DOCUMENT RESUME

ED 128 426	95	TM 005 611

AUTHOR	Karlson, Alfred L.; Stodolsky, Susan S.
TITLE	Predicting School Outcomes from Observations of Child Behavior in Classrooms.
INSTITUTION	Chicago Univ., Ill. Chicago Early Education Research Center.
SPONS AGENCY	Office of Economic Opportunity, Washington, D.C.; Office of Education (DHEW), Washington, D.C.
PUB DATE	[Feb 73]
GRANT	OEO-CG-8469
NOTE	24p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, Louisiana, February 1973)
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	Class Activities; *Classroom Observation Techniques; Cognitive Development; Cognitive Tests; Disadvantaged Youth; Evaluation Methods; *Individual Differences; Prediction; Preschool Education; *Preschool Programs; Program Effectiveness; *Program Evaluation; Social Class; Statistical Analysis; *Student Behavior; Test Results
IDENTIFIERS	*Montessori Method; Project Head Start

ABSTRACT

A modified Montessori nursery school program operating in a private urban school serving a racially heterogeneous population was evaluated. The project called for investigating the educational effects of the program on both Head Start and middle class children. The impression was that children participated in the project differentially. That is, they had unique patterns of work and play when they were given a choice of different activities. The evaluation procedures spanned the entire school year. Tests were administered to a total sample of 58 children, divided equally by social class during the first and last month of the school year. The classroom observations were recorded for a subset of this group, which contained 43 children. This sample of 43 children was also equally divided by social class. Each Head Start child was pair-matched to the extent possible on age, sex, classroom, and previous school experience with a middle class child. Twenty six of these children were above four years of age and received the Wechsler Preschool and Primary Scale of Intelligence and the Stanford-Binet, L. M., the remaining children received the Merrill Palmer Scale of Mental Tests and the Stanford-Binet. Classroom observations were conducted during the period between testings. Findings are discussed in terms of the observational data, the test data, and the relationship between the two. (RC)

Predicting School Outcomes from Observations of Child Behavior in Classrooms[1,2]

Alfred L. Karlson, University of Massachusetts, Amherst and Susan S. Stodolsky, University of Chicago

Recently we assumed responsibility for evaluating a modified Montessori nursery school program which was operating in a private urban school. The school served a racially heterogeneous population. Most of the children came from middle and upper-working class homes, but in addition approximately thirty poor black children from the neighborhood were participating in the regular classroom program through a unique Head Start program. The evaluation project called for investigating the educational effects of the program on both the Head Start and middle-class children.

It was our hope to understand the objectives of the program well, to assess how it was implemented, and to measure some aspects of its effectivenss on the children who participated in it. These modest goals led us considerably beyond a traditional pre- and post-test study, to an enormous expenditure of time and energy systematically watching and recording what individual children did in the preschool classrooms of the program which we were evaluating.

Today I would like to share with you some aspects of this experience and to more specifically describe a strategy which we developed for doing evaluation which suggests that time spent watching child behavior in the classroom can lead to a better understanding of what children learn from their school experience.

When we first considered evaluation strategies, we were struck with what we knew about these Montessori classrooms from our previous study of child behavior in them. We had just completed a research project which utilized observational methods to describe how children use the free play periods of nursery and kindergarten settings (Stodolsky, in press). The particular Montessori school we were now evaluating had been used as a site for this earlier work, because, in accordance with Montessori tradition, most of the day was spent in work sessions in which the children were encouraged to spontaneously choose their own activities. While the focus of our observational studies had been on how children selected activities, we were impressed with the differences we saw among children with respect to the actual activities they selected to pursue. It was our distinct impression that children participated in the program differentially. That is, they had unique patterns of work and play when they were given a choice of different activities as is the case when the curriculum is implemented under "free play" conditions. This impression of different patterns of activity was supported in a pilot study completed in the summer session prior to the evaluation study. It is also supported by earlier literature dealing with the behavior of preschool children under free play conditions (e.g., VanAlstyne, 1932; Bott, 1928; Bridges, 1929).

This pattern of differential participation suggested to us that a meaningful evaluation of the regular school year program must take child behavior into consideration. We found ourselves making this argument: Evaluation is essentially a type of treatment effect research. In this case we were interested in the effects of the program on the students who participated in it, but we recognized that the treatment of attending a Montessori school would be varied depending on what the individual children actually did in the classroom. For us this represented a major problem with the traditional pre- and post-test evaluation model for research on educational effects. This problem obviously also applies to evaluations of preschool programs which have significant free choice components in their curriculum, as well as to open classrooms or other programs which stress individual selection from the curriculum.

