-	-	
NNS 4	12	86.58	14.92	94.0	13.37	7.42
NNS 4*	1			84.	-	
1965-66						
NNS 3	3	90.0	5.19	99.0	1.73	9.00
NNS 3*	2	-	-	78.5	27.57	
NNS 4-1	5	99.80	22.07	85.0	26.83	-14.80
NNS 4-1*	3			98.0	15.72	
NNS 4-2**	10	85.90	20.94	86.80	9.17	.90
NNS 4-2*	1			101.		
1966-67						
NNS 3	17	76.82	19.89	79.53	16.97	2.70
NNS 3*	7			88.57	13.26	
NNS 4-1	11	71.00	23.97	81.82	17.67	10.82
NNS 4-1*	1			65.	-	
NNS 4-2***	6	92.50	9.06	99.67	6.02	7.17
NNS 4-2*	1			99.		

* No Pre-Test
** Pre-Test 1964
*** Pre-Test 1965

TABLE II

STANFORD-BINET MEAN I.Q. SCORES AND STANDARD DEVIATIONS
FOR
NEW NURSERY SCHOOL CHILDREN WHO ATTENDED THE SCHOOL
DURING THE SCHOOL YEAR OF 1964-65, 1965-66 or 1966-67

GROUP	N	PRE-TEST MEAN	PRE-TEST S.D.	POST-TEST MEAN	POST-TEST S.D.	DIFF.
1964-65						
NNS 3	11	96.64	15.14	99.18	5.21	2.58
NNS 3*	3			94.00	11.27	
NNS 4	11	92.00	8.00	94.73	5.46	2.73
NNS 4*	1			82.0	-	
1965-66						
NNS 3	6	99.83	6.94	95.33	6.15	4.50
NNS 3*	-			-	-	
NNS 4-1	6	89.33	24.05	93.67	11.17	4.34
NNS 4-1*	3			87.33	18.48	
NNS 4-2	10	-	-	96.10	12.72	3.90
NNS 4-2*	2			92.00	4.24	
1966-67						
NNS 3	17	89.65	14.42			
NNS 3*	10	-	-			
NNS 4-1	17	81.76	14.02			
NNS 4-1*	1	-	-			

* No Pre-Test Score

TABLE III

MEAN I.Q. SCORES AND STANDARD DEVIATIONS ON THE PPVT AND STANFORD BINET FOR REN CHILDREN 1965-66 AND 1966-67

GROUP	N	PRE-TEST PPVT MEAN	S.D.	POST-TEST PPVT MEAN	S.D.	DIFF.
1965-66						
REN 3	6	95.50	25.13	110.33	23.28	14.93
REN 3*	5			91.40	27.32	
REN 4	9	105.89	24.40	111.44	15.07	5.55
REN 4*	2			94.0	41.01	
1966-67						
REN 3	5	122.80	8.90	125.00	6.60	2.20
REN 3*	3			90.33	18.77	
REN 4-1	10	120.20	17.94	114.60	12.09	-5.60
REN 4-1*	7			122.00	13.10	
REN 4-2**	5	98.80	28.03	111.40	25.07	12.60

GROUP	N	PRE-TEST RSB MEAN	S.D.	POST-TEST RSB MEAN	S.D.	DIFF.
1965-66						
REN 3[1]	3	116.00	22.87	113.67	25.33	2.33
REN 3*	-			-	-	
REN 4	8	114.63	17.73	115.38	17.34	.75
REN 4*	1			86.	-	
1966-67						
REN 3	5	115.40	14.59			
REN 4-1	10	113.20	16.25			
REN 4-2	5	111.20	22.03			

* No pre-test score
** Pre-test 1965

TABLE IV

MEANS I.Q. SCORES ON STANFORD-BINET AND STANDARD DEVIATIONS
FOR
NEW NURSERY SCHOOL CHILDREN WHO ENTERED KINDERGARTEN
IN 1965 AND 1966 AND FOR TWO COMPARABLE GROUPS

	NNS			CONTROL			DIFF.
	N	MEAN	S.D.	N	MEAN	S.D.	
1965	12	93.67	6.37	22	85.63	10.31	8.04
1966	21	93.76	12.18	28	83.75	18.66	10.01

The testers reported that the NNS children were generally more responsive and talked more in the testing situation than the other children did.

Our tentative conclusions are:

1. The test in its present form is not adequate and some other measure of self-concept is needed. During the 1967-68 school year, we will be testing other procedures.

2. What little data we have seems to indicate that first grade children who attended the NNS have fewer negative responses to images of themselves and are more self-assured in the testing situation than children from a similar background who did not attend the school.

EVALUATING THE DEVELOPMENT OF THE SENSES AND PERCEPTUAL ACUITY

The evaluation of the senses and perceptual acuity has for the most part been a part of the teaching process and no special measuring devices have been used. The exception is measuring the children's ability to identify and name the colors. This test utilizes the *Language Master*, but it can be administered without special equipment. The colors of black, white, red, orange, yellow, brown, green, purple, and blue are painted on nine different *Language Master* cards, one color on each card. The name of the color is recorded on the instructor's channel, "The color of paint on this card is blue." The tester presents one card at a time and asks, "What color is the paint on this card?" The teacher waits for three to five seconds for a response. If the child does not respond or gives a wrong answer, we assume that he cannot name that color. In any event, the tester allows the child to run the card through the machine himself to hear the correct answer. (If the *Language Master* is not used, the tester tells the child the correct answer.)

The obvious weakness of the test, as described, is that it requires a verbal response. We have revised it so that the child is first tested on his ability to identify the color by pointing to it and later by asking him to give the name for a specific color. We have also extended it to test one form of problem solving--obtaining the right answer by eliminating wrong responses. As the test was used in 1966-67, it does provide some useful information. Table V shows the pre- and post-test scores for three groups of NNS children and REN CHILDREN. It is obvious that the three-year-old REN (Ren-3) and the four-year-old REN children who are attending the school for the first time (REN 4-1) can name more colors than the comparable groups of NNS children. The NNS children improve over the course of the year and by the end of the second year can name all of the colors.

We will be devoting more time to developing other more formal measures of specific development of the senses and perceptual acuity but many of the activities involved in developing the senses and perceptual acuity are an integral part of language development and concept formation and can be evaluated in that context, so that most of our effort has gone into experimenting with more general tests.

TABLE V

The Mean Pre- and Post-Test Scores and Mean Change for the NNS and REN Children on the Test of a Young Child's Ability to Name the Colors.

GROUP	N	Pre-Test	Post-Test	Mean Change
NNS 3	15	2.0	5.2	3.2
NNS 4-1	11	1.45	6.63	5.18
NNS 4-2	5	7.8	9.0	1.2
REN 3	6	4.5	7.8	3.3
REN 4-1	10	6.4	8.0	1.6
REN 4-2	4	7.3	8.3	1.0

EVALUATING THE DEVELOPMENT OF LANGUAGE ABILITY

METROPOLITAN READING READINESS TEST

Probably the best single indicator of language development available is the results on the Metropolitan Reading Readiness Test that is administered to all kindergarten children in the Greeley Public Schools at the end of the 1966-67 school year (including the sixteen children who attended the New Nursery School and the twenty-eight comparable children). The two groups are comparable in the following ways: (1) both groups were composed of Spanish-surnamed children and (2) both groups came from low income homes.

The two groups are not comparable in one way--in the selection of the children to attend the New Nursery School we made an effort to select those children who lived in the most deprived circumstances so the comparable group of children come from poor homes but are, in our judgment, somewhat less deprived than the NNS group. We used sociological and anthropological data as a base for making this judgment.

We consider the Metropolitan Reading Readiness Test a good measure of the program because of the delay of almost a year from the time the children left the nursery school and the time the test was administered. None of the children could score on the test when they entered the nursery school.

Table VI shows how the two groups scored on the test. The mean score of 79 for the NNS children is at the 70th percentile on the test, and the mean score of 64 for the control group is at the 35th percentile. Based upon the knowledge of the children's background and IQ scores on the Stanford-Binet, the comparable group scored about as one would expect. The mean IQ's for the NNS group when they entered the nursery school was somewhere between 83 and 93 (some could not be tested), and when the two groups were tested at the beginning of their year in kindergarten, their mean IQ scores were--NNS's, 94; and the comparable group, 84. So a mean score on the test that falls at the 70th percentile is considerably better than one would predict for the NNS children if they had not attended the New Nursery School.

