DOCUMENT RESUME

ED 026 142 PS 001 535

By-Hayweiser, Lois; And Others

Evaluating Behavioral Change During a Six-Week Pre-Kindergarten Intervention Experience. Research Project Number 5 of Project Head Start Research and Evaluation Center, Syracuse University Research Institute. Final Report, November 1, 1967.

Syracuse Univ., N.Y. Research Inst.
Spons Agency-Office of Economic Opportunity, Washington, D.C.
Report No-OEO-4010
Pub Date 1 Nov 67
Note-43p.
EDRS Price MF-$0.25 HC-$2.25
Descriptors-Analysis of Variance, Behavior Rating Scales, Cognitive Measurement, Correlation, Individual Characteristics, *Intelligence Quotient, Motivation, Preschool Children, *Preschool Evaluation, Self Control, Teacher Attitudes, *Testing Programs, *Test Interpretation
Identifiers-Adaptive Behavior Rating Scale, Draw A Line Slowly, Head Start, Perceptual Speed Test, Stanford Binet, Walk A Line Slowly, Zigler Behavior Inventory

The objective of this study was the development of evaluation techniques and the assessment of these techniques when compared with standard procedures of the national Head Start program. Assessments of cognitive behavior, social behavior, and teachers' perceptions--as opposed to aides' perceptions--of children were made. The subjects were 33 children of broad socioeconomic levels. The following conclusions were drawn from the comparative and intercorrelational analyses: (1) a meaningful proportion of the variance in Stanford-Binet performance is related to performance on the "impulsivity measures" and suggests that impulsivity has deleterious effects on children despite the degree of their cognitive ability; (2) findings with Draw-a-Line and Walk-a-Line indicate no relationship between "fast" condition and the Stanford-Binet score; (3) increases in percent work responses are not especially related to increases in Stanford-Binet scores; and (4) teachers' perceptions of children's social adaptiveness is positively correlated with intelligence. Further analyses will be reported later. Future research should concentrate on identifying dimensions of variability and then concentrate on variations in programing for individual children. (DO)

Project Head Start Research
and Evaluation Center
Contract # OEO 4010
Final Report
November 1, 1967

William J. Meyer
Center Director

Cooperating Agency
Syracuse University
Research Institute
201 Marshall Street
Syracuse, New York 13210

Evaluating Behavioral Change During a Six-Week Pre-kindergarten Intervention Experience

Lois Hayweiser
David Massari
W. J. Meyer

Evaluating Behavioral Change
During a Six-Week Pre-kindergarten Intervention Experience

Lois Hayweiser, David Massari, and W.J. Meyer[1]

This project grew out of a unique set of circumstances occurring in the City of Syracuse and Onondaga County during the summer of 1967. Essentially, there occurred a number of disagreements between the local Community Action Agency and the Office of Economic Opportunity making it unclear whether or not a Summer Head Start program would be funded. In addition it seemed apparent that the community of professional child development specialists would have minimal involvment in the Summer Head Start program. One consequence of the discord was the refusal of several communities in Onondaga County to submit proposals for Summer Head Start programs. Nevertheless, at least one of these communities was sufficiently concerned about their "culturally deprived" children that we were contacted and asked if we could help support a six-week pre-kindergarten program for approximately forty-five children. After receiving approval from the Institute for Educational Development (IED) to expend funds for such a project, and after agreement by the school officials that we could perform an extensive evaluation of the program and use the children for other research purposes, we agreed to undertake the project.

Responsibility for the content of the six-week program was given to the teachers and aides. Our primary objective was the development of evaluation techniques and the assessment of these techniques against certain standard procedures, some of which were used in the national evaluation of Project Head Start.

[1]The authors wish to express their appreciation to the entire staff of the Syracuse Evaluation and Research Center for their contributions to this project. A particular note of appreciation is due to Mrs. Jacqulina Morrow Massari who contributed many ideas to this project and who served, on a voluntary basis, in the collection of the data.

Thus, instruments such as the Zigler Behavior Inventory and the Stanford Binet were included as part of the evaluation program. Of equal concern were the changes in the children's behavior during the six-week period. Since the children continue their education in the same School District, it also will be possible to maintain a longitudinal assessment of their accomplishments.

Because of the extensive number of measures collected for each child and the consequent enormity of the data analyses, it was impossible to complete the data analysis in time for this report. The salient features of our findings are included and focus on those issues of primary concern to us. Subsequent reports will provide additional analysis.

Several considerations provided direction for the instruments employed in the evaluation. There was every reason to expect that the six-week experience would bring about significant gains in Stanford Binet performance. The reasons for the gains, and their meaning in terms of changes in intellectual competence, were unclear. Thus, it was our hypothesis that the higher I.Q.'s typically found on post-test Binet performance do not reflect change in cognitive structure. Our position, which is consistent with that proposed by Zigler (1967), is that observed increases in I.Q. reflect change in some aspects of the child's motivation. More specifically, it was assumed that children are more willing to emit responses, both verbally and motionally, because of an increased confidence in adults derived from positive encounters with adults during the intervention experience. One aspect of our assessment procedure, therefore, was to evaluate changes in the children's willingness to emit responses to the various cognitive demands of the Stanford-Binet.

Another aspect of our assessment of cognitive behavior focused on "cognitive style". Studies by Kagan, Moss, and Seigel (1963), Maccoby, Dawley, Hagan, and Deguman (1965) and Hess and Shipman (1964) clearly suggest that children approach cognitive tasks differently. Maccoby et al. (1965) examined cognitive style in terms of a lack of "motor inhibition" and reports a positive correlation (r = .41) between her task, Draw A Line Slowly, and Stanford-Binet I.Q. It would appear that a lack of "motor inhibition", or "impulsivity", is a general style pervading cognitive behavior and which may also influence social behavior.

In descriptive terms, the impulsive child seems to respond to stimulus situations at a too rapid rate in order to process the information presented . Components of the stimuli may be missed altogether or responses may be made only to the most dominant or familiar features of the stimuli. There is the additional possibility that the child does not even consider the stimulus characteristics of the problem but responds on some totally irrelevant basis unique to the child. Impulsive children, whether the antecedents are biological, environmental, or a combination of both, find complex tasks, which require processing time, more difficult. Thus, lower performance on a measure of intelligence such as the Binet would be anticipated.

The inability to inhibit responses should also negatively influence how teachers view the adequacy of the social behaviors of impulsive children. The failure to "hear" directions, excessive wiggling at the desk, or running in the hallway, are behaviors generally viewed as negative by teachers. It is just such behaviors, however, which apparently characterize the "impulsive" child.

A second broadly defined area of the evaluation focused on the adequacy of the children's social behavior. Of concern here is the relationship between the teacher's perception of the adequacy of the child's social behavior and Binet gain.