Two important methodological issues had to be resolved in order to conduct an evaluation which would take differential participation into account. First, having acknowledged that each child might experience a highly unique educational experience we had to grapple with the fact that the probability of finding meaningful group effects on outcome measures was seriously threatened. It seemed imperative to try to use measures of outcome which seemed well-articulated with a variety of experiences available in the classroom. Second, we had to develop a method for observing child behavior which would lead to a valid and reliable estimate of the treatment or patterns of participation for each child. If we could meet these two criteria (empirical descriptions of the child's behavior in the classroom and outcome measures which were in fact sensitive to the curriculum) it was our ultimate purpose to be able to predict outcome from our knowledge or variation in treatment.

I will discuss our rationale for instrument selection first. Ideally, we would have liked to develop a wide variety of measures which would articulate with the presumed educational outcomes of various parts of the Montessori curriculum. However, we already planned a large investment in the observational component of the evaluation study and were concerned that new instruments, given the time constraints involved, would not be sufficiently reliable and valid for our purposes. We searched, therefore, for instruments which appeared to articulate reasonably well with the curricular emphases and which were appropriate for the age range being studied. Since there had been earlier evaluation studies conducted in the school (Kohlberg, 1968) we also wanted to maintain some continuity with the earlier work.

From our reading on the Montessori method, and from previous experience observing in the classrooms, we hypothesized that the children participating in the program ought to gain in certain types of cognitive and related psychomotor skills, but would not show sustained growth in language areas. While there seemed to be posible socio-emotional effects from participating in the program, we chose to base our evaluation in the cognitive domain because many of the Montessori activities are clearly designed to teach specific cognitive skills for which there existed standardized measures.

We therefore chose to pre and post test the children above 4 years of age with the following sub-scales of the Weschler Preschool and Primary Scale of Intelligence (WPPSI): Animal House (a sorting and matching task); Arithmetic, which stressed concrete presentation of numbers; Geometric Design which was an eye-hand coordination task with the use of a pencil; Block Design, which tapped visual motor integration; and Mazes which measured psycho-motor skill in eye-hand coordination. Each one of these tests could be directly related to groups of specific Montessori and traditional activities. In addition, the sub-tests (with Arithmetic omitted) could also be used to compute a prorated performance I.Q. score. We also chose to test all children with the Stanford Binet, L.M., to provide a contrast to the performance emphasis in the WPPSI and to provide research continuity with earlier work at the school (Kohlberg, 1968). For children under 4 years of age, we administered parts of the Merrill Palmer Scale of Mental Tests.

With these measures we hoped to be able to describe cognitive gains made by the children in a somewhat specific manner. We assumed that overall, the performance measures were more appropriate to the curricular emphases and would show sustained effects in the children, whereas the Stanford-Binet would generally not show significant effects as it could not be meaningfully linked to activities in the classroom.

Our next problem was to design the observational component of the evaluation project with which we would empirically define treatment. We had expected that the children would participate in the program differentially, and would develop individual patterns of work and play and we now sought a way to record and demonstrate this differential participation so that we could predict cognitive gain from it. Previously we had developed a systematic method for describing child behavior in the free play periods or open choice learning situations of classroom settings. This method made several distinctions between kinds of behaviors seen in free play sessions. For this purpose, the most important distinction was to differentiate transition times or times between activities from actitity segments themselves.

Our previous work had already established methods for making these distinctions. From a narrative record of the stream of behavior, activities are coded when the child exhibits focused behavior with a beginning, middle and end which lasts at least one minute. Behaviors not meeting the activity criteria are transition or inbetween behaviors. This rather straight-forward definition of activity behavior has been easy to apply and leads to inclusion of behaviors one would ordinarily class as activities such as building with blocks, drawing a picture, doing a puzzle, etc.

Our strategy was to collect data on the activity behavior of each child during the school year and to make a taxonomy of different kinds of activity behavior. Differential participation in the program, or differences in how the children used the curriculum would then be compared in terms of percent of time distributed among different activity categories. We decided that five-minute time samples would be appropriate because we had found that five minutes closely approximated the average length of an activity segment in this age group when we were studying transition behaviors and recording in fifteen minute time samples. It was more difficult to establish an appropriate number of such samples because little was known about the true stability of activity behavior over time, and we wanted a large margin for measuring this stability. We therefore decided to observe the children on a near daily basis.