To illustrate this point, the results on the Metropolitan Reading Readiness Test for the six children who probably had the lowest IQ scores at the time they entered the nursery school are as follows:

> Two of the children could not be tested on either the Stanford-Binet or the PPVT when they started at the New Nursery School, three could not be tested on the Stanford-Binet but had PPVT-IQ at 79, 75, and 50. One child had an IQ score of 57 on the Stanford-Binet and 61 on the PPVT. Their percentile ranks on the Metropolitan Reading Readiness Test were 31, 93, 45, 53, 79, and 51. The 31 percentile is considered a low normal; all the other scores are average or better. These six NNS children constituted the lowest 38 per cent of that group based upon pre-test scores. We studied the lowest 38 per cent (11 children) from the comparable group based upon their IQ scores when they

TABLE VI

MEAN SCORES AND STANDARD DEVIATION ON
METROPOLITAN READING READINESS TEST FOR
NEW NURSERY SCHOOL AND CONTROL CHILDREN IN KINDERGARTEN
DURING THE 1966-67 SCHOOL YEAR

	NNS			CONTROL			
	N	MEAN	S.D.	N	MEAN	S.D.	DIFF.
READING	16	55.31	6.10	28	46.57	15.03	8.74
ARITHMETIC	16	17.00	3.98	28	12.89	5.83	4.11
TOTAL	16	78.87	9.98	28	64.43	22.12	14.44

entered kindergarten and their percentile scores on the Metropolitan Reading Readiness Test were 9, 5, 0, 3, 10, 11, 68, 61, 13, and 66. The mean of their total score was 49, which is at about 12 percentile compared to 72 (53 percentile) for the lowest NNS group; this is a difference of 23 points.

Our tentative conclusion is that the experience at the New Nursery School did improve the language ability of this group of children, as measured by the Metropolitan Reading Readiness Test taken at the end of kindergarten.

THE PRE-SCHOOL INVENTORY

The Pre-School Inventory by Bettye Caldwell and Donald Soule (memo. Children's Center, Department of Pediatrics, Upstate Medical Center, State University of New York, Syracuse, New York, no date) is still in the developmental stage, and our use of the test was designed to evaluate the test as well as our program. The authors gave the following description of the pre-school inventory.

> The Pre-School Inventory is a brief assessment procedure designed for individual use with children in the three-to-six range. It was developed to give a measure of achievement in areas regarded as necessary for success in school. It is by no means culture free; in fact one aim of the instrument is to permit educators to highlight the degree of disadvantage which a child from a deprived background has at the time of entering school in order to help eliminate any observed deficits. Another goal in the development of the procedure was to make available an instrument that was sensitive to experience and could thus be used to demonstrate changes associated with educational intervention. (p. 1).

The norms are currently based upon 171 children who attended a Head Start program during the summer of 1965. Through a process of factor analysis, the authors identified four factors; (1) Personal-Social Responsiveness - Factor A; (2) Associative Vocabulary - Factor B; (3) Concept Activation, Numerical - Factor C; and (4) Concept Activation, Sensory - Factor C_2. The following is a brief description of each factor's composition:

> Factor A. *Personal-Social Responsiveness*. This factor appears to involve knowledge about the child's own personal world (name, address, parts of body, friends) and his ability to establish rapport with and respond to the communications of another person (carrying out simple and complicated verbal instructions given by an adult). Perhaps more than any other factor, it represents the type of eminently practical ability which the Inventory was originally designed to assess.
>
> Factor B. *Associative Vocabulary*. This factor requires the ability to demonstrate awareness of action or by associating to certain intrinsic qualities of the underlying verbal concept. Item units having high loadings include simple labeling of geometric figures, supplying verbal or gestural labels for

certain functions, actions, events and time sequences, and being able to describe verbally the essential characteristics of certain social roles. Many of the specific deficits frequently attributed to culturally deprived children cluster in this factor.

Factor C. *Concept Activation*. This is the factor that accounted for the greatest amount of common variance. The concepts involved seem to represent two major categories: ordinal or numerical relations, and sensory attributes such as form, color, size, shape, and motion. The activation involves either being able to call on established concepts to describe or compare attributes (relating shapes to objects, color-names to objects or events) or to execute motorically some kind of spatial concept (reproduction of geometric designs or drawing the human figure). High scores on this factor involve being able to label quantities ("How many" questions), to make judgments of more or less, to recognize seriated positions (first, last, middle), to be aware of certain sensory attributes (shape, size, motion, color), and to be able to execute certain visual-motor configurations (geometric forms, draw a man).

As this factor accounted for the greatest amount of common variance on the initial version of the instrument and as it appeared to be composed of two subfactors (numerical and sensory concepts), it was given double representation in the standardization version of the instrument. The items measuring numerical concepts were separated for the convenience of test users from those sampling sensory concepts. (Caldwell and Soule, no date, p. 2.)

It should be obvious to the reader that this test purports to measure more than language ability, but it does depend to a great extent upon the child's ability to understand and use language. We have therefore chosen to include it under our assessment of language development and we can refer back to it as we discuss the evaluation of concept formation ability.

We administered the test to the NNS children at the end of the school year, 1966-67. Since we do not have pre- and post-test data, our analysis was limited to comparing the subgroups of NNS children with each other and to the norms on the test, and with correlating the results on this test with the results on other tests; however, we will be able to measure the predictive ability of the test as the study progresses. Table VII shows the results on the four factors, the total scores, and the percentile that would be given to an individual score at the mean. Table VIII shows the chi-square values that result from comparing the scores of the NNS 3's, 4-1's and 4-2's on the Pre-School Inventory.[1] Table IX shows the intercorrelations on the Pre-School Inventory and the PPVT post-test raw score for all 1966-67 NNS children.

---

[1] We have made limited use of statistics in this report because of the small number of children we are dealing with and problems of meeting the underlying assumptions that a given statistical procedure is based upon. When statistics have been used, their only purpose is to help the reader interpret the findings. We are not trying to generalize from our data.

TABLE VII

THE MEAN SCORES AND CORRESPONDING PERCENTILE SCORES FOR THE NNS CHILDREN ON THE PRE-SCHOOL INVENTORY

GROUP	N	FACTOR A		FACTOR B		FACTOR C_1		FACTOR C_2		TOTAL	
		SCORE	PERCENTILE	SCORE	PERCENTILE	SCORE	PERCENTILE	SCORE	PERCENTILE	SCORE	PERCENTILE
NNS 3	25	14.52	--- *	7.36	---	7.56	---	11.64		41.08	
NNS 4-1	14	16.07	35	9.93	25	10.50	60	14.50	75	51.00	60
NNS 4-2	7	19.00	65	15.28	70	12.28	80	16.85	95	63.42	95

Factor A Personal-social Responsiveness
Factor B Associative Vocabulary
Factor C_1 Concept Activation, Numerical
Factor C_2 Concept Activation, Sensory

* No percentile given for this age group

TABLE VIII

THE CHI-SQUARE VALUES[1] WHICH RESULT WHEN THE SCORES FOR NNS 3'S, 4-1's, AND 4-2's ARE COMPARED ON THE PRE-SCHOOL INVENTORY

GROUP	FACTOR A		FACTOR B		FACTOR C_1		FACTOR C_2		TOTAL	
	NNS 4-1	4-2	NNS 4-1	4-2	NNS 4-1	4-2	NNS 4-1	4-2	NNS 4-1	4-2
NNS 3	1.09	6.72	5.42	20.16	8.07	12.63	5.34	11.34	5.10	13.21
NNS 4-1		3.43		6.09		2.38		6.09		6.09

1. Using the Kolmogorov-Smirnov Two Sample Test df = 2 .10 - 4.60 .05 = 5.99 .01 = 9.21

TABLE IX

Inter-correlations Between the Pre-School Inventory Factors, Pre-School Inventory Total Score and the Post-test Raw Score on the PPVT for the 1966-67 NNS Children

	Pre-School Inventory				
	A	B	C_1	C_2	TOTAL
PPVT	.41	.55	.41	.42	.55
A		.56	.40	.45	.77
B			.40	.51	.80
C_1				.54	.70
C_2					.80

With an N of 42 all of these correlations exceed the .05 level of significance.