Our assumption was that children who conform more to the demands of the classroom situation are preferred by teachers and they therefore provide them with a more conducive learning environment. Rosenthal (1967), for example, shows that the teacher's initial perception of her pupils influences the amount of gain reported. A scale concerned with children's adaptive classroom behaviors was developed, in collaboration with the teachers, and administered on a pre- and post-test basis. At no time were the teachers, or the aides, aware of the child's Binet I.Q. so that all ratings can be considered independent of any obvious biasing effect.

A third consideration in designing this project was an examination of the teachers' as opposed to the aides' perceptions of the children. One might speculate that teachers view children differently from aides because of their more academic orientation. Aides, it might be presumed, accept this role for a variety of motives, one of which is most likely to be that they enjoy children. One might anticipate, therefore, that the aides' perception of children is less influenced by their estimate of the child's intellectual ability, or by his level of achievement. An alternative reason for expecting differences derives from the fact that aides generally are not as experienced in judging groups of children in a classroom situation as are the teachers. Since the three aides for this project were sophomore or junior students at state teacher colleges in upstate New York it would be expected that any differences would more likely be based not so much upon their differences in academic orientation as their lack of an experiential standard for a classroom type situation against which to compare the children.

A final consideration in the design of this project was to provide an assessment of the effects of a six-week intervention experience in comparison with children from the same population not having the experience. In this connection, more

children were invited to participate in the program than we anticipated would finally attend. Unfortunately, in some respects, this procedure worked better than we had hoped for in that more children than anticipated either failed to attend altogether or attended for only one or two days. Fortunately, it was possible to assess the general level of intellectual functioning of these children who had been invited to attend the program but who, for a variety of reasons, did not do so.

In summary it can be stated that our major focus was on social cognitive components related to behavioral change. Although our interest was focused somewhat on demonstrating change as a function of the intervention experience, our primary concern was with the relationship of cognitive, social, and teacher-aide variables and observed behavioral change.

Method

Subjects: The subject pool from which the sample was drawn was defined by the Director of Elementary Education (Mr. Paul Anderson) of the Liverpool School District. This school district includes a broad range of socio-economic levels. The sample identified met the poverty criterion with respect to family income and were further known to the school district officials either through social service agencies or through prior encounters with the family because of problems arising with older children in the family. Of the approximately 85 children identified, 45 children were selected as being in greatest need for a pre-kindergarten program. The parents of all 45 of these children were personally contacted by the school officials and the program described to them. A total of 33 of the invited children finally attended the pre-school. Characteristics of these children are summarized in Table 1 from which it can be noted that their pre-test Binet scores are not

Table 1

Characteristics of Sample

	CA(Mos.)	Binet	Zigler
M	64.8	90.0	138.5
SD	6.9	16.8	20.6

unlike those typically reported in other Head Start classes. The somewhat higher range of chronological ages reflects the inclusion of two children who had been denied entrance to the first grade after a kindergarten program on the grounds that they were insufficiently ready for that experience.

Staff

The three teachers were comprised of two females and one male. Prior teaching experience was 3, 7, and 10 years. Each of the three aides, all females, were enrolled in college and were in teacher preparation programs.

All Stanford-Binets were administered by Examiners with considerable experience in administering the Binet to young children. Other cognitive assessments and classroom observations were conducted by a core of 8 advanced level graduate students.

Instruments

A. Cognitive Measures.

The Stanford-Binet, Form LM, was administered on a pre- and post-test basis. This test, rather than some other measure, was selected simply because it has been used in other similar studies making comparisons with our data easier. Each of the Examiners gained extensive experience with the instrument and the age level because of their having examined the children in our Head Start Evaluation sample. Children were randomly assigned to Examiners. In order to maximize reliability within measurements, the same Examiner was used for each child on the pre- and post-tests. Independence between measures was assured by using different Examiners for the different measures. Standard manual procedures were followed.

The second measure focused on the children's responses to the cognitive demands imposed by the Stanford-Binet. More specifically, children's responses were rated by the Examiner during the testing, in terms of "work" or "non work" following a

procedure adapted from Birch (1967). The scales are: Work-No Work Categories (Adapted from Hertzig, Birch, Thomas, and Mendez, 1967).

A. Work Responses (child must attempt what was asked of him)

1. Work Verbal Spontaneous Extension -- S spontaneously extends his response. For example, in completing the analogy "birds fly and fish ____" the child says "bite-like one once bit me on the leg."

2. Work Verbal Delimited -- S's response is restricted to defined requirements of task.

3. Work Non-Verbal Spontaneous Extension -- S responds to demands of task and then spontaneously performs some additional behavior. For example, child strings beads and then places around her neck.

4. Work Non-Verbal Delimited -- S responds strictly to the demands of the task.

B. Non-Work Responses (Child does not work at task)

1. Non-Work Verbal Competence -- S refuses to attempt task stating a lack of competence to handle task. For example, "I am too young for that."

2. Non-Work Verbal Negation -- S verbally refuses to work. For example, "I won't."

3. Non-Work Verbal Substitution -- S offers an irrelevant verbalization such as "I want to leave now," or "I want to play with the toys."

4. Non-Work Verbal Aid -- S verbally requests aid from examiner.

5. Non-Work Non-Verbal Negation -- S refuses to work by indicating with a motor response. For example, head shake, pushing materials away, or turning away from Examiner.

6. Non-Work Non-Verbal Substitution -- S engages in an irrelevant motor response. For example, when asked to build a bridge, the child walks away

or pushes the blocks back and forth on table.

7. Non-Work Non-Verbal Passive -- S does not respond in any way to task or Examiner.

Prior to testing the sample children, each Examiner administered a minimum of three Binets to preschool age children with two other raters present. Observer reliability was estimated in terms of the per cent of instances where all three observers were in agreement with respect to the 11 categories. The median agreement was approximately 74%. Agreement in terms of the broad categories of "work" "non-work" was approximately 90%. On the basis of these data, it was felt that the Examiners could rate the "work" "non-work" responses of the children while administering the Binet test.

The "work" "non-work" ratings were included to measure change in children's responses to cognitive demands under, at least, somewhat stressful conditions. Partially, these measures should reflect the child's willingness to attempt responses where the probability of failure is high. Assuming that the child responds correctly to some items that were, in various ways, avoided on the pre-test, it may be possible to account for gains in Binet I.Q. The "work" "non-work" measure may also reflect a child's "style" in approaching difficult tasks. Thus rather than attack a difficult task, a child may prefer either not to respond at all or respond in a completely irrelevant fashion. Finally, the "work" "non-work" measure may also reflect a motivation to, in a sense, please the Examiner.