All of these decisions, of course, led us to the realization that this evaluation of a small Montessori based Head Start and nursery school program would involve an extraordinary expenditure of time and energy with certain built in risks. But we proceeded from the assumption that the children would, as a group, improve on the standard psychometric measures we had chosen, and that by taking a close look at the treatment which we assumed would be effecting this change, we would find differences. If both of these assumptions turned out to be true, then we could relate treatment to outcome and perhaps in the larger picture, demonstrate a model for evaluation that takes into consideration a free choice or open classroom learning setting.

Our procedures, briefly described here, spanned the entire school year. The tests were administered to a total sample of 58 children, divided equally by social class during the first and last month of the school year. The observations were recorded for a subset of this group, which contained 43 children. This sample of 43 children was also equally divided by social class. Each Head Start child was pair-matched to the extent possible on age, sex, classroom, and previous school experience with a middle-class child. Twenty-six of these children were above four years of age and received the WPPSI and Stanford-Binet, the remaining children received the Merrill-Palmer and the Stanford-Binet. Observations were conducted during the period between testings (Approximately 7 months).

All children were observed by four trained observers on a near daily basis or at least on four separate days of the school week. Observers rotated classrooms, and randomly selected children for observation. Each child was observed for five minutes. These five minute samples were made up of narrative records of child behavior of the kind described by Barker and Wright (1955), and Wright (1967). Activity behavior was coded from the records using Stodolsky's method, previously described. Each activity segment was also given a brief content label. Both field and coding reliability were established and maintained at a high level by observers throughout the school year using standard methods of blind double recording and double coding. The coded and labeled activity segments for all children were then content analyzed forming a natural taxonomy of activity behavior actually observed. To some extent the categories reflected aspects of the Montessori curriculum; however, the basis for classifying or grouping activities was also made by judging the activities potential for facilitating the acquisition of skills.

It is possible to discuss the findings in terms of the observational data, the test data, and the relationship between the two, which was, of course, our primary concern. The first condition necessary to demonstrate a relationship between the observations and the test gains was to find change on the psychometric measures believed relevant to the curriculum. This condition was met for the sample in regard to the WPSSI prorated performance I.Q. which is a combination of most of the scales we gave. Table 1 presents the results of the WPPSI administration for the 26 children who received this test. Significant change on this measure was not associated with sex or social class; however the Head Start children as a group had lower initial and final scores than the middle class-children. The children did show significant gains on some of the separate scales as well, but sufficient change was only available for the entire scale for the purpose of predicting gain from the observational data.

The Stanford-Binet results presented a different picture, one we had anticipated. The only group of children who gained significantly on the Stanford-Binet were young children in their first year of school attendance. We attributed this finding to the initial schooling effect which has so frequently been reported in other preschool evaluation studies and not to specific curricular effects. The Merrill-Palmer data showed significant gains for the Head Start children tested, but the middle-class children did not show such gains. It appeared in retrospect that the Merrill-Palmer was too easy for the young middle-class group as a clear ceiling effect was present in the data. Thus we could not use these data in conjunction with the observational material.

A full report of the test results for the entire sample studied is available and includes results on all of these measures and the more detailed rationale for their use (Stodolsky and Karlson, 1972). To recap, the WPPSI Performance I.Q. and the Stanford-Binet measures were appropriate for use in further analyses in connection with the observation data.

In the observational data we found the second necessary condition: As we had expected the children had utilized the program differentially. The content analysis of activity behavior revealed that the children had unevenly distributed their time among fifteen different categories of activities. The fifteenth was the only "catch all" category which was called non-curriculum relevant activities, and included activities such as running the water in the bathroom, watching the janitor sweep the hall, or discussing bubble-gum cards from home. The children did this the least amount of time, spending an average of 1.9 percent of their activity behavior this way. The other categories included practical life exercises, eating snacks, eye-hand coordination in use of the pencil activities, art, construction toys, unit and table blocks, sensorial sorting and matching activities, picture and Montessori puzzles, mathematics exercises, reading, socio-dramatic play, and social interaction.