The results as they appear on Table VII are as we would have predicted: (1) There is a consistent pattern of increases in scores from NNS 3's to NNS 4-1's to NNS 4-2's. As the chi-square values in Table VIII indicate, except for Factor A, Personal-Social Responsiveness, most of these differences are probably significant. That is also where the NNS 4-2's scored the lowest according to the percentile rank, and they scored significantly better than the normative group on the other three factors. This is consistent with the program emphasis of the New Nursery School.

All of the factors have a fairly high and uniform correlation with each other and the raw scores on the PPVT post-test (see Table IX).

Our tentative conclusions from the results on the Pre-School Inventory are:

1. If the Pre-School Inventory proves to be a good predictor of school success, the results indicate that the NNS program is helping children become more successful in school; and

2. The Pre-School Inventory seems to be a promising test and we will continue to experiment with it.

EVALUATING THE DEVELOPMENT OF CONCEPT FORMATION AND PROBLEM-SOLVING ABILITY

Our evaluation to date is based upon the use of three experimental measures: some of the tests from the Cincinnati Autonomy Test Battery (CATB), the "C" Test, a test of a child's ability to relate objects that can be placed in the same classification, and the Children's Categories Test, a test of a child's ability to discover a category. All of these tests are in the early stages of development; we do not have pre- and post-test data, and we do not know how they relate to school success.

THE CINCINNATI AUTONOMY TEST BATTERY

The Cincinnati Autonomy Test Battery is being developed at the University of Cincinnati by Thomas J. Banta (memo paper to appear in Volume I of Cognitive Studies to be published in 1968). The test grew out of a study of the development of children from three to five years of age--especially the development of autonomy in thinking, perceiving, and social behavior. Autonomy is defined by Banta as "the self-regulating abilities which facilitate effective problem solving behavior." John Holt's description of intelligence, "not how much we know how to do, but how we behave when we don't know what to do" (*How Children Fail*, 1964) succinctly describes the focus of the Cincinnati Autonomy Test Battery.

The test battery consists of six tests. A brief description of each follows:

1. *Curiosity*: Tendency to explore, manipulate, investigate,

and discover in relation to novel stimuli. A curiosity box is used in the test. Attached on the sides and on the top of a box about the size of a small orange crate are all sorts of curiosities--a chain lock, a light switch, a pull chain, a window lock, a bolt, etc. There are also two peep holes and a hole big enough for a child's hand to fit into. The strange, interesting-looking thing is placed in front of the child; the examiner says, invitingly, "This is something for you to play with." For five minutes the child is left in this unstructured situation. The examiner places himself out of the line of vision of the child. The child is scored on his amount of involvement--manipulatory, tactual, and visual--and on the amount of verbalization. He receives a point for each type of involvement during a 30-second period. This test measures the way in which a subject explores or does not explore a new complexity within an unstructured situation.

2. *Impulse Control*: Tendency to restrain motor activity when the task demands it. The subject is shown how to draw a a line fast; then he is told to draw a line "very fast." He is then shown how to draw a line slowly and then told to draw three lines "very slowly." His score is based on the amount of time he takes to draw an eight-inch line very slowly. A ratio of the length of line divided by the time in hundredths of a second is obtained.

3. *Incidental Learning*: Tendency to acquire information not referred to in the instruction stimuli. The child is presented ten simple, familiar drawings; each drawing has one part colored green. The remainder of the drawing is colorless. As each drawing is presented, the child is asked to point to the green area. At the conclusion of the presentation, the subject is asked to recall the drawings, that is, recall that which was seen incidentally.

4. *Intentional Learning*: Tendency to acquire information specified in the instructional stimuli. After the incidental learning test, the same series of drawings is presented again and each named by the subject. After this practice in naming the objects, the child tries to recall the drawings. Two scores are obtained: one is based upon the number of right incidental recalls, and the second is based upon the number of right intentional learning recalls.

5. *Innovative Behavior*: Tendency to generate alternative solutions to problems. The dog and bone test used to measure innovative behavior consists of a cardboard the size of a chess board with four "houses" near the corners of the board, but with about two or three inches of space between them on the edge of the board, and a small toy dog and bone. The toy dog is placed in front of the child, and the bone is placed at the other end of the board. The examiner shows the child how the dog can go get the bone in two ways: (1) a straight line and (2) by going around one house near the child. Then the child is asked to show the examiner another way the dog can go to get the bone. Each

child has ten trials, and the scoring is based upon the number of unique ways he finds for the dog to get the bone. He can score one or two points depending upon the complexity of the route; two is for routes of greater complexity.

6. *Field Independence*: Tendency to separate an item from the field or context of which it is a part. An embedded figures' test is used. In this test the child is given a cone-shaped piece of paper and asked to place it over the same shape in a series of pictures of nature, things, people, and geometric drawings. The cone may form a valley between mountains, or a part of a cowboy's body or be hidden in a design of circles. The child sees fourteen pictures, and one point is scored for each correct response.

7. *Persistence*: Attention to a problem with solution-oriented behavior where the goal is specified. The replacement puzzle test that is used is a wood-inlay puzzle in which the pieces do not fit together, and each piece is a complete object. For example, one piece is in the shape of an airplane, and another is in the shape of a horseshoe. Some of the pieces are nailed to the tray; the others are taken out, and the child is asked to replace them. The difficulty is that there is only one way in which all the pieces will fit back into the tray. It soon becomes an overwhelming task for the child. Two points are added to the score every 20 seconds that the subject is attending to the task. A negative point is given for inattention during the same time interval. The range of scores is from minus six to plus twelve.

8. *Resistance to Distraction*: Persistence with distracting stimuli present. The replacement puzzle test is used as described above. After two minutes a distracting stimulus (blocks) is presented. The subject is told that he can either play with the blocks or finish putting the pieces in the tray. Scoring is based on two factors. The second score is based on the subject's attention to the task when a distracting stimulus is offered. A score of +3 is given for attention to the original task during three twenty-second time periods (one minute). A score of -1 is given for inattention to the task.

A tester trained by Banta administered all six tests but the reliability of the tests varies widely. According to the data supplied by Banta, the reliability of the tests of innovative behavior, curiosity, impulse control, field independence and intentional learning was acceptable, so we have limited our analysis to these tests.

Table X shows the mean scores for the group of deprived Negro children tested by Banta, the NNS children, and the REN children. In order to give the reader some indication of the significance of the difference in scores we have shown in Table XI the chi-square values that result when comparisons are made between the comparable groups of NNS and REN children. Table XII contains the inter-correlations on the CATB and the correlations of the CATB with the Pre-School Inventory and the PPVT.

TABLE X

THE MEAN SCORES FOR NNS AND REN CHILDREN (1966-67) ON THE CINCINNATI AUTONOMY TEST BATTERY

GROUP		CATB TESTS 1	2	3	4	5
CATB Group*	N M	84 13.83	84 .69	84 2.81	84 4.80	84 8.36
NNS-3	N M	22 17.32	19 .83	20 2.30	19 3.58	20 5.05
NNS 4-1	N M	14 11.36	14 .64	13 2.62	13 4.85	13 7.00
NNS 4-2	N M	6 18.17	6 .57	6 4.50	6 8.00	6 6.00
REN-3	N M	5 27.80	5 .69	5 2.80	5 9.00	5 5.80
REN 4-1	N M	12 21.50	12 .48	12 2.33	12 8.92	12 8.58
REN 4-2	N M	4 29.75	4 .30	4 2.75	4 9.50	4 10.00

1. Curiosity Box
2. Impulse Control (Measured in minutes - low score is best)
3. Intentional Learning
4. Innovative Behavior
5. Field Independence

* The CATB group consisted of deprived Negro children in Cincinnati

TABLE XI

THE CHI-SQUARE VALUES THAT RESULT FROM COMPARING NNS AND REN CHILDREN ON THE CATB

		NNS		REN		
		4-1	4-2	3	4-1	4-2
Curiosity Box	NNS 3	2.30	.52	2.97		
	4-1		1.61		7.50	
	4-2					2.40
	REN 3				2.62	.36
	4-1					1.75
Impulse Control	NNS 3	4.89	1.80	1.13		
	4-1		1.59		1.93	
	4-2					3.20
	REN 3				1.41	3.20
	4-1					3.00
Intentional Learning	NNS 3	.60	4.62	1.96		
	4-1		4.53		.86	
	4-2					3.27
	REN 3				2.45	.55
	4-1					.75
Innovative Behavior	NNS 3	3.66	3.65	3.71		
	4-1		4.10		5.55	
	4-2					2.40
	REN 3				2.62	.80
	4-1					.61
Embedded Figures	NNS 3	3.24	1.72	.93		
	4-1		2.11		4.47	
	4-2					4.27
	REN 3				3.53	3.20
	4-1					1.33