Several impulsivity measures were used in this study. The "Draw a Line Slowly" (DAL) Test, reported by Maccoby, et al. (1965) and Hess, et al. (1966) was used with certain modifications. This task requires the child to draw a line, beginning at the top of a plain 8½ x 11" piece of paper and proceeding to the bottom of the page, as slowly as possible.

The following instructions were used: "I am going to draw a line on this paper as fast as I can. I will start here at the top and go to the bottom. Now you try it, take the pencil and go as fast as you can to the bottom. Very good. Now, this time I am going to draw a line from the top to the bottom of the page as slow as I can. Watch. Now you try it. Draw the line as slow as you can. That was very good." These were practice trials and to show and explain to the children the meanings of "slow" and "fast". The children were then given the test trials in the following sequence of conditions: 1. "Draw the line slow"using an 8½" x 11" piece of paper with no markings on it. Here the direction was to draw the line as slow as possible starting at the top and going to the bottom. 2. "Draw the line slow" using an 8½" x 11" piece of paper on which there was an X at the top and bottom. The child was told to draw the line "even more"slowly and to connect the Xs. 3. This condition repeated number two. 4. "Draw the line fast" using an 8½" x 11" piece of paper with no markings. 5. "Draw the line fast" and connect the X at the top to the X at the bottom.

The second impulsivity measure was the Walk a Line Slowly Test (WAL). In this task two six feet long parallel lines of adhesive tape, five inches apart, were placed on the floor. The child was instructed to place one foot on each tape. The task was given under three conditions, in the following order: 1. "I want you to walk to the end of the tape making sure you do not step off the lines." No instructions concerning speed were given. 2. "Now I want you to walk the line as slow as you can." 3. "Now I want you to walk the line as fast as you can."

Two measures, not previously reported by other investigators, were used. One measure was an adaptation of the "Perceptual Speed Test"(PST) subtest from the Primary Mental Abilities Test, Primary I Battery. This test requires the child to match a standard against four alternatives. In addition to using the standard

scoring system of recording number of correct responses, a latency to first response measure was used and a correction procedure in which the children continued responding to each item until making a correct choice. A short latency was assumed to indicate a lack of careful analysis of the response alternatives (One problem with this assumption is that a very bright child might be able to rapidly process the alternatives giving the impression of impulsive behavior.) In terms of responses-to-correct-choice measure, it was assumed that a child requiring many responses over the test items is responding more or less at random without considering the pertinent components of the stimuli.

The PST was administered following the verbal directions given in the manual. For the correction procedure, the children were told: "No, that is not the correct one, try again." "That one is wrong too, try again," and so forth. The more difficult items were (those appearing on page 13 of the test booklet) in order to reduce familiarity with the stimuli.

The final cognitive measure involved a modification of two WPPSI mazes, numbers 2A and 3A. Each maze was administered as it appears on the test and then with the blind alleys deleted. Following a practice maze, (Maze 1A from the WPPSI) each subject was given the test mazes in the following sequence: 1. no cul-de-sac, 2. easy maze with cul-de-sac, 3. no cul-de-sac, but longer than maze 1, and 4. harder maze with cul-de-sac. This task was designed to measure impulsivity with a more difficult paper-and-pencil task than was involved with the DAL. Performance was assessed in terms of errors, defined as crossing lines and/or entering a cul-de-sac (mazes 2 and 4, only), and elapsed time to complete the maze.

Social Behavior Measures

The main objective of the instruments described in this section is to provide data relevant to the adequacy of the children's social behavior as perceived by the teachers and aides. Although there was concern for behavioral change, our analysis also focused on the correlates of perceived social adequacy.

Since teacher perceptions probably effect, perhaps to a significant degree, the school environment of the child, it seemed reasonable to ask teachers to describe the behaviors influencing these perceptions. Each of our three teachers were asked, in separate interviews, to describe a maximally adapted and maximally maladapted kindergarten child. All statements were probed until descriptions were given in reasonably objective behavioral terms. For example, the statement "is well behaved" after probing resulted in statements such as "does not grab", "waits in line and stays in his place", or "waits for directions before rushing in". A total of 62 such statements comprise the Adaptive Behavior Rating Scale (ABRS), (see Appendix A for a copy of the scale.)

The teachers and aides completed the scale for each of 33 children. Both pre- and post-test measures were obtained. An index of internal consistency was obtained using a procedure described by Flanagan (1965) resulting in a reliability estimate of .86.

A second measure of teacher perception involved a man-to-man rating technique adapted from procedures described by Gardner and Thompson (1956). Essentially, this technique involves naming a child, from some population which includes the group to be rated but which is larger than that group, who has the most and least of the attribute to be subsequently rated. A third point is identified in terms of the most average child. All members of the group are then rated in terms of this

reference population. Separate reference scales are established for each rated attribute. Each rater defines the children for his scale providing, at least theoretically, scales across raters that have psychological equivalence.

Two hypothetical situations were constructed and rated by both teachers and aides. Situation 1 was: "Suppose you were asked to select children for a special kindergarten class where academic superiority was important." Situation 2 was: "Suppose you were asked to select children for a kindergarten class in terms of their qualities as children that teachers enjoy." Ratings were made on a five point scale.

A third instrument used to assess social-emotional behavior was the Zigler Behavior Inventory. This instrument was employed in the national evaluation of Project Head Start and its properties are generally well known.

Finally, our assessment of social behavior included a very limited number of observations on approximately one-half of the sample. These children were randomly selected prior to the first day of classes and were observed for ten randomly selected one-minute periods over two mornings. This procedure was followed during the first and last sixth weeks of the program.

Categories of behavior were broadly defined and include the following:

1. Conformity-non-conformity: Observers judgment that child was doing what was expected, or not, during observation time;

2. Non-verbal communications: Child points, shoves, signals;

3. Incomplete verbal communications: Child uses one word or fragmented sentences;

4. Complete verbal communications: Child uses complete sentences or phrases;

5. Attentional changes: Child shifts attention from one task to another,

shifts from one situation to another (person entering room).

Observer agreement was determined by having the three observers observe and simultaneously code the behavior of similar aged children in one of the city preschools. These practice sessions were done during both structured and free play situations for a number of children. Practice was continued until the raters achieved 90-100% agreement for the number of responses in the categories of conformity-nonconformity, verbal vs. non-verbal and 70-80% agreement for attentional changes and incomplete vs. complete verbal communications.

Classroom and Home Stimulation

Two measures of input were secured on each child in the sample. In an effort to assess the level of stimulation of the child's home, the teachers accompanied by an aide or research assistant, made home visits using the Inventory of Home Stimulation (STIM) developed by Caldwell (1967). This instrument assesses such variables as orderliness, number and kinds of play materials and verbal interactions of the mother. A copy of our modification can be found in Appendix A.