Table 2 shows the average percentage of time spent in each category of activity for the sample of 43 children observed and includes a brief description of each activity category. As can be seen in Table 2, the mean percent time engaged in these different categories varied from a high of 14.9 in practical life activities to one percent in doll corner play. Construction toys, blocks, sensorial sorting and matching, and math all averaged around 7.5 percent, while others were under 5 percent. It is important to note that the standard deviations associated with these means were high in all cases. Using a multivariate analysis of variance, an extension of the matched t test, differences among the children's distributions were found to be significantly different beyond the .0001 level. (The F statistic was 1527.5 with 13 and 27 degrees of freedom.) This test supported statistically the fact that each child was making differential use of the curriculum.

The average number of observations (five minute time samples) on each child was 64. Variations in this number could be attributed to attendance rather than sex, age or social class. Spearman Brown odd-even correlations for mean percent time engaged in each activity category are consistently above $r = .65$, significant at the $p < .01$ level for the fifteen categories. Split half correlations are somewhat lower, with nine of the fifteen reaching significance; however, these data suggest that activity behavior is stable for most children over the school year.

As a group, the sample children participated in the curriculum differentially, that is there were differences in popularity of activities. Some variation in this differential utilization of the curriculum was demonstrated to be related to age, sex, and initial level of ability as measured by the Stanford-Binet pretest. Some illustrative group differences can be presented here, but extensive data on this issue is available elsewhere (Karlson, 1972).

As a group, younger children were observed doing more practical life exercises, construction toy activities, sorting and matching exercises, and puzzles, while older children did more math, reading, and eye-hand coordination exercises. No activities were totally age dependent although a developmental trend did seem to exist which seemed to reflect differences in difficulty of activities of the curriculum. In an examination of sex and social class influences on differential participation, cultural stereotyping was seen in the girls' preference for practical life exercises and the boys' preference for construction toys and blocks, but art and sociodramatic play were uneffected by sex differences which runs contrary to previous findings. It was found that middle-class children spent more time doing reading and math exercises, but in an analysis of covariance, using initial level of ability (Stanford-Binet pre-test) as a covariate, the social class effect is not present. This suggests that children who have high Stanford-Finet I.Q. choose to do these activities regardless of social class. Considering all of these differences, the primary importance of the findings was the empirical demonstration of total group and individual differences in differential participation in the program. This finding was essential to the issue of defining treatment and eventually allowed us to see the amount of time engaged in an activity category as an empirical predictor of change on the psychometric outcome measures.

The prediction of measured cognitive change from observational variables was executed by designating the percent total time engaged in the activity categories as independent variables and change on the WPPSI prorated Performance I.Q. as the dependent variable in a multivariate regression analysis. As indicated earlier we had hoped to use sub-scales of the WPPSI but there was not sufficient range in change scores to use them singly. We performed a similar regression analysis for the Stanford-Binet change scores, but did not expect a significant result as there appeared to be little connection between the curricular activities and the skills tapped by the Stanford-Binet. In separate analyses, the activity predictors accounted for 85% of the variance in WPPSI performance I.Q. change, significant at the $< .02$ level; while the same predictors account for only 27% of the variance on the Stanford-Binet change scores and the analysis is not statistically significant. Table 3 presents the results of these two regression analyses.

These analyses are based on a within group correlation matrix which takes into consideration differences in the sub-class means (sex within social class) in the sample. Two small categories, noncurriculum relevant activities, and practical life in the doll corner, were eliminated to remove the linearity of the relationship between the independent and dependent variables. Because the independent variables were proportional, that is, they were percentages, an arc sine transformation was used to stabilize their variances.

Before discussing the meaning of the regression analysis presented there are some important additional issues as points of information. First of all, there is no regression effect in the test data, that is, the change reported in the test data is not a function of a relationship between the initial WPPSI score and the amount of change. Similarly, when WPPSI pre-test is used as a predictor in a regression analysis with the observational variables, it does not account for a significant amount of the variance. Moreover, the interpretation of the contribution of each predictor in the stepwise analysis must be done with caution because of the conditional nature of each step. That is, each percentage reported is a function of first removing the percentage attributable to those preceding it. Returning to the data in Table 3, it is apparent that we were very successful in accounting for variation in cognitive growth by estimating the child's "treatment" through activity sampling.

The lack of power of the observational variables in relation to the Stanford-Binet demonstrates the importance of choosing pre- and post-test measures which are sensitive to the program if one hopes to use variables generated from observing child behavior to predict to gain or in fact to get a meaningful evaluation at all. In contrast to the time spent in activity categories, the only significant relationship to gain made on the Stanford-Binet was the child's age and lack of previous school experience ($r = -.32$). Neither of these variables were related to change on the WPPSI.