TABLE XII

THE INTER-CORRELATIONS ON THE CATB, THE PPVT AND THE PRE-SCHOOL INVENTORY

	CATB					PPVT	PRE-SCHOOL INVENTORY				
CATB	1	2	3	4	5		A	B	C_1	C_2	Total
1		-.21	.17	.01	-.16	-.01	-.26	-.08	-.25	-.08	-.18
2			.03	.03	.21	.38	.04	.15	.35	.10	.17
3				.14	.36	.44	.25	.51	.20	.41	.46
4					.23	.30	.45	.41	.42	.45	.53
5						.23	.13	.22	.48	.44	.40

N = 35

CATB
1 Curiosity Box
2 Impulse Control
3 Intentional Learning
4 Innovative Behavior
5 Embedded Figures

All three of these tables will be used in discussing the results. The first thing to note about the test of curiosity, the Curiosity Box, is that if any relationship exists between it and the other tests, it is a negative one. Since we don't know whether a relationship exists or not we will drop this test from our analysis at this time, but continue to use the test because it appears to be reliable; if it proves to have some significant negative correlations with other measures it may be useful.

The test of impulse control does seem to have a positive correlation with other measures (PPVT .38 and PSI-C_1 .35). The lower the score on this test the more control the child had in drawing a line. The pattern of the scores in both groups is uniform with the 4-2's in each group demonstrating more control and the three-year-olds demonstrating the least control. The REN children as a group have more control than the NNS children, with the REN 3's scoring about the same as the NNS 4-1's. Judging from this pattern in the data and the chi-square values in Table XI, the differences in these scores probably have some meaning.

The Intentional Learning Test was the least reliable test (test-retest .60 and internal consistency .40) but it correlates with the PPVT .44, P-SI-B .51, P-SI-C_2 .41. The scores the various groups made on the test are all fairly uniform except for the NNS 4-2's who scored significantly higher than the other groups; since the reliability of the test is low and the probability of this one difference occurring by chance is high, this test will not be used in the evaluation of the program.

The Dog and Bone test of innovative behavior is certainly one of the most interesting tests in the CATB and it may prove to be one of the most productive. The test-retest reliability as reported by Banta was .73 for one group and .82 for another and internal consistency is .76. It correlates with all the factors on the Pre-School Inventory. The scores on the Dog and Bone test for the NNS children go from the lowest score for the NNS 3's to the highest for the NNS 4-2's; scores for the REN group are about the same. The fact that the NNS 4-2's score about the same as the REN group while the NNS 3's and 4-1's score much lower may be significant.

The Embedded Figures test also seems promising, although it appears to be less reliable. Banta reports a test-retest reliability of .41 and measures of internal consistency from .48 to .80. It correlates with factors C_1 and C_2 on the Pre-School Inventory. The difference in scores among the three NNS groups are small and may be meaningless; difference in scores among the REN groups are greater and may be more significant. The REN 3's scored about as well on the test as the NNS 4-2's.

Our tentative conclusions from the results on the CATB are:

1. The main value of the CATB will be in the future evaluation of the program. However, when combined with other test results, the results on the CATB can provide some information that can be used in the current evaluation of the program.

2. The tests of curiosity and intentional learning may be valuable later on, but we cannot use them at this time.

3. The tests of Impulse Control, Innovative Behavior, and Field Independence will be used in the evaluation in this section.

THE "C" TEST

One assumption made about environmentally deprived children is that their ability to use categories and classification systems is underdeveloped. In order to test this assumption and to measure changes in such behavior we developed the "C" Test, an individually-administered test using the concrete objects listed below:

Stimulus	*Response*
orange	apple
doll	toy car
cup	bowl
glove	shoe
comb	toothbrush
pencil	crayons
cigarette	cigar
penny	dollar
hammer	screw driver
flashlight	lightbulb

All the response items are placed on a table in front of the child and named by the tester. Then one stimulus item at a time is held by the tester who says, "Show me the thing that goes with this orange." if the child names or points to the matched item he has made a correct response. A pilot test was given to 34 NNS children and to 6 REN children.

As one might expect, we discovered several errors in the construction of the first test. Associations were made which we neither expected not credited. Many associations were based upon color: the green toothbrush was matched with the green crayon, and the red apple was matched with the red glove. Based upon these findings, we revised the test to eliminate the possibility of matching by color.

Other associations were made; while these reflected ability to classify they did not meet the requirements of the test. The doll was matched with the comb because the doll's hair obviously needed combing. One child said that the lightbulb and screw driver went together because "they both screw." In the revision we placed the lightbulb and the screw driver in the response items and the comb and doll in the stimulus items to avoid this kind of matching. We also found that many children did not know what a candle was, so we substituted a flashlight.

The revised test was administered to all the NNS and REN children. Table XIII shows the mean scores for the NNS children and the REN children, and Table XIV shows the chi-square values when the scores of the various groups are compared. Table XV shows the intercorrelations of the "C" test with the Pre-School Inventory and the CATB.

TABLE XIII

THE MEAN SCORES ON THE "C" TEST FOR
REN AND NNS CHILDREN (1966-67)

Group	N	Mean
NNS 3	23	1.67
NNS 4-1	13	3.77
NNS 4-2	7	5.86
REN 3	7	3.71
REN 4-1	7	4.71
REN 4-2	3	5.00

TABLE XIV

CHI-SQUARE VALUES WHICH RESULT WHEN SUBGROUPS OF
NNS AND REN CHILDREN ARE COMPARED ON THE "C" TEST

Group	NNS 4-1	NNS 4-2	REN 3-1	REN 4-1	REN 4-2
NNS 3	5.62	16.23	5.04		
4-1		5.27		5.27	
4-2					1.22
REN 3				2.57	1.54
4-1					.30

x^2 obtained by using Kolmogorov-Smirnov Two Sample Test df = 2

.10 = 4.60 .05 = 5.99 .01 = 9.21

TABLE XV

CORRELATIONS OF THE "C" TEST WITH THE PRE-SCHOOL INVENTORY AND THE CATB

	Pre-School Inventory					Cincinnati Autonomy Test Battery								PPVT
	A	B	C_1	C_2	Total	1	2	3	4	5	6	7	8	
"C" TEST	.41	.72	.40	.37	.60	-.01	-.26	.14	.27	.14	.15	.05	.01	.47

N = 29

The pattern of the scores on the "C" Test is the same as we have observed before. The REN children as a group have higher scores than the NNS children, and the scores improve within each group, from three-year-old children to four-year-old children who are in their second year. There are significant differences between all three groups of NNS children and between NNS 3's and REN 3's, and NNS 4-1's and REN 4-1's. There are differences between NNS 4-2's and REN 4-2's.

There are correlations of the "C" Test with all factors and the total on the Pre-School Inventory. The high correlations with Factor B, Associative Vocabulary, is a predictable one. The "C" Test also correlates with Innovative Behavior and the PPVT.

Our tentative conclusions are:

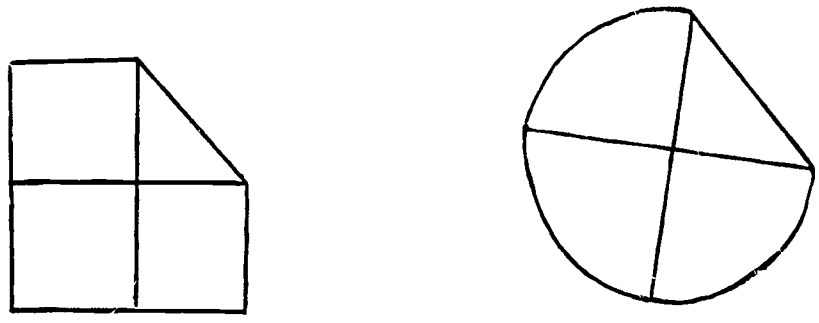
1. If the test is reliable, the high positive correlations plus the apparent ability to discriminate between groups suggest that this simple test may be very useful. We are in the process of testing the reliability of "C" Test; if it proves reliable and the correlations remain high in subsequent research, we will have a simple, easily administered test that requires no verbal response, (yet is highly correlated with Factor B, Associative Vocabulary, on the Pre-School Inventory) that can be easily translated into other languages, and that can be used in a test battery for three- and four-year-old deprived children.

