DOCUMENT RESUME

ED 023 453 PS 001 181

By-Holmes, Monica Bychowski; And Others
Interaction Patterns as a Source of Error in Teachers' Evaluations of Head Start Children. Final Report.
Associated YM-YWHAS of Greater New York, N.Y.
Spons Agency-Office of Economic Opportunity, Washington, D.C.
Report No-OEO-4152
Pub Date 5 Aug 68
Note-67p.
EDRS Price MF-$0.50 HC-$3.45
Descriptors-*Behavior Patterns, Behavior Rating Scales, *Behavior Standards, Culturally Disadvantaged, Factor Analysis, Intelligence Differences, *Intelligence Level, Middle Class, Middle Class Values, Peer Relationship, Predictive Measurement, *Preschool Children, Teacher Evaluation, *Teacher Rating
Identifiers-Caldwell Preschool Inventory, PPVT, *Project Head Start, Stanford Binet

In an effort to determine whether intelligence and school readiness could be estimated from observin g behavior and if middle class Head Start teachers ranked highly those pupils whose behavior resembled middle class behavior, the behavior patterns of 36 Head Start pupils were compared with their teachers' ranking of perceived intelligence and school readiness and with behavior patterns of 32 middle class and 40 upper-middle class children, all approximately age 4. The children's intelligence and school readiness were measured by standardized tests. Behavior patterns were observed on two schedules: continuous and summary. The results showed that (1) intelligence and school readiness were not correctly predicted from different types of behavior patterns, (2) while there was no difference in behavior among those Head Start pupils with actual intelligence differences, the behavior of those perceived to be brighter closely resembled middle class behavior, and (3) behavior patterns of all Head Start pupils were similar, regardless of actual school readiness. The data were tabulated, and suggestions were made for a replica study. A manual for using the observation schedule is included. (JS)

INTERACTION PATTERNS AS A SOURCE OF ERROR IN TEACHERS' EVALUATIONS OF HEAD START CHILDREN*

(FINAL REPORT)

Submitted by:

Monica Bychowski Holmes, Ph.D.
Project Director and
Principal Investigator

Douglas Holmes, Ph.D.
Director of Research

Arlene Friedman
Senior Research Assistant

Associated YM-YWHAs
of
Greater New York
33 West 60th St.
New York, N.Y. 10023

Research Department

August 5, 1968

* The research of which this is a report was supported, in part, by a research grant from Operation Head Start, Office of Economic Opportunity (Grant OEO-4152).

INTRODUCTION AND BACKGROUND:

Since the summer of 1965, Associated YM-YWHA member centers have operated a number of "Operation Head Start" programs for disadvantaged pre-school age children. Using some of these groups as a subject population, the Research Department of the Associated Ys conducted an evaluation (OEO-550) of the 1965 summer Head Start programs. A grant received in the Spring of 1966 (OEO 1271) made possible an investigation into possible differences among different classes of Head Start participants. The results of both these investigations have been reported elsewhere (Holmes, 1965, 1966) and will not, therefore, be discussed here other than as is relevant to this research.

As part of the procedure in the first study conducted by the Associated Ys, trained interviewers and observers were in daily contact with both the Head Start participants and their teachers. Although an evaluation of the teachers was not called for in either study, day-to-day contact with the Head Start Personnel gave research staff the impression that there were vast differences in terms not only of instructional skill, but in terms also of the accuracy with which the different teachers evaluated the children's status and progress. Indeed, it seemed that there was no common approach to evaluation which became, therefore, erratic and idiosyncratic.

At the end of the second study (OEO-1271), the Head Start teachers were asked to rank the children in their groups according to five dimensions: (1) school readiness; (2) amount of change during the course of program; (3) general appeal to others; (4) social adjustment; (5) intelligence. The teachers were given a set of cards each of which had written on it the name of one of the children in the teacher's group. The teacher was asked to rank these name cards along each of the dimensions, one at a time, each child receiving a score equal to his rank in the group. This procedure was followed for each of the scales. The data were then correlated for each teacher, using the Spearman Rho rank-order correlation.

The results showed that the patterns of interrelationships among the rankings were not only highly idiosyncratic, i.e., lacking any consistancy among the different teachers, but also that they were highly inaccurate. Specifically, the correlation between perceived IQ and actual IQ ranged from highly negative, for some teachers, through non-significant to highly significant for others. Similarly, the relationship between actual and perceived scores in terms of such variables as "adjustment" and "school readiness" fluctuated markedly, as did the relationship between, for example, "adjustment" and "appeal."