After demonstrating that it is possible to use observational variables as tools for evaluation, we sought to examine more closely the relationship between participation in individual categories of activity and cognitive change. Here we came to an unexpected finding. In Table 4 we present some correlations between activity categories and WPPSI gain. The art, block, construction toy and math activities are positively related to change, but the sorting and matching exercises, the eye-hand coordination exercises, and the puzzles are all negatively related to change. That is, the Montessori activities that were clearly meant to give children training in the specific skills we measured seemed to be interfering with acquisition of the skills they were supposed to teach. If we can be permitted to hypothesize a plausible causal relationship between the amount of time engaged in the specific activity category and WPPSI gain, then a more critical question is important: What is it about the activities that contribute to or detract from learning cognitive and psychomotor skills as measured by the SPPSI? While any explanation is inferential and subject to future verification, we made the following admittedly post-hoc argument: Puzzles and sorting and matching activities interfere with the development of these cognitive skills because they teach the child a set to respond to problem solving that is <u>solution oriented</u>. That is, these activities always have a fixed solution and from practice with them, they teach the child the solution rather than an approach to a problem. On the other hand, activities such as construction toys, art, and block activities, require the child to set his own problem and his own solution to it. In contrast to fixed solution activities, these activities teach children a flexibility in approaching new problems. This is to say that a child who spends most of his time doing predominately sorting and matching exercises or puzzles, develops a routinized and fixed view of any similar problem, but apparently not the ability to approach it with the requisite interest or curiosity of how it should be done, while the reverse is true of the child who controls his own problems and solutions in activities that do not have fixed solutions. A good example is found in the Montessori Geometric Puzzles. These toys are a series of geometric shapes which progress from simple triangles to more complex forms. The child achieves the correct solution to these puzzles by matching black lines which are marked at the correct edges of the broken shapes. The most complex puzzles can be solved by matching these lines. Therefore this particular set of puzzles has a fixed solution. A college art project of pasting various geometric shapes together may be very similar in terms of elements, but it offers a flexibility in solution and teaches the child no fixed response. When a child is faced with a new problem, such as the classic block design task found in the WPPSI, his performance will be somewhat a function of his past experience with the manipulation of geometric forms. It seems that the child who has had previous practice in developing his own solutions will be better able to transfer his skill to the new problem.

In any case, it is findings like this, based on exploring the relationship between observations and outcome, that give us the notion that variables generated from direct observation of child behavior can lead to a more meaningful evaluation of preschool programs, or of any program that incorporates variation in its implementation. The initial findings from this small evaluation study have demonstrated the feasibility of using observational variables in understanding how children utilize a program and how they learn from it. What seems to be needed now are research approaches which refine the method in its three essential components; these being, unitizing observable activity behavior or other classroom events, finding or developing external criterion to be related to the observational data, and exploring new methods of demonstrating the relationship between the observational data and outcome. In this way it would seem possible to meaningfully compare different kinds of programs and assess the effectiveness of their different elements.

In view of the substantive findings of this study, it would also be very important to attempt replications and extensions of this work both in the naturalistic mode we chose and using experimental designs. If the distinction between fixed solution and open-ended tasks could be found to be generally supported under conditions of free choice and in conditions in which such activities were imposed on the child, important educational implications would emerge.

We have tried elsewhere (Stodolsky, 1972; Karlson, 1972) to explore in more depth the evaluation implications of this study. It seems clear that these results call into question evaluations of programs with significant free choice components which do not take into account the particular course the child pursues in the program. Data such as these would go a long way to resolving and interpreting the differential effects of programs which operate at various points along the structure continuum.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF WPPSI PRORATED I.Q. SCORES AND STANDARDIZED SCALE SCORES BY SEX AND SOCIAL CLASS
(N=26)