2. The "C" Test will be used in the present evaluation but it can only be suggestive in its present state of development.

CHILDREN'S CATEGORIES TEST

This test is a measure of a child's ability to discriminate among abstract concepts such as the largest circle from among several circles. The test uses a set of colored slides that are projected one at a time onto a rear-vision screen about the size of a 14-inch TV screen. The child can indicate his response by pulling one of four levers which are color coded (red, blue, yellow, and green). The child matches the color of the lever with the color of his chosen response. For example, if the concept to be discriminated is "the largest circle" and that circle is colored red, the child will be correct if he pulls the red lever. In that event he hears a bell. There is no sound for a wrong response. The only data the child receives are the results of his previous selection; thus, problem-solving processes such as elimination of wrong previous approaches are involved. The first slide shows a red triangle and the task is to match color alone. The next nine slides are simply color matching. The next ten slides deal with the problem of two-of-a-kind; one slide shows two green circles, a yellow circle, and a blue circle. On the next series of ten slides the problem is to discriminate the largest shape. This is followed by a series which presents several of the same shapes with one different shape, such as a triangle among several different of same-sized squares. The next series of ten deal with incomplete parts. For example, the

child sees a square or circle divided into four colored quarters, but one quarter is incomplete. In the following examples the upper right quadrant represents the incomplete one:



On the next ten slides a complete quadrant of the square or circle is missing, and the task is to identify the missing part so that the correct response is to pull the lever for the color not seen on the screen. The final ten slides are reviews of the previous series and are designed to test for short-term memory. In scoring the test, the child receives a score on the first ten items (color), the next ten (two-of-a-kind), the ten largest, the next twenty (partial or missing parts), the last ten, and a total score. The test was originally developed by Halstead to test biological intelligence in children and has been adopted for use at the New Nursery School by John Meier.

We first administered the test to 39 NNS and 17 REN children in September and October of 1966. For the post-testing in May, 1967, we eliminated 20 slides on the basis of an item analysis in order to shorten the test and eliminate non-discriminating items. The results on the post-test are reported in Table XVI. The chi-square values for the comparison of the groups are shown in Table XVII, and the correlations of the Children's Categories Test with other measures is shown in Table XVIII.

The first observed difference in Table XVI is that the rather consistent pattern we have noted before on other measures is not evident. The REN children scored higher than the NNS children, but the NNS 4-1's scored as well as the NNS 4-2's, and the REN 4-1's scored better than the REN 4-2's. The NNS 4-1's scored higher than the NNS 3's. The REN 4-1's scored significantly higher than any other group, and the REN 4-2's scored higher than REN 3's.

The correlations with the Pre-School Inventory are fairly high and the Categories Test correlates with Intentional Learning (.48), Innovative Behavior (.37), and the Field Independence test (.37). Our tentative conclusions are:

1. Since the Children's Categories Test is closely related to the objectives at the New Nursery School and seems to be discriminating between groups and to be correlated with other measures, we should continue to experiment with the test and obtain some measure of its reliability and validity; and

2. Like the "C" test, the Categories Test has some value in our current evaluation.

TABLE XVI

THE MEAN SCORES FOR NNS AND REN CHILDREN ON THE CATEGORIES TEST

Group	N	Mean
NNS-3	10	21.14
NNS 4-1	12	26.75
NNS 4-2	6	25.67
REN-3	4	22.25
REN 4-1	10	40.90
REN 4-2	3	37.33

TABLE XVII

THE CHI-SQUARE VALUES RESULTING FROM COMPARING NNS AND REN SUBGROUPS ON THE CATEGORIES TEST

	NNS's		REN's		
Group	4-1	4-2	3	4-1	4-2
NNS-3	4.82	2.92	.57		
NNS 4-1		.44		10.69	
NNS 4-2					8.00
REN-3				5.60	6.86
REN 4-1					4.52

x^2 obtained by using Kolmogorov-Smirnov Two Sample Test df = 2

df = 2 levels of significance are .10 = 4.60 .05 = 5.99

.01 = 9.21

TABLE XVIII

CORRELATIONS AMONG TOTAL SCORES ON THE
CHILDREN'S CATEGORIES TEST AND OTHER TESTS ADMINISTERED
DURING 1966-67 TO NNS AND REN CHILDREN

	"C" TEST	PRE-SCHOOL INVENTORY					CATB					PPVT
		A	B	C_1	C_2	TOTAL	1	2	3	4	5	
CHILDREN'S CATEGORIES TEST	.27	.29	.37	.34	.58	.56	.07	-.17	.40	.37	.37	.25

CATB
1. Curiosity Box
2. Impulse Control
3. Intentional Learning
4. Innovative Behavior
5. Field Independence

SUMMARY OF THE TESTING OF CONCEPT FORMATION

AND PROBLEM-SOLVING ABILITY

All of the tests used are in early stages of development and therefore have limited value, and we do not have any estimate of their ability to predict school success. Their main value at this time is that they have provided a variety of ways to compare the NNS and REN children. If we assume that the REN children are more likely to be successful in school and the four-year-old REN children are more advanced in their development, then we can assume that these children represent a standard; the more closely the other children approximate that standard the more likely they are to be successful in school and school related-endeavors.

To make it easier to see the results we have organized them in a graphic form. The reader should be aware of the fact that some visual distortion is included because not all of the graphs start at zero. The graphs do illustrate the point that the REN children consistently do better than the NNS children and the REN 4's score better than the REN 3's. It also appears that the REN 3's are scoring as well as, or better than, NNS 4-1's.

It should be noted that on the "C" test the NNS 4-2's score as well as or better than the REN 4's. On the Dog and Bone test of innovative behavior the NNS 4-2's scored about the same as the REN groups and on Impulse Control they scored about the same as the REN 4-1's but not as high as the REN 4-2's. On the Children's Categories Test and Embedded Figures they score much lower than the REN 4's and about the same as REN 3's.

Our tentative conclusions are:

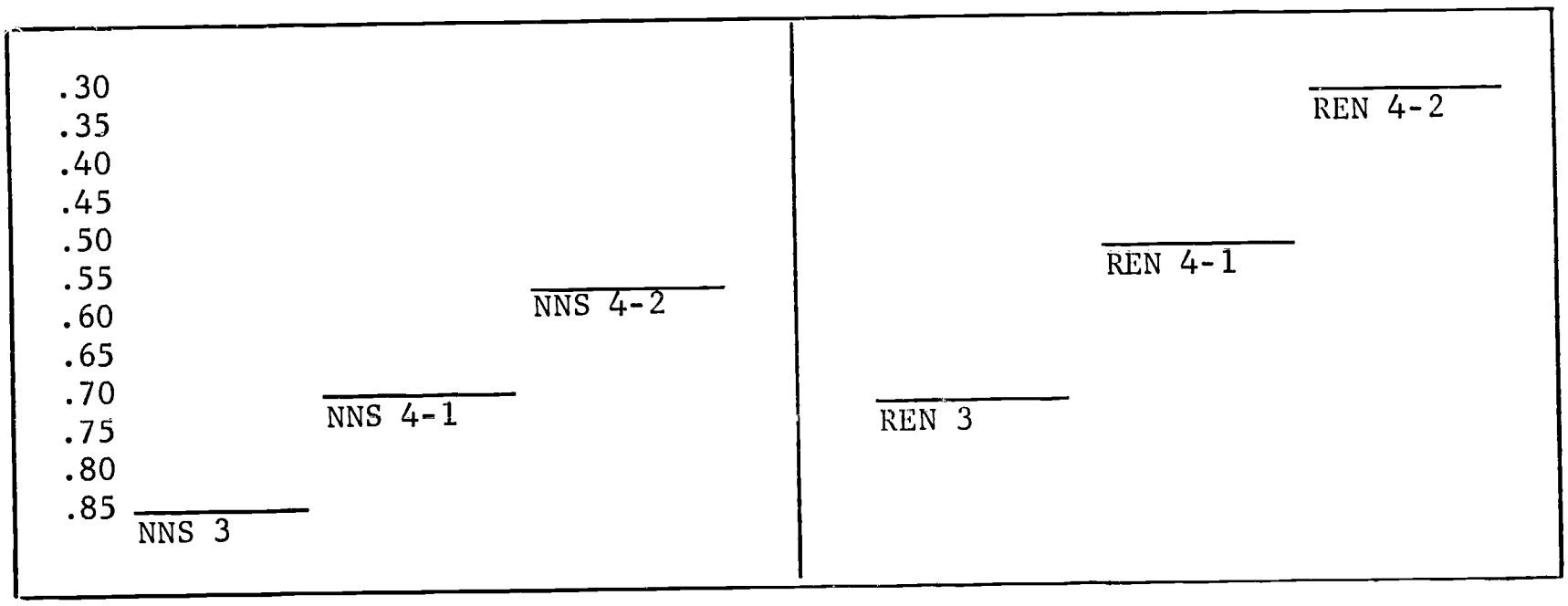
1. Considerably more research and study is needed to determine the significance of the test results, its relationship to the program at the New Nursery School and its relationship to school success.