It is apparent that the teachers were responding to factors other than those ostensibly being rated, in

evaluating the status of the children. It is the basic hypothesis of the current study, that since most of the Head Start teachers come from a middle-class background, they will be most familiar with, and tend to value, middle-class behavior patterns. Thus it was predicted that they would evaluate their pupils more in terms of the middle-class model than in terms of the particular child's individual characteristics, i.e., they would respond to inter-personal rather than to intra-personal data. Therefore it could be predicted that the teachers would rate most highly those children whose behavior most closely resembled the behavior of middle-class children.

Further, the experience of the research interviewers during the follow-up phase of the initial research conducted by the Associated Ys (OEO-550) was that when the Head Start children entered public school, their teachers were apt to discourage those behaviours which were most clearly associated with cognitive growth, e.g., aggressive questioning, demanding attention and demand for creative tasks, etc. The deleterious impact of such negative reinforcement hardly needs be discussed. For instance, as Swift (1964) has noted, there is general consensus that the most important variable in the effectiveness of any nursery program is the personality and behavior of the teacher. Thus, the lack of valid infor-

mation concerning an individual child will result not only in the teachers not identifying and dealing with individual weaknesses, but can also result in their counteracting what may indeed be individual strengths.

The principal aim of the present research was to relate the behavior patterns of the Head Start children to the rankings given them by their teachers along the dimensions "perceived intelligence" and "perceived school readiness" and to compare the behavior patterns of those children to those of a sample of the middle class children enrolled in Associated Y nursery school centers. In addition, a sample of upper middle-class children enrolled in a private school was observed, using the same schedule, so as to compare the behavior patterns of the young disadvantaged child with those of the highly advantaged child. This group was not included in the original proposal and was therefore studied at agency cost. They were included, as part of this study because no objective comparison of the behaviors of disadvantaged and of highly advantaged children had yet been made. If it is true that teachers estimate intelligence and school readiness on the basis of behavior, it is important to know more about the behavior of the advantaged child.

II. METHOD AND PROCEDURE:

A. Sample:

The study population consists of the following samples: (1) Head Start sample: this group was comprised of 10 children in one class and 9 children in another class at the East Tremont YM-YWHA and of 17 children in one class at the Coney Island YM-YWHA. This makes for a total N = 36 children. At the Coney Island Y it was possible to include all the children in the class in the study, however at the East Tremont Y there were two problems. There were many Spanish speaking children who were excluded from the study because of the difficulty in testing them. Moreover, the turnover was so great that it was difficult to find children who had been there for at least five months. As was stated in the original proposal this was the minimal time period set in order to allow the children to stabilize their behavior patterns somewhat.

(2) Middle-class sample: this group with a total N of 32 was comprised of 18 children in one class and 14 children in the other class at the Emanu-El Midtown Y.

(3) Upper middle-class sample: this group with a total N of 40 was comprised of 10 children in one class, 14 in a second class, and 16 in a third class.

Table 1 represents the number of children in each sample and the mean age of each group.

Table 1. Distribution of Children Among Different Classes and the Mean Age of Each Group.

	N	MEAN AGE	SD in yrs.
HEAD START			
Class 1	9	4 yrs., 5 mos.	.269
Class 2	10	4 yrs., 9 mos.	.434
Class 3	17	4 yrs., 6 mos.	.244
Combined 1, 2, 3	36	4 yrs., 7 mos.	.346
MIDDLE-CLASS			
Class 1	14	5 yrs., 1 mo.	.276
Class 2	18	4 yrs., 5 mos.	.233
Combined 1 & 2	32	4 yrs., 8 mos.	.394
PRIVATE SCHOOL			
Class 1	10	4 yrs., 6 mos.	.230
Class 2	14	5 yrs., 1 mo.	.545
Class 3	16	4 yrs., 3 mos.	.477
Combined 1, 2, 3	40	4 yrs., 7 mos.	.549

Table 2. ANOVA for Ages of Head Start v. Middle-Class v. Private School Samples.

	ss	df	MS	F	P
Total	21.491	107			
Between	.159	2	.079	.389	NS
Within	21.332	105	.203		

As can been seen from an inspection of Table 2, there is no significant proportion of variance attributable to age and it is clear, therefore, that the groups are comparable in terms of age.

B. Data-gathering instruments and techniques:

Instruments used measured the children's behavior patterns and their intelligence and school readiness, in terms both of teacher's ratings and of objective test scores.