	Performance I.Q.				Animal House				Mazes			
	Time 1	Time 2	t	p	Time 1	Time 2	t	p	Time 1	Time 2	t	p
Head Start boys Mean Age = 53.8 N=6 (4.6)	89.00 (10.63)	97.33 (12.56)	2.91	.05	6.66 (3.72)	8.66 (2.16)	1.38	ns	8.83 (1.36)	8.33 (1.36)	--	ns
Head Start girls Mean Age = 55.8 N=6 (9.0)	89.25 (3.55)	93.33 (8.89)	.92	ns	8.33 (1.75)	9.50 (2.25)	1.23	ns	6.83 (.98)	7.83 (1.16)	1.37	ns
Middle Class boys Mean Age = 55.6 N=7 (5.0)	99.00 (11.80)	108.57 (10.81)	2.48	.05	[illegible]	11.71 (1.38)	1.69	ns	10.57 (2.87)	11.85 (2.67)	5.43	.01
Middle Class girls Mean Age = 55.6 N=7 (6.8)	111.14 (18.07)	118.57 (10.81)	5.49	.01	11.28 (2.81)	13.14 (2.54)	2.25	ns	11.00 (3.21)	12.71 (3.25)	2.76	.05
Total Mean Age = 55.1 N=26 (6.5)	97.73 (14.93)	105.15 (16.54)	4.30	.01	8.96 (3.72)	10.88 (2.68)	4.30	.01	9.42 (2.85)	10.34 (3.09)	3.04	.01

	Arithmetic				Geometric Design				Block Design			
	Time 1	Time 2	t	p	Time 1	Time 2	t	p	Time 1	Time 2	t	p
Head Start boys	8.00 (.89)	9.66 (1.50)	2.06	ns	8.83	11.33	1.81	ns	9.16 (2.31)	10.00 (2.60)	.99	ns
Head Start girls	5.33 (1.96)	7.16 (1.94)	1.83	.05	9.33	8.88	--	ns	9.66 (2.16)	10.00 (2.09)	.33	ns
Middle Class boys	10.00 (1.52)	10.57 (1.81)	.83	ns	9.42	9.85	--	ns	10.00 (1.52)	11.71 (3.03)	1.72	ns
Middle Class girls	10.46 (4.09)	12.14 (3.53)	1.98	ns	11.71	12.42	.54	ns	12.42 (3.03)	14.00 (2.58)	1.54	ns
Total	8.57 (3.44)	10.00 (2.88)	4.03	.01	9.88	10.61	1.23	ns	10.38	11.53	2.55	.05

TABLE 2

DESCRIPTION OF ACTIVITY CATEGORIES AND
AVERAGE PERCENT TIME IN EACH (N = 43)

1. Practical life exercises ($\bar{X}$ = 14.9, s.d. = 7.6) cleaning, peeling and cutting carrots, slicing apples, cracking nuts, making toast, spreading peanut butter on crackers, watering the plants, sweeping the floor, washing tables, polishing silver, and tying, zippering, buckling and buttoning the Montessori practical life exercise frames.

2. Eating snacks ($\bar{X}$ = 3.2, s.d. = 3.4) taking time out from the usual activities to have milk and cookies or other food.

3. Doll corner play ($\bar{X}$ = 1.0, s.d. = 3.4) playing with the equipment in the doll corner without socio-dramatic play or social interaction. Dressing up in play clothes, using kitchen utensils, doing house work, a sub-category of practical life.

4. Eye-hand coordination in use of the pencil activities ($\bar{X}$ = 8.6, s.d. = 7.0) drawing with the Montessori geometric insets, tracing maps, tracing pictures, writing using a model, copying numbers, copying models of shapes and numbers on the blackboard, and doing connect the dots drawing exercises.

5. Art ($\bar{X}$ = 13.6, s.d. = 7.2) painting at the easle, making collages from cut paper, finger painting, using clay, coloring with crayons.

6. Construction toys ($\bar{X}$ = 6.5, s.d. = 5.9) these toys are like puzzles but have no prescribed solution, the child can put them together as he likes or as well as he can. Many of these toys are known by their trade names which have been avoided here. In general there were four types of toys of this kind, the first utilized rods or sticks and connectors, the second utilized knotched blocks of various shapes, the third utilized ordinary blocks to build mosaic patterns, and the final types were building toys which used various gears, nuts and bolts, and wheels and axles.

7. Unit and table blocks ($\bar{X}$ = 6.0, s.d. = 7.4). The unit blocks are used for the construction or building of structures in the block corner. They are heavy, wooden blocks of multiple sizes. Table blocks are small unknotched cubes.

TABLE 2 cont.

8. Sensorial sorting and matching activities ($\bar{X}$ = 7.76, s.d. = 5.0). These activities included all of the Montessori sensorial matching exercises and supplementary matching games which are based on visual discrimination. Examples are: picture lotto, sorting beads by color or size, classifying natural objects, certain matching puzzles, the picture phonic cards, the Montessori color tablets, the Montessori cylinders, or putting away the colored pencils in their color coded holders.