2. At four years of age these deprived children are already a year behind in their concept formation and problem-solving ability.

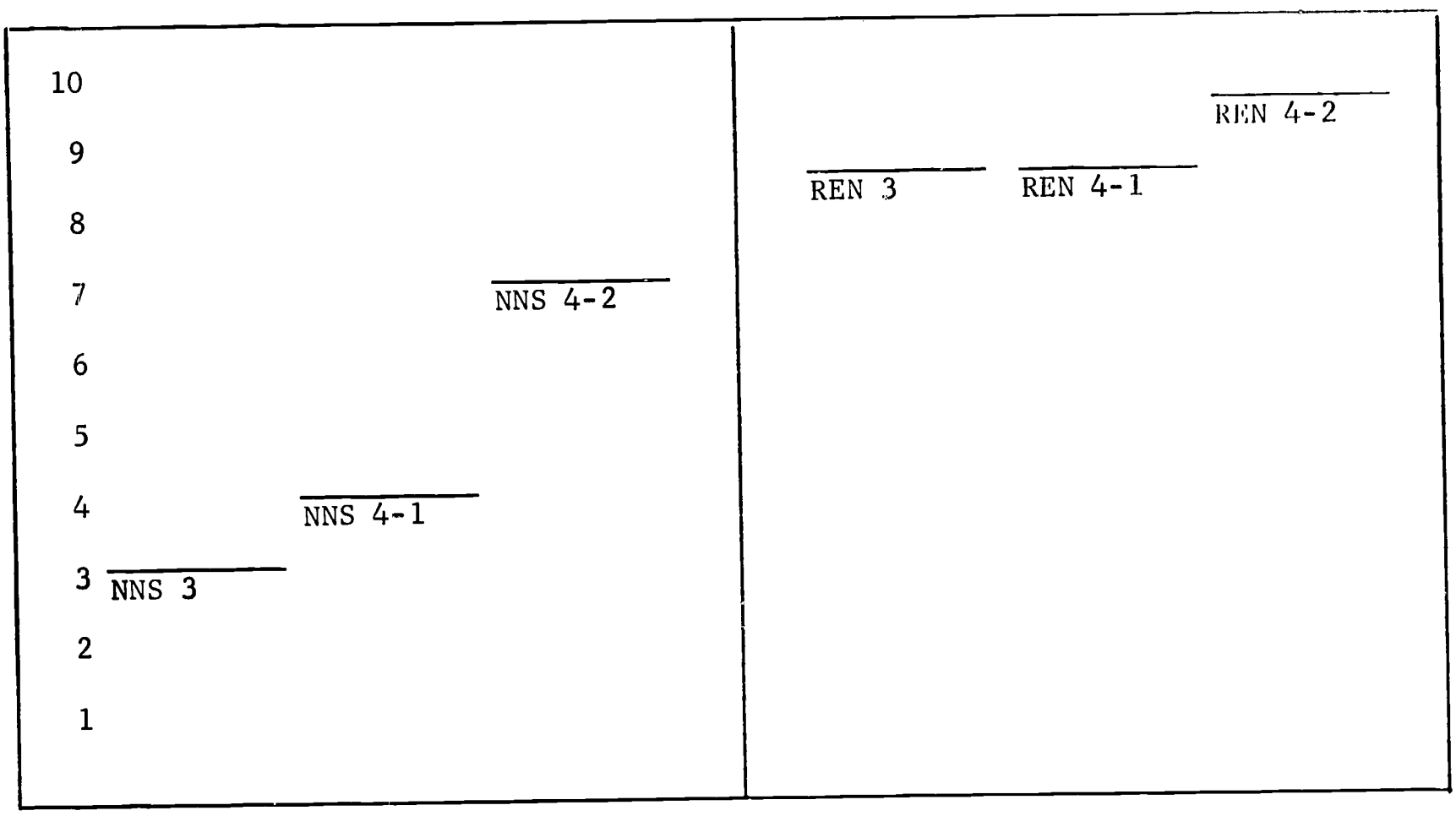
3. The four-year-olds are attending the school for the second year have made significant advances towards closing the gap between their ability and the ability of middle-class children to form concepts and solve problems. But on a more difficult problem-solving test like the Children's Categories Test, the NNS children still are testing much lower than the middle-class children. The fact that the NNS 4-2's seem to be responding more like the REN 4's and better than the NNS 4-1's who respond better than the NNS 3's, indicates that the New Nursery School is achieving some of its objectives. When one considers the fact that there is still a tested I.Q. difference of about 20 to 25 points (approximately 90 I.Q. for NNS 4-2's versus approximately 112 for the REN 4's) between the two groups of