The classroom environment was assessed in terms of the teachers and aides use of "praise" and "blame". We were especially interested in assessing the initial impact of the teachers on the children so that 30 minutes of observations were made for each teacher and aide over the first hour of the first day of classes. Teachers and aides were observed in alternate periods of 10 minutes each. Praise, blame, object and situation content were recorded. An additional one-half hour was obtained for both teachers and aides on one additional day during the first week and on one day during the last week of the program. Further observations were made of both teacher and aide during a field trip to a farm and during one-half hour of the lunch period.

Observer agreement was obtained by having the observers simultaneously rate two teachers in a city nursery school for two five minute periods. Observer agreement was between 90 and 100% for praise or blame, the object and essential content of the situation.

Results

The results are organized in three general categories: 1. behavioral changes during the six-week program with three subsections consisting of cognitive changes, changes in social behaviors, and changes in teacher behaviors; 2. the interrelationships among the variables; and 3. a summary of two special projects within the overall program concerned with arithmetic skills.

Although the various indices reflect specific theoretical biases, it should be made clear that this is an exploratory study. The Syracuse Center is currently attempting to cross-validate some of the more promising relationships and hope that other investigators may wish to work with some of the measures. Until such validating studies are conducted the measures should not be used for other than research purposes.

One additional comment may serve to reduce reader confusion. Obviously our measures were obtained over several days and were scheduled with phenomenal precision. An absent child often could not be rescheduled so that the N varies somewhat over analysis. Most of the change score and correlational analyses are based on N's of between 29 and 33 children. All instances of an N below 29 will be specifically noted.

A. Effects of Intervention Program

1. Cognitive Behaviors.

All analyses in this section, unless otherwise noted, consist of correlated *t* tests following standard procedures. Summarized in Table 2 are the means, SDs,

Table 2
Summary of Pre-Post Means, SDs and *t* tests

	Stanford-Binet		PST (Responses)		PST (Latency)		DAL Rates	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
M	90.0	96.1	18.3	15.7	44.3	44.8	2.0	2.2
SD	16.8	19.7	4.0	4.1	12.5	22.2	1.9	1.9
	t = 6.01**		t = 3.96**		t = <1		t = <1	

	Mazes (Time Scores)		Mazes Errors		WAL		Teacher Academic Ratings	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
M	60.4	50.7	7.2	5.4	10.5	11.0	1.1	2.6
SD	17.1	12.7	4.5	4.4	6.7	4.6	.8	1.2
	t = 2.5*		t = 1.8		t = <1		t = <1	
							df = 30	

	Aides Academic Ratings	
	Pre	Post
M	2.7	2.6
SD	1.2	1.4
	t = <1	

** $p < .01$
* $p < .05$

and ts for each cognitive measure. The average gain in Binet I.Q. was 6.1 and is statistically significant. This average amount of gain is certainly not spectacular but is not inconsistent with gains reported in other intervention studies. Consistency and relative position between pre- and post-test is remarkably high as reflected in a product moment correlation of .98. As anticipated, there is a significant relationship ($r = .52$) between pre-test IQ and change score showing that the brighter children made the greatest IQ gains.

Performance on the Perceptual Speed Test (PST) also improved significantly over the six-week period. It should be noted that improvement on this measure is denoted by a reduction in the average score which reflects, because of the correction procedure used, a reduction in the number of responses made to each item before selecting the correct response. This result suggests that the children were somewhat more careful in making their choices on the post-test, but it cannot be argued that they were necessarily taking more time in their responses. This latter argument is supported by the very slight gain in latency to score in making their initial choice on this task. Relative position with respect to number of responses made on the pre- and post-tests has a moderately high relationship as reflected by a product moment correlation of .68. The relationship between pre-test performance and change score is lower than that observed for the Stanford-Binet ($r = .39$), but statistically significant. The lower correlation probably reflects a ceiling effect in that children obtaining an initially low score could only improve by one or two points. Again, the correlation suggests that the intervention experience is more profitable for those children with initially high scores. The test-retest correlation for the latency measure on the PST indicates essentially no relationship between the two measures ($r = -.14$, $df = 27$, $p = > .05$). Undoubtedly this correlation reflects the

Table 3
Work Non-Work Behaviors

Total Performance

Work Verbal Items		Work Performance Items	
Pre	Post	Pre	Post
83.1%	85.0%	92.8%	97.0%
$t = .90$ using arc sin trans		$t = 2.8$ using arc sine	

Ceiling Year Items

Work Verbal Items		Work Performance Items	
Pre	Post	Pre	Post
69%	74%	75%	96%
$T = 117$ (Wilcoxon Matched Pairs)		$T = 0$	

general problem of obtaining reliable latency measures, but as will be seen in the next section of this report the post-test latency measure, but not the pre-test latency measure, is related to performance on the Stanford-Binet.

Analyses of the Draw-A-Line Slowly (DAL) Test is based on a rate measure; that is, length of line divided by time to draw the line. (Analyses involving the straight latency measure yield identical results.) Thus in this measure a high-score is indicative of impulsivity. Scores on the three trials were highly inter-correlated (median r = .89) therefore scores for each child were pooled. It will be noted from Table 2 that the change in performance from pre- to post-test was not statistically significant. There was, however, considerable consistency in relative position on the pre- and post-test as reflected in a product moment correlation of .84. The correlation between pre-test performance and change score was .29 again indicating a tendency, which is not statistically significant, for greater improvement to occur among the initially better performers on this measure. That this correlation is low may reflect a ceiling effect in that initially low performance by some of the children could just not be improved upon. It is noteworthy, however, that among the high impulsive children there was apparently little general gain. A similar result was found on the Walk-A-Line Slowly Test where again the difference between the pre- and post-test measures was not statistically significant. The measure of consistency for the WAL was lower (but statistically significant) than that found for the DAL (r = .41). There is no apparent reason for this difference unless, in fact, the WAL is a less reliable measure. The correlation between pre-test performance on the WAL and difference score is negatively related (r = -.55) indicating that initially slow performance is associated with slower (less impulsive) performance on the post-test.

Performance on the WPPSI mazes in terms of pre- and post-performance indicated a significant increase in the speed with which the tasks were performed but there was no significant decrease in the number of errors (the decline in the number of errors approached statistical significance, $t = 1.8$; $df = 18$; $p = < .08$). These results suggest that when children are not given specific instructions to perform a task slowly, there is a tendency, at least with respect to the maze task, to speed up performance. The correlation between error scores on the pre- and post-tests was $r = .50$ and the relationship between pre-test score and change score was $r = .49$. With respect to the latency measure for the mazes, the correlation between pre- and post-test performance was .33, which for 19 df is not statistically significant, and the correlation between pre-test performance and change score was $r = -.73$.