9. Picture puzzles and Montessori puzzles ($\bar{X}$ = 7.7, s.d. = 5.0). These activities stressed visual motor integration. They included a variety of wooden and cardboard picture puzzles (jig-saw), geometric shapes from simple to complex, which were sectioned geometrically, three dimensional cube puzzles, and block design tasks where there were specific models to copy.

10. Mathematics exercises ($\bar{X}$ = 7.1, s.d. = 5.3). These exercises include the entire Montessori mathematics curriculum, described previously, and written number problems. The emphasis in the math exercises is on arithmetic computations using concrete representation of numbers.

11. Reading and reading related ($\bar{X}$ = 3.6, s.d. = 3.4). These exercises included looking at books in the reading area, practice reading with the teacher, mimic reading, and being read to. Exercises with the Montessori sand-paper letters were also included here. However, the phonic picture exercises were included in the category of sorting tasks because it involved sorting pictures of objects by their first sounds.

12. Sociodramatic play ($\bar{X}$ = 10.2, s.d. = 7.7). This type of activity involves role-taking play in the doll corner or other settings in which children assume the roles of adults or others and center their play around a dramatic theme. Television characters often were prominent.

13. Social interaction ($\bar{X}$ = 1.9, s.d. = 2.1). These activities are of a social purpose to the exclusion of all other purposes. That is they may be associated with minimal participation in other activities, but are easily recognizable as making friends, exchanging information about personal matters, or seeking affection.

14. Special projects ($\bar{X}$ = 5.5, s.d. = 3.5). In this case, the activity is usually unique in the sense that it is something special that the teacher has decided the child might do (like going to the office with the attendance record) or it is an activity of special interest brought in for a single day. For example, playing with a musical instrument on display, or examining a beef heart, or making a special art project.

TABLE 2 cont.

15. Non-curriculum relevant activities ($\overline{X}$ = 1.9, s.d. = 2.6). No matter how "prepared" or pre-planned the Montessori settings were, children created their own activities. This included such activities as watching the maintenance man sweep the hall, running water in the lavatories, watching the trains outside of the window or discussing bubble gum cards brought from home.

TABLE 3

MEANS AND STANDARD DEVIATIONS OF WPPSI PRORATED I.Q. SCORES
STANDARDIZED SCALE SCORES BY SEX AND SOCIAL CLASS
(N=26)

	Performance I.Q. Time 1	Performance I.Q. Time 2	t	P	Animal House Time 1	Animal House Time 2	t	P	Mazes Time 1	Mazes Time 2	t	P
Head Start boys Mean Age = 53.8 N = 6 (4.6)	89.00 (10.63)	97.33 (12.56)	2.91	.05	8.66 (3.72)	8.66 (2.16	1.38	ns	8.83 (1.36)	8.33 (1.36)	—	ns
Head Start girls Mean Age = 55.8 N=6 (9.0)	89.25 (3.55)	93.33 (8.89)	.92	ns	8.33 (1.75)	9.50 (2.25)	1.23	ns	6.83 (.98)	7.83 (1.16)	1.37	ns
Middle Class boys Mean Age = 55.6 N=7 (5.0)	99.00 (11.80)	108.57 (10.81)	2.48	.05	9.14 (4.87)	11.71 (1.38)	1.69	ns	10.57 (2.87)	11.85 (2.67)	5.43	.01
Middle Class girls Mean Age 55.6 N=7 (6.8)	111.14 (18.07)	118.57 (10.81)	5.49	.01	11.28 (2.81)	13.14 (2.54)	2.25	ns	11.00 (3.31)	12.71 (3.25)	2.76	.05
Total Mean Age = 55.1 N=26 (6.5)	97.73 (14.93)	105.15 (16.54)	4.30	.01	8.96 (3.72)	10.38 (2.68)	4.30	.01	9.42 (2.85)	10.34 (3.09)	3.04	.01

	Arithmetic Time 1	Arithmetic Time 2	t	P	Geometric Design Time 1	Geometric Design Time 2	t	P	Block Design Time 1	Block Design Time 2	t	P
Head Start boys	8.00 (.89)	9.66 (1.50)	2.66	ns	8.83	11.33	1.81	ns	9.16 (2.31)	10.00 (2.60)	.99	ns
Head Start girls	5.33 (1.96)	7.16 (1.95)	1.83	.05	9.33	8.88	--	ns	9.66 (2.16)	10.00 (2.09)	.33	ns
Middle Class boys	10.00 (1.52)	10.57 (1.81)	.83	ns	9.42	9.85	—	ns	10.00 (1.52)	11.71 (3.03)	1.72	ns
Middle Class girls	10.00 (4.09)	12.14 (3.53)	1.98	ns	11.71	12.42	.54	ns	12.42 (3.03)	14.00 (2.58)	1.54	ns
Total	8.57 (3.44)	10.00 (2.88)	4.03	.01	9.88	10.61	1.23	ns	10.38	11.53	2.55	.05