GRAPHIC PRESENTATION OF RESULTS ON TESTS OF SENSORY DEVELOPMENT, CONCEPT FORMATION AND PROBLEM SOLVING

IMPULSE CONTROL

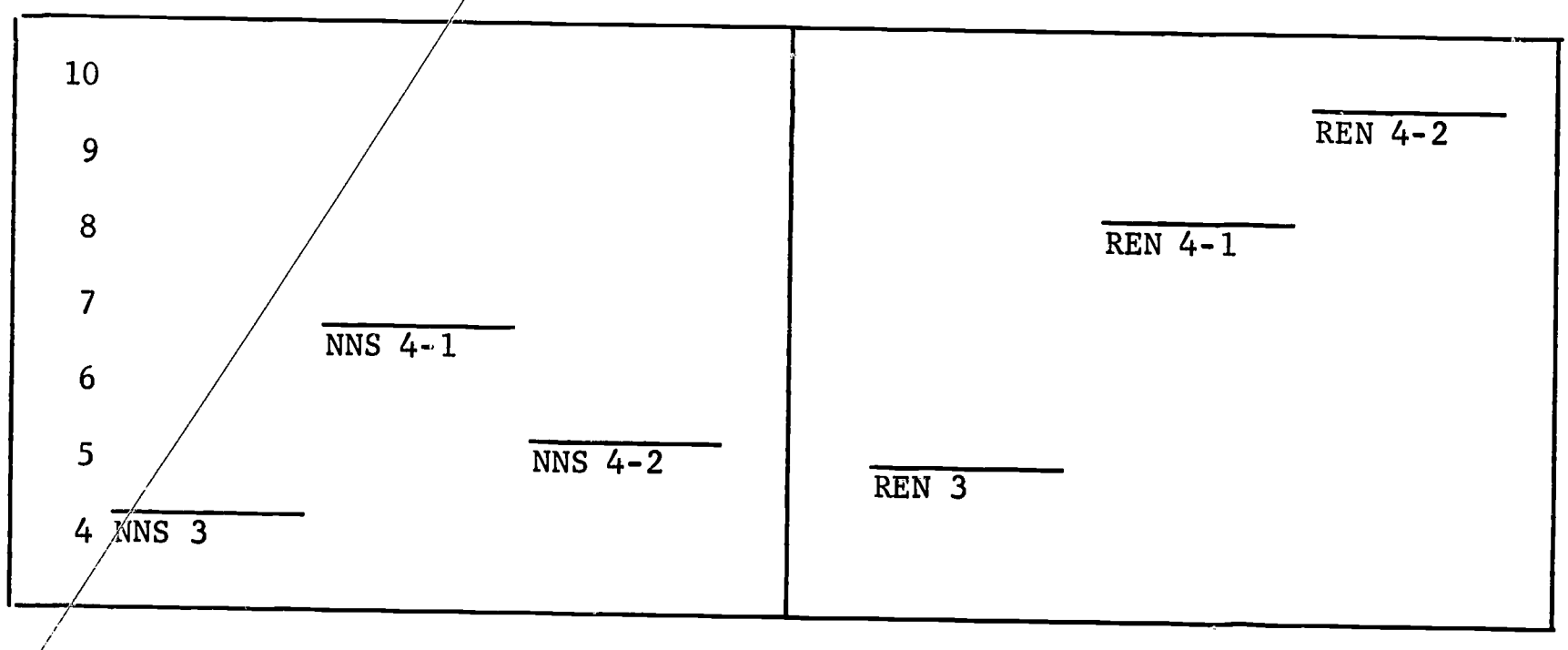


INNOVATIVE BEHAVIOR

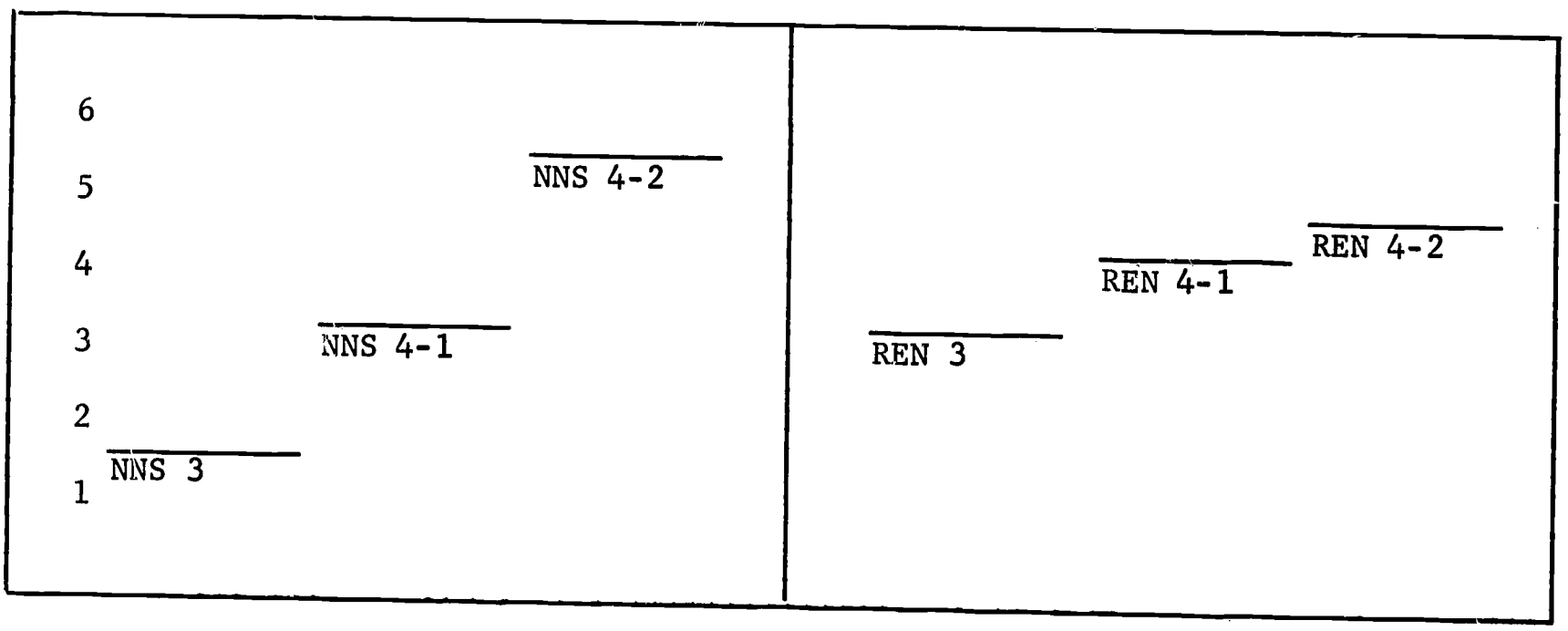


GRAPHIC PRESENTATION OF RESULTS ON TESTS (cont'd)

FIELD INDEPENDENCE

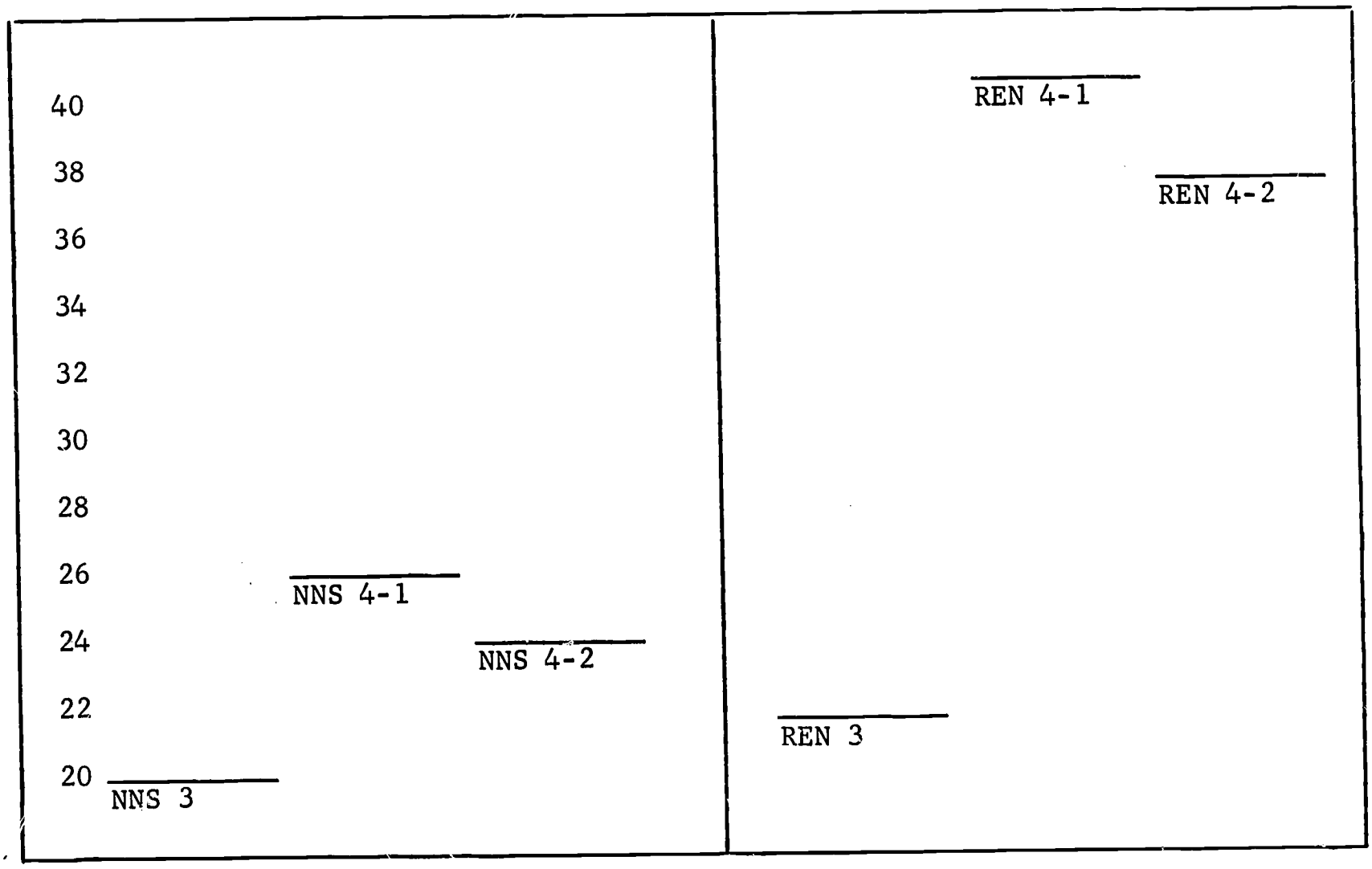


"C" TEST



GRAPHIC PRESENTATION OF RESULTS ON TESTS (cont'd)

CATEGORIES TEST



children, the results are encouraging. But we cannot assign much significance to this finding at this time because the number of cases is small and there may be some variable other than attending the nursery school that would account for the fact that the NNS 4-2's achieve better than the NNS 4-1's and in some instances as well as REN 4's. As the study progresses we can verify these findings.