One of the primary concerns in this project is the teachers perception of the children. Recall that we administered an adaptation of the Syracuse Scales of Social Relations asking each teacher and aide to rate each child with respect to academic capability. Examination of the results of this procedure, shown in Table 2, indicate that neither the teachers nor the aides felt that there was significant change in the youngsters over the six-week period. It should be noted that, on the average, both the teachers and the aides rated the children to a considerable degree at the low end of the scale. As anticipated, there was a significant tendency for the aides to rate the children higher on the pre-test than did the teachers ($t = 7.8$; $df = 66$; $p = < .01$). This difference disappears on the post-test with teachers and aides rating the children approximately the same. Both teachers and aides tend to rank the children with considerable consistency from the pre- to the post- measure; $r = .68$ for teachers and .74 for aides. Since the predominant change score was zero, it was not feasible to run a correlation between pre-test score and change

score. In order to determine if any relationship exists between pre-test score and change score, a 3 x 3 contingency table was established consisting of a "gain," "no change," "loss" categories for the change score, and "above 3", "below 3" and "3" as pre-test ratings. The results of this analysis indicates no significant change on the part of the teacher ($X^2 = 3.14$; $\underline{df} = 4$; $\underline{p} = > .05$) but a significant change for the aides ($X^2 = 10.6$; $\underline{df} = 4$; $\underline{p} = < .05$). These data suggest then that the aides were more likely to shift their perceptions of the children and that their initial impressions of the children were less related to the change score.

Table 3 summarizes the results of the work vs. non-work responses to the items on the Stanford-Binet. These data are reported in terms of percentages of work responses to total responses and are summarized in terms of responses made to all of the items administered to a particular child and, in the bottom half of the table, the work responses made during the administration of the ceiling year tests for each individual child. Consistent with the findings of Hertzig et al, there is a greater tendency to emit work responses with respect to the performance items on the Stanford-Binet as opposed to the verbal items. It should also be noted that, as anticipated, there are somewhat fewer work responses during the ceiling year than on overall performance. Comparisons of pre- and post- performance indicate that there was no significant increase in the percentage of work responses for the verbal items of the Stanford-Binet, either in overall performance or with respect to ceiling year items. There was, however, a significant increase in the percentage of work responses to the performance items on both total performance and ceiling year items. Since the general trend for both verbal and performance items is in the direction of a greater emission of work responses, it seems reasonable to conclude that post-test performance on the Stanford-Binet is characterized by a greater

tendency to attempt items rather than to avoid the demands of the situation. It is quite conceivable that the average gain in Binet I.Q. (6.1) is attributable to the greater number of task relevant responses made by the children on the post-test. In an effort to test this assumption, a correlation was run between Binet change score and the change score for work responses. This analysis resulted in a product moment correlation of zero indicating that a change in per cent work responses is not related to Binet change score. Our original formulation appears to have been oversimplified. Whereas we had anticipated that an increase in work responses would, even on a chance basis, lead to a gain in I.Q. score, a post hoc examination of the data indicates that this was not the case among the low pre-test Binet children; that is, their per cent work responses increased but they did not get more Binet items correct. Among the high pre-test Binet children, there was little gain in work responses, relative to the total group, but they made better use of their additional work responses than the low pre-Binet children. A similar case can be made for those children with average pre-test Binet's. Thus, it appears that an increase in work responses is more beneficial to children of average or above average ability than to children of low ability.

II. Social Behaviors

It will be recalled that four measures of social behavior were used: the Adaptive Behavior Rating Scale (ABRS), the Zigler Behavior Inventory, an adaptation of the Syracuse Scales of Social Relations with respect to personal attributes, and direct observations of the children's verbal behavior, conforming behavior, and attentional behavior.

With respect to the observations, we were not able, unfortunately, to secure enough data to warrant analyses in terms of pre-post measures. This occurred because we

were only allotting ten minutes of observation time to a subject and, frequently, children were absent on the days that observations were being run.

Table 4 summarizes the three measures for which we have analyzed the data. All t analyses are based on correlated means following standard procedures. It will be observed that the only significant changes with respect to our measures of social behavior occurred for the ABRS. These changes are in the direction of higher ratings on the post-tests and suggest that the children developed more adaptive behaviors, as defined by the teachers. Since this is the first time the ABRS has been administered anywhere, it is not possible to make comparative statements about this sample. However, to put the mean score in some perspective it might help to know that the maximum score possible is 310 (5 points is allotted to the "always" category and 1 point to the "never" category) and the minimun score possible is 62. If all of the children had received a scale score of 3 ("sometimes") on all of the items their score would have been 186. Thus, it appears that on the pre-test the teachers viewed the children as performing the behaviors on the average, somewhere between the categories of "sometimes" and "most of the time". The aides, it will be noted, rated the children lower; that is, somewhere between "once in a while" and"sometimes". On the post-test, the teachers ratings shifted in the direction of the category "most of the time" and this change is statistically significant. The aides, who show a greater degree of change, place the children, on the average, in the same relative position as the teachers. In terms of consistency of ratings over the intervention period, the product moment r for teachers was .55 (df = 31; $p < .01$) and for aides .31 which with 31 df is not statistically significant. The correlation between the teachers' pre-test ratings and change score was .57 (df = 31; $p < .01$) the similar correlation for aides was .31 degrees of freedom is not statistically significant. It is clear from these correlations that the teachers'

Table 4
ADAPTIVE BEHAVIOR

Inventory

	Teachers			Aides		
	M	SD	N	M	SD	N
Pre	196.7	33.6	33	183.4	31.2	33
Post	213.8	50.8	33	213.1	51.4	33
	t = 5.4**			t = 7.6**		

Zigler Behavior Inventory

	Teachers			Aides		
	M	SD	N	M	SD	N
Pre	138.5	20.6	34	139.7	21.9	32
Post	134.5	20.8	31	137.9	23.2	32
	t = 1.7			t = 1.2		

Personal Attributes

	Teachers		Aides	
	M	SD	M	SD
Pre	3.09	.91	3.00	1.29
Post	3.10	1.06	3.03	1.47
	t = < 1		t = < 1	

** p = < .01

post-test score and consequent change score is more closely related to their initial rating of the children than is the case for the aides.