TABLE 3A

STATISTICS FOR REGRESSION ANALYSIS WITH 13 PREDICTORS AND WPPSI PRORATED PERFORMANCE I.Q. CHANGE BASED ON WITHIN GROUP CORRELATIONS N=26

Square Mult'R	Mult'R	F	P less than
.8545	.9244	4.0652	.0207

The following five categories account for 76% of the variance in the stated order in the stepwise analysis.

Category	% of Variance Accounted For
1. Art (5)	13.5
2. Construction Toys (6)	16.8
3. Blocks (7)	7.0
4. Sorting and Matching Exercises (8) (-)	16.6
5. Math (10)	22.6
6. All Others	16.0
TOTAL	92.4%

TABLE 3B

STATISTICS FOR REGRESSION ANALYSIS WITH 13 PREDICTORS AND STANFORD BINET I.Q. CHANGE N=43

Square Multiple R	Mult'R	F	P less than
.2749	.5243	.8459	.6125

The following categories accounted for 15.46% of the variance in the stepwise analysis.

Category	% of Variance
Math	10.38
Reading	5.08
All Others	36.97#
Total	52.43

#No other category by itself contributed over 3 percent additional variance.

TABLE 4

CORRELATIONS BETWEEN WPPSI PRORATED PERFORMANCE I.Q. CHANGE AND AMOUNT OF TIME ENGAGED IN APPROPRIATE ACTIVITY CATEGORIES, N=26

Practical Life	Eye-Hand Coordination Exercises (insets)	Art	Block Activities	Construction Toys	Sorting & Matching Exercises	Puzzles
.083	-.208	.370#	.243	.289	-.519##	-.538##

Math	Reading	Social Interaction	Socio-Dramatic Play	Special Projects
.246	-.050	-.030	-.314	-.065

#$p < .05$
##$p < .01$

REFERENCES

Barker, R. G., Wright, H. F. The midwest and its children. New York: Harper & Row, 1955.

Bott, H. M. Observation of play activities in nursery school. Genetic Psychological Monographs, 1928, 4, 44-88.

Bridges, K. M. B. Occupational interests of three-year-old children. The Pedagogical Seminary, 1927, 34, 415-423.

Karlson, A. L. A naturelistic method for identifying behavioral aspects of cognitive acquisition in young children participating in preschool programs. Doctoral dissertation. Chicago: University of Chicago, 1972.

Kohlberg, L. Montessori with the culturally disadvantaged: A cognitive-developmental interpretation and some research findings. In Early education: Theory, research and action. Edited by R. Hess and Roberta Bear. Chicago: Aldine Press, 1968.

Stodolsky, S. S. How children find something to do in preschools. Genetic Psychological Monographs, in press.

Stodolsky, S. S. Defining treatment and outcome in early childhood education. In Rethinking urban education. Edited by H. Walberg and A. Kopan. San Francisco: Jossey-Bass, 1972.

Stodolsky, S. S. & Karlson, A. L. Differential outcomes of a Montessori curriculum. Elementary School Journal, 1972, 72, 419-433.

VanAlstyne, D. Play behavior and choice of play materials of preschool children. Chicago: University of Chicago Press, 1932.

Wright, H. F. Recording and analyzing child behavior. New York: Harper & Row, 1967.

Notes

1. The research reported herein was supported in part by a grant (CG-8469) from the Office of Economic Opportunity, Executive Office of the President, Washington, D.C. and in part by a grant from the Office of Education, U.S. Department of Health, Education and Welfare to the Early Education Research Center, University of Chicago, a component of the National Program on Early Childhood Education. The opinions expressed are those of the authors and should not be construed as representing the opinion or policy of any agency of the United States Government. The OEO grant was made jointly to the University of Chicago and the Ancona School, Chicago, Illinois. We wish to express appreication to Christine Halfar, Nancy Miller, and Ellen Potter for assistance in test administration and observing. We also wish to express appreication to the staff and students who participated in the study at the Ancona School.

2. Presented at Annual AERA Meeting, February, 1973, New Orleans.