4. The fact that the REN 4-2's also tend to score higher than the REN 4-1's is an indication that the program is effective not only for deprived children but for non-deprived children. The initial mean of the I.Q. scores for the two groups was approximately the same--REN 4-1, 113 and REN 4-2's 111.

5. The Innovative Behavior appears to discriminate best among the NNS groups and it may have limited value for more advanced groups. The reverse appears to be true of the Children's Categories Test and the Field Independence Test. It may be that on these two tests the younger or more deprived children do not understand what is expected of them.

6. We will continue to experiment with all five tests.

EVALUATION OF SUCCESS IN SCHOOL

Other than the tests that have been reported on in other sections of this report, we have the following information.

RESULT OF FOLLOW-UP OF CHILDREN WHO ATTENDED THE SCHOOL DURING 1964-65 AND 1965-66

Twelve NNS children entered kindergarten in 1965 after completing one year at the school. During the fifth week in kindergarten we asked the school principals to have the kindergarten teachers rank all of the children in their classrooms as (1) probably going to be successful in school, (2) probably going to have average success in school, or (3) probably going to have difficulty in school. In the teachers' judgment eight out of twelve NNS children would probably do average work, two were above average, and two would probably have difficulty in school. Since these children were in classrooms where the majority of the children were not from deprived homes, we considered this a plus in evaluating the effectiveness of the school.

The next year when the same children were in the first grade we asked the first-grade teacher to rate the same children and some of the children from the control group who were in the same classrooms. The teachers were asked to rate the child as being in the upper 10 per cent of his class, the next 20 per cent, the middle 40 per cent, the next lowest 20 per cent, or the lowest 10 per cent on reading, arithmetic, independence, attention span, good behavior, and total success in school. These ratings were made in February, 1967. The two groups were not significantly different, and the groups were judged to be at the mean or slightly below the means of the classrooms

they were in. This could mean that the positive effects of the NNS experience were washed out by the middle of the first grade, or, it could mean that the teachers were not perceptive enough to note the differences. In any event we will know considerably more at the end of this year because we will have data on former NNS children in kindergarten, the first and second grades.

Another encouraging indication is the fact that none of the NNS graduates have repeated a grade, nor have any been recommended for special education for the mentally retarded. When one considers the fact that even after making four or five attempts over a two-month period of time we could not test eighteen of these children on the Revised Stanford-Binet when they entered the program, and three other children had initial RSB I.Q. scores of 45, 57 and 60, the probability of four of five failing to the extent of being retained in the kindergarten or first grade or being considered for M-R classrooms was high.

SUMMARY AND DISCUSSION

The major criterion for evaluating the success of the New Nursery School is the performance of its graduates in school. The specific objectives of the school are:

1. Affective
   a. To develop a positive self-image

2. Cognitive
   a. To develop the senses and perceptual acuity;
   b. To develop language ability; and
   c. To develop concept formation and problem-solving ability.

To measure the development of a positive self-image we used the Brown-IDS Self-Concept Referents Test with first grade children who had attended the New Nursery School and a comparable group of children. The NNS children had fewer negative responses about themselves than the other group had, but in our judgment the test does not discriminate as well as it should. Consequently, we will continue to experiment with other measures of self-concept.

Our tentative conclusion was that the NNS program probably helps the children to develop a better self-concept in relationship to their school experience.

As an index to the persistence and degree of language and perceptual development effected in the NNS children, we cite the results on the Metropolitan Reading Readiness Test which was administered to kindergartners at the end of the 1966-67 school year. The NNS children (our graduates) who took the test were compared to a similar group of children who had not had nursery school experience. The mean score of the NNS group was at the 70th percentile and the mean score for the comparable group was at the 35th percentile. The NNS children who took this test scored considerably better than one would expect if they had not attended the NNS. This is an encouraging indicator of the initial success of the program in achieving a major objective of enabling the environmentally deprived children to perform better in regular school.

At the end of the school year 1966-67, we tested the NNS group on the Pre-School Inventory. The four-year-old children who were completing their second year at the school (NNS 4-2's) scored better than the four-year-olds who were completing their first year at the school (NNS 4-1's) and the 4-1's scored better than the three-year-olds (NNS 3's). The NNS mean total score on the test was at the 95 percentile (norms based upon children in Headstart programs); on personal-social responsiveness, the NNS mean score was at the 65th percentile; on associative vocabulary (labeling) the NNS mean score was at the 70th percentile; on concept activation, numerical (quantitative attributes of objects), the mean score was at the 80th percentile; and on concept activation, sensory (qualitative attributes of objects), the mean score was at the 95th percentile. This test was developed to give a measure of achievement in areas regarded as necessary for success in school. It involves more than language development but it depends to a large extent upon the child's ability to understand and use language. If the Pre-School Inventory proves to be a good predictor of school success, the NNS program is evidently helping children to become more successful in school.

To evaluate the development of concept formation and problem-solving ability, we have experimented with the Cincinnati Autonomy Test Battery (CATB), the "C" Test, and the Children's Categories Test.

We administered all of the tests in the CATB but found the results to be noteworthy on only the Curiosity Box, Intentional Learning, Innovative Behavior, and the Field Independence Tests. The results on the Impulse Control subtest (drawing a vertical line more and more slowly) indicate that the older children and the children who have been in the school two years have more control than younger children and those who are in school for their first year; the REN middle-class children have more control than comparable NNS children. On the Innovative Behavior Subtest the NNS 4-2's mean score was about the same as the REN children's mean score and higher than the means for the other NNS children. On the Field Independence Subtest (finding figures embedded in complex backgrounds), the NNS groups' mean scores were about the same; the REN 3's mean score was about the same as the NNS 4-2's but the other two REN groups' mean scores were significantly higher. The Curiosity Box and Intentional Learning Subtest results were mixed and non-significant.

The "C" Test is a test of a child's ability to relate objects that belong in the same category. On this test the REN 3's and 4-1's scored better than the NNS 3's and 4-1's but the NNS 4-2's scored as well as the REN 4-2's.

The Children's Categories Test requires a child to discover a concept or principle and use it to mediate the solution to subsequent non-verbal problems. On this test the mean scores of the NNS groups are similar and the mean score of the REN 3's is about the same as the NNS's but the REN 4's score significantly higher than the other groups.

The fact that the NNS 4-2's scored better than the NNS 4-1's and 3's and in some instances as well as the REN 4's indicates that the program may be helping to close the gap between the deprived children and the middle class children insofar as concept formation and problem-solving are concerned. The fact that the REN 4-2's score better than the REN 4-1's reinforces this notion and suggests that the program is also effective with middle-class children.

The combined results on the Metropolitan Reading Readiness Test, the Pre-School Inventory, and on the various experimental tests of concept formation and problem solving are encouraging. Furthermore, we have considerable case study material that supports the test findings. We will be able to place more confidence in the findings as the project continues and we are able to more precisely evaluate the strengths and weaknesses of the program.

The only negative data was the report of the first grade teachers that the children who had gone to the New Nursery School were performing no better than a comparable group of children who had not attended the school. This indicates that the effects of the program may wash out by the middle of the first grade. There is considerable evidence to support this notion. Other researchers have reported similar findings. The NNS experience for these children only represents three hours a day for about 170 days and each day they return to the same homes and environment that contributed to their deprivation in the first place. Unless the nursery school program is followed up with additional help it is plausible to reason that the effects will wash out over time. It is conceivable that the teachers' judgments are erroneous and that some of the teachers were not perceptive enough to note differences that will become evident over time. For example, if we have improved the child's concept formation and problem-solving ability and the teacher provides no experiences in which this ability can be demonstrated, the teacher may overlook the fact that the child has or has not developed the ability that will be required of him later on in school. We recognize that this may be wishful thinking on our part and will wait for verification.

Another very real possibility is that one year of nursery school experience is not enough to even start to offset the effects of severe deprivation in many of our children. Some of our data suggest that this is a reasonable notion but we can only wait for more children to enter school and for others to reach higher grades before we can answer either of these questions.