As will be noted in the next section of this report, the ABRS is significantly related to many of our cognitive measures suggesting that the instrument is measuring behavior of significance to the child's experience in the intervention program. As a measure of validity, we correlated the ABRS post-test scores with the Zigler Inventory post-test scores, resulting in a product moment r of .59 (df = 27; p = < 01). (We did not run a similar correlation with the pre-test scores because of some pecularities observed on the pre-test scores for the Zigler.) It is clear that the Zigler and the ABRS share a significant portion of variance, but also, obviously, there is a good deal of unique variance between the two instruments. Unfortunately for the purposes of this report, we were unable to complete an item factor analysis of the ABRS, and are thus unable to report the behavioral categories reflected in our items. We anticipate completing this analysis soon making it possible to compare ABRS behavioral categories with those being identified by other investigators using the Zigler items. There are plans to readminister the ABRS by having the children's current teachers rate them as an estimate of between-teacher consistency. There are also plans to cross-validate the instrument using another nursery school sample.

The data summarized in Table 4 for the Zigler Behavior Inventory indicates a decline in ratings by both teachers and aides but the difference is not statistically significant. One possible reason for the lower post-test scores may be attributable to the difficulty reported by the teachers and aides in making the judgments required on the Zigler so early (first week) in the intervention program. They also reported difficulty in understanding the more "global" items. Despite

these comments it should be noted that the correlation between the pre-test ratings and post-test ratings for the teachers was .80 (df = 31; p = <.01) and a similar correlation for the aides was .72 (df = 31; p <.01). Among our teachers and aides, therefore, there was a tendency to rate the children slightly lower on the post-test with relatively little shift in the rank ordering of the children.

The third set of data in Table 4 relates to the 'personal' attributes ratings using the Syracuse Scales of Social Relations. 'Personal' attributes, as defined here, refers to the general social attractiveness of the children and not to any particular attribute of social behavior. Recall that these ratings are made on a 5-point scale where the raters defined the two end points and the mid point in terms of all the kindergarten children they had ever known. The resulting data reflect the relative position of our sample of children with respect to the reference populations defined by the teachers and aides. Examination of the means for both teachers and aides indicates that the average child in the sample was placed in the almost exact middle of the scale. This is in contrast to the very much lower ratings given to the children on the 'academic' attributes scaling. This discrepancy suggests that despite the low academic attributes of the children, their personal attributes are seen as average. As shown in Table 4, there were no statistically significant changes between pre- and post- ratings for either teachers or aides on the 'personal' attributes scale. The correlation between the initial rating and the post-rating was .47 (df = 31; p <.05) for teachers and .47 (df = 31; p <.05) for aides on the personal attributes scale. In contrast, the identical correlations for the academic scale ratings were substantially higher (.68 and .74 for teachers and aides respectively). Thus, for both groups of raters there is a greater relationship between pre- and post-test scores for the academic attributes as opposed to the personal attributes.

III. Input Measures

There were two measures of stimulus input used in this study: 1. the inventory of home stimulation, and 2. observations of the teacher and aide behavior in the classroom.

Table 5 summarizes the means and SD's for both measures. Because of the relatively little interaction occurring between the children and the teachers and aides it was decided to pool their praise and blame statements to each of the children. Despite the few observations that were recorded, it can be seen that the teachers and aides were tending to use more blame in their approach to the children than praise. The difference between the praise and blame received by the children is statistically significant ($t = 2.2$; $df = 31$; $p = <.05$). It should also be noted that the variation in blame received is greater than for praise received. This occurred because three or four of the children received a comparatively large number of the blame contacts from the teachers. These results are not unlike those reported by others (Meyer and Thompson, 1956) and suggest that some children learn very early to expect disapproval from school personnel.

The mean and standard deviation for the STIM is also reported in Table 5. Since there are no norms available for this version of the STIM, it is not possible to comment in any intelligent way about the meaning of the mean and the standard deviation. Instead, we will use this measure in the next section of this report that discusses the inter-relationships among the measures.

IV Inter-relationships of Measures

This section of the report is concerned with the inter-correlations among the variables used in evaluating this intervention program. In view of the fact that the data were not available for analysis until mid-September, it will only be

Table 5
Means and SDs for
Stimulus Input Measures

	Praise (Teacher and Aide)	Blame (Teacher and Aide)
M	2.03	3.44
SD	1.68	3.70

STIM
M = 46.4
SD = 10.4

possible to report a few of our findings. As further analyses become available, they will be included in subsequent quarterly reports and other written materials. One other comment is necessary before examining the data. We are only too well aware that correlational data are subject to a large number of variables that may have occurred fortuitously as a result of sampling peculiarities among the teachers and/or among the children. In reading this material, much of which strikes us as exciting, it should always be remembered that the data are derived from one sample of children whose totality of characteristics are largely unknown to us.

Recall that the Draw-A-Line Slowly and Walk-A-Line Slowly Tests were designed to measure "impulsivity". It was assumed that impulsive children are less likely to attend to the stimulus characteristics in their environments and are more likely to respond to specific characteristics or components of the stimulus situations without adequate analyses of the total stimulus situation. If the DAL and WAL reflect the described cognitive style there should be a significant relationship between these measures and the Binet. Impulsivity should also be related to the adequacy of social behavior in the classroom. Thus, the impulsive child may talk too much, may become involved in disputes with other children more frequently, or he may run down hallways instead of walking.

The first set of correlations focus on the relationship between the impulsivity measures and performance on the Stanford-Binet. The correlations involving the DAL are based on the "rate" measure; that is, length of line/time. (Correlations were run between absolute time elapsed on the DAL and Stanford-Binet and they are the same magnitude or greater, than those to be reported now.) With this measure, a high score signifies a greater degree of impulsivity than a low score. Thus, if impulsivity is related to Binet performance, we would expect negative correlations.

Scores on the WAL are direct time measures, and they would be expected to correlate positively with Binet performance. This discussion also includes an analysis of both the error and latency scores on the perceptual speed test. A negative correlation between errors on the PST and Binet would be anticipated, whereas a positive correlation between latency on the PST and Binet is anticipated.

Table 6 provides the relevant correlations showing the pre-test correlations in the upper part and the post-test correlations in the lower part of the Table. Examination of Table 6 indicates that both the WAL and the DAL are significantly related to performance on the Stanford-Binet, for both pre- and post-test. The PST measures, for the pre-test, are not significantly related to any of the other measures. For the post-test assessment, however, the PST measures are significantly related. Indeed, it should be noted that the correlations for the post-test measures are generally higher than those for the pre-test measures.

One obvious question that can be raised concerning the correlations between the DAL and Binet and the WAL and Binet is the degree to which performance on the impulsivity measures reflects the ability to understand the instructions. (This is an interesting question in that it implies, from our data at least, that understanding instructions accounts for something like 36% of the variance on Stanford-Binet). The question can be answered, at least partially, from procedures used in this study with the WAL. Prior to giving the children instructions about going "slow" and going "fast", we asked the children to walk the line without any instructions. The correlation between performance with no instructions and pre-test Binet was .43 and when the same procedure was repeated on the post-test the correlation with post-test Binet was .51 (both correlations are statistically significant at the .05 level). Of greater interest is the correlation of .81 between the "no instruction"

Table 6
Inter-Correlations of the DAL, WAL, PST and
Stanford-Binet for Pre and Post-Test
(Decimal points have been deleted).

	1 WAL	2 DAL	3 errors	PST 4 latency	5 S-B
1		-40*	-10	-06	44*
2	-72**		17	-23	-45*
3	-39*	44*		-29	-26
4	51**	-57**	-10		-11
5	60**	-56**	-43*	43*	

* p => .05
** p => .01

condition and the instruction to walk "slowly", but only a correlation of .38, which is not statistically significant, between the "no instruction" condition and the instruction to walk "fast". These data suggest that children who are impulsive under the "no instruction" condition are unable to conform to the "slow" condition but that the more controlled children under the "no instruction" condition can conform to the "fast" instructions. We have already noted that there is a statistically significant difference in performance under the "slow" instructions as opposed to the "fast" instruction condition. Finally, it should be noted that all the correlations involving the WAL and the DAL in the "fast" condition and the Stanford-Binet measure are essentially zero. Thus, it would appear that the correlations for the DAL and WAL with the Stanford-Binet are not a function of the childrens' ability to understand the instructions; namely, the distinction between "slow" and "fast". Rather, it would appear that those children who are able to conform to the "slow" instruction are able to utilize environmental information than more impulsive children. This latter assertion receives some support from the fact that the correlation between the WAL pre-test and change score on the Binet is .56 which indicates a significant relationship between I.Q. point gain and motor control.

One of the interesting, and unexpected, findings in this study were the generally higher correlations among the post-test measures; particularly where latency measures were involved. It was found, for example, that correlations between pre- and post-test latency measures were very low, and in some instances were even negative. The pre-test latency measures, however, generally do not correlate with any of our other measures but the post-test latency measures correlated significantly. Our interpretation of these results is that, perhaps as

a result of the intervention experience, or as a result of their pre-test experiences, the within subject variability over tests was sharply reduced; that is, the post-test measures reflect the child's general test taking style. The alternative explanation that between subject variability increased from the pre- to the post-tests is not supported by the data. It would appear, therefore, that our post-test measures provide a more accurate picture of the child's behavioral capability, especially with respect to latency measures. These data also suggest that too early administration of a pre-test battery may not provide the best estimate of the child's range of capabilities.

One final comment on the DAL and WAL tests. Our results are in general agreement with those reported by Maccoby et al. (1965) whose subjects were quite different from this sample, at least in terms of average intellectual ability. The results are not in agreement, however, with Hess et al. (1966) and Banta (1967) who failed to find significant relationships between the DAL and Binet performance. Obviously, more research is needed to define the correlates of impulsivity and, more importantly, to determine under highly controlled conditions the behavioral consequences of impulsivity.

Table 7 summarizes the correlations between the DAL, WAL, Binet, and the ABRS. Our concern here was with the degree to which the teacher's perception of the adaptive behaviors of the children were related to our impulsivity measures and Stanford-Binet performance. These correlations indicate that both the teacher's and aide's perceptions of the adaptive behaviors of the children are significantly related to the children's performance on the Stanford-Binet. With one exception, the impulsivity measures are also significantly related to the ABRS, but only for the pre-test scores. It is quite possible that during the six-week intervention program the more maladaptive social behaviors of the more impulsive children were

Table 7
Intercorrelations of ABRS with DAL, WAL, and
Stanford-Binet for Pre- and Post-Tests
(Decimals have been deleted)

	Teacher		Aides	
	Pre ABRS	Post ABRS	Pre ABRS	Post ABRS
Binet	43*	43*	46*	52*
WAL	22	-05	44*	14
DAL	-58**	-30	-55**	-29

* $p = >.05$
** $p = >.01$

extinguished, and that they learned to make the more elementary responses considered important by the teachers and the aides. It is also clear, however, that as perceived by the teachers and aides the brightest children tend to make the most socially adaptive responses. Incidentally, it should be noted that none of the cognitive measures were significantly correlated with change score on the ABRS.

V. The Arithmetic Project

In addition to the regular classroom porgram, the research staff conducted two special programs; 1. a Montessori Program, and 2. a Bereiter type program focusing on arithmetic only. The Montessori class was conducted by Mrs. Jacquline Morrow Massari who had previous experience with a Montessori class but with no formal training in the techniques. The Bereiter group was conducted by Miss Gwen Simpson, under the direction of Dr. Vernon Hall.

Two Montessori groups of five each were established by randomly selecting the 10 children from among the 33 enrolled in the program. The Bereiter group consisted of five randomly selected children from the sample. A total of eight randomly selected children served as control subjects.

Since the Bereiter group only worked on arithmetic, comparisons among the groups were based only on arithmetic achievement (undoubtedly other perhaps more important variables should have been assessed but time did not allow for anything more. Besides, there is considerable confounding because of the design used. Indeed, the entire project must be considered as exploratory.) All groups were administered an arithmetic achievement test developed for this specific purpose. A pre-test post-test design was used.

The children in the three experimental groups, two Montessori and one Bereiter, went to their particular rooms for one-half hour each day. The control group remained in their own classrooms. Montessori materials were available to the

children and no special emphasis was arithmetical operations or concepts. The Bereiter group followed the program used by Bereiter.

Although statistical analysis are not completed, there may be some interest in how the children performed on the test (see Appendix A for a copy of the test). Table 8 summarizes the means and SDs for the pre- and post-tests for each group. These data indicate that the largest gain occurred in the Control Group and is statistically significant ($t = 4.6$; $df = 7$; $p < .01$). The average gain for the pooled Montessori groups is not statistically significant ($t = 1.8$; $df = 8$; $p > .05$) nor is the gain for the Bereiter group statistically significant ($t = 1.6$; $df = 4$; $p > .05$).

Perhaps the most significant aspect of the data in Table 8 is the substantial variation in performance on the test. Examination of pre-test scores for the entire group shows a range of 0-23 and a post-test range of 4-29. These data suggest that some preschool children are capable of handling some fairly sophisticated arithmetical concepts and operations.

Discussion

Although all of the comparative and inter-correlational analyses are not completed, several implications would seem to emerge from the available data.

Considering first, the various measures of cognitive ability, it is clear that in our sample a meaningful proportion of the variance in Stanford-Binet performance is related to performance on the "impulsivity measures". In particular, we are impressed with the relationships between the DAL, the WAL, and Binet performance, because neither the DAL or the WAL would appear, on the surface, to be a measure of what is commonly called "g". This is particularly clear with respect to the WAL test where no relationship was found between walking a line fast

and Binet performance but where correlations of .50 were found between walking a line "slow" and Binet and walking a line without any particular instructions and Binet. The "no instruction" condition would seem to indicate how children respond to a variety of demands in a more naturalistic setting. Thus a hypothesis that impulsive children respond to a broad variety of stimuli, both relevant and irrelevant, in their environment without integrating them or establishing appropriate response patterns to these stimuli seems to be indicated. Our data, along with those reported by Maccoby et al. (1965), suggest that the deleterious effects of impulsivity are as apparent among children of superior intellectual ability as among those with low normal ability.

Perhaps the most significant aspect of our findings with the DAL and WAL is the fact that no relationship was found between "fast" condition and the Stanford-Binet. This finding coupled with the fact that the difference in time under the "slow" and "fast" condition was statistically significant suggests that the correlation between the "slow" conditions and the Binet is not an artifact of understanding the instructions.

Our findings with respect to the children's responses to the cognitive demands of the Stanford-Binet were somewhat disappointing in that increases in per cent work responses were not especially related to increases in Stanford-Binet I.Q. Apparently, our hypothesis was only accurate for those children in the middle I.Q. range of our distribution, but not for those children on either extreme. This raises certain interesting questions. One possibility is that improved motivation leads to a greater tendency to give relevant responses to the Stanford-Binet. This assumption was supported by our data but, as already noted, increased work responses were not demonstrably related to I.Q. gain. In this context it should be kept in

mind that the designation of a "work" response is not determined by whether or not the answer given to the question is correct. Thus a child who increases in "work" responses is now putting out a necessary but not sufficient response in order to get a Binet item correct. It may be hypothesized then that in order to get an item correct a child must both have a set to work at answering and also have acquired the necessary content. It would be expected that a general "non-work" mode of response would mitigate against learning throughout the child's development. A change in tendency to work, therefore, would not result in an increased Binet score unless or until the work tendencies had operated in the acquisition of sufficient content to enable the child to have the "right answer" response in his repetoire. A change in "work" tendency may then be seen as a very real improvement with which other cognitive inputs will interact.

Another possibility, more difficult to examine empirically, is that initially poorly motivated children in fact emit work responses but make little or no effort with respect to the accuracy of their response. Indeed, examiners have been known to report that children sometimes seem to make errors on purpose. If this were the case, our work-non-work categories would be unable to detect such subtle changes in behavior.

Several aspects of the results of our examination of the teachers' perceptions of the children merit comment here. As just about everyone would expect, on the basis of studies such as Rosenthal's (1966), the teachers perceptions of children's social adaptiveness is significantly and positively correlated with intelligence. This finding occurs despite the substantial precautions taken to avoid communicating to the teachers about the intellectual abilities of the children. The issue may not be as simple as the notion that teachers like bright children only because of

their intellectual competence. Our data suggest, for example, a tendency for the less able children to display a greater degree of impulsivity. Many of the adaptive behaviors specified by the teachers require, for example, substantial motoric control. In other words, it is important to teachers that children stand quietly in line or sit quietly in their seats while they are giving instructions. The high impulsive child is less likely to perform this task, thus becoming a source of irritation to the teacher. Given a broad range of intellectual abilities, it would be surprising indeed to find no relationship between intellectual and social abilities. These data suggest that it may be well to formulate our theoretical models of child socialization in terms of cognitive structures. It could be argued that children who can grasp and act upon conceptualizations will be in a better position to understand and accept the demands made by their teachers. It also follows that the more impulsive child, who may find his relationships with teachers more difficult even with mature cognitive structure, is more likely not to have such structures available to him. Thus, this child is placed in a much more difficult situation as suggested by the low ratings they tend to receive from their teacher. We do not mean to imply that all of the variance in teacher attitudes in children can be accounted for in terms of variation in cognitive structure. Certainly there are variations in the social adequacy of children's behavior even where there is homogeneity with respect to intellectual ability and cognitive style of functioning. Such variations undoubtedly result from important influences shaping the child's social behaviors, such as parent-child and teacher-child relationships. Our data provide only the most minimal cues as to what happens to the less positively perceived children, namely, that they tend to receive a substantial proportion of the teacher's blame statements without necessarily a concomitant

increase in praise for more controlled behavior. Manipulation of unapproved behaviors is apparently attempted almost entirely by punishment albiet verbal and not particularly severe.

A final observation. Throughout the arduous business of data analysis, one could not help but feel overwhelmed by the variation among the children on each of the measures and the concomitant gain scores. This feeling of concern about the variation is certainly not unique to our research group, but one can't help wondering why classroom programs continue to have a uniformity of procedures over all children. Perhaps future research efforts should concentrate on identifying the salient dimensions of variability (multi-variable statistical techniques are now available) and, then concentrate on variations in programming for individual children.

Table 8
Means and SDs for Arithmetic Test

	Pre-Test		Post-Test		Gain
	M	SD	M	SD	
Montessori I	10.4	7.6	13.2	7.3	2.8
Montessori 2	7.7	2.5	9.7	2.8	2.0
Bereiter	9.2	8.2	12.4	6.8	3.2
Control	9.9	7.2	15.2	9.9	5.3
Total	9.4	6.5	13.1	7.5	

References

Banta, T. Tests for the evaluation of early childhood education. The Cincinnati Autonomy Test Battery. Mimeo, 1967.

Caldwell, Bettye, Heider, Jaye, Kaplan, Barton. The Inventory of Home Stimulation Delivered at APA, New York, September 1966. Mimeo, 1966.

Flanagan, J. A proposed procedure for increasing the efficiency of objective tests. Journal of Educational Psychology, 1937, 28, 17-21.

Gardner, E. and Thompson, G.G. Social relations and morale in small groups. N.Y.: Appleton-Century-Crofts, 1956.

Hertzig, Margaret E., Birch, Herbert G., Thomas Alexander, and Mendez, Olga Aran. Class and Ethnic Differences in the responsiveness of pre-school children to cognitive demands. Mimeo, 1967.

Hess, R., Kramer, R. Slaughter, Diane, Torney, Judith, Bury, Carla, and Hall, Ethel. Techniques for assessing cognitive and social abilities of children and parents in Project Head Start. Mimeo, University of Chicago, 1966.

Kagan, J., Moss, H. and Siegel, I. Psychological significance of styles of conceptualization. Monographs Society Research Child Development, 1963, 28 (#86).

Maccoby, Eleanor, Dawley, Edith M., Hagen, J.W., and Degerman, R. Activity level and intellectual functioning in normal pre-school children. Child Development 1965, 36, 761-770.

Meyer, W.J. and Thompson, G.G. Sex differences in the distribution of teacher approval and disapproval among sixth-grade children. Journal Educational Psychology, 1956, 47, 385-396.

Zigler, E. Personal Communication