1. Children's intelligence measures:

As set forth in the study proposal, only the Peabody Picture Vocabulary Test (PPVT) was to be used as a measure of intelligence. However, due to the fact that the Stanford-Binet offers a measure of intelligence that does not rest solely on verbal ability, the complete Stanford-Binet was administered to each child whose intelligence was measured.

Hence, each Head Start child and each private school child was given both the PPVT and the Stanford-Binet, on an individual basis.

The Head Start teachers were asked to rank the children in their classes according to their estimates of the individual child's intelligence. They did this without having any knowledge of the children's measured IQs.

2. Measures of school readiness:

The Caldwell Inventory was presented to the Head Start and private school children as an objective measure of school achievement which, as Caldwell (1966) has suggested, appears to be related to what is generally termed school readiness.

The Head Start teachers were asked to rank the

children in their classes in terms of school readiness.

3. Behavior profiles:

An observation schedule developed by the Research Department of the Associated Ys for use in the structured observation of children's play groups has been reported upon elsewhere (Holmes, 1964, 1965, 1966). Some changes were made in this schedule on the basis of its prior use among Head Start children. The revised schedule and manual for its use are appended to this report. Each child in the study was observed for five 20-minute periods. The observations were scheduled on a random basis, so that the day and time of observation, and the person observing were randomly assigned. This was done so as to avoid, insofar as possible, bias stemming from raters and occasions.

All of the observations were completed prior to the administration of the other testing procedures, so that observations would not be biased on the basis of known IQ and Caldwell scores. Moreover, the observations were conducted by research assistants other than those who did the formal testing. The inter-rater reliability estimates between the two observers was .73 for Part A of the observation schedule and .87 for Part B, using Horst's correction.

The observation schedule consists of Part A and Part B. Part A requires that the observer rate every action in which the child engages, on an ongoing and continuous basis for the 20-minute period. Part B requires that the observer make one summary rating of behavior over the entire 20-minute period, using a modified Likert seven point scale, along a variety of dimensions.

For Part A of the schedule, scores consist of the number of instances in which each child is observed as engaging in each of the types of activity. The scores given each child consist, for the Likert-type items of Part B, of the scale value for each of the items.

III. STUDY HYPOTHESES:

Essentially there were two major hypotheses underlying those aspects of the study which deal with intelligence, and two parallel hypotheses underlying those aspects of the study which bear on school readiness.

Hypothesis I: The perceptions and evaluations of the Head Start children in terms of intelligence, by their teachers may be influenced erroneously by the degree to which the behavior patterns of the disadvantaged child approximate those of the middle-class children.

The specific predictions deriving from this hypothesis were as follows:

1. The r between the Head Start teacher's ratings of intelligence and actual intelligence as measured

by the Stanford-Binet will be low and non-significant.

2. In those instances where this prediction holds, there will be a significant difference between the behavior profile of those Head Start children rated as being the most intelligent and those children rated as least intelligent.

3. Those children who are rated by their teacher as being of high intelligence will show a behavior profile which is more similar to the middle-class profile than is the case with the children who are perceived as being of lower intelligence.

Hypothesis II: There is a relationship between intelligence and behavior, so that it is possible to estimate intelligence on the basis of observed behavior.

The specific prediction deriving from this hypothesis is as follows:

4. It will be possible, using multiple regression analysis, to develop a multiple correlation between actual IQ and the various behavior factors which emerge.

Hypothesis III: The perceptions and evaluations of the Head Start children, in terms of school readiness, may be influenced erroneously by the degree to which the behavior patterns of the disadvantaged child approximate those of the middle-class child.

The specific predictions deriving from this hypothesis are as follows:

5. The r between the Head Start teacher's ratings of school readiness, and of actual readiness as measured by the Caldwell Inventory, will be low and insignificant.
6. In those instances where Prediction 5 holds, there will be a significant difference between the behavior profile of those children rated as high on school readiness, and those rated as low on that variable.
7. The children who are rated by their teacher as being of high readiness will manifest a behavior profile which is more similar to the middle-class profile than is the case with the children who are perceived as being of relatively lower readiness.

Hypothesis IV: There is a relationship between school readiness and behavior, so that it is possible to estimate school readiness on the basis of observed behavior.

The specific prediction deriving from this hypothesis is as follows:

8. It will be possible, using multiple regression analysis, to develop a multiple correlation between actual school readiness, as measured by the Caldwell Inventory, and the various behavior factors which emerge.

These were the central predictions of the study. In addition the behavior profiles of the private school children were developed and compared to the Head Start and middle-class samples.

An examination of the above predictions would suggest that there are two global areas of investigation in this project: first, the relationship between intelligence, both actual and

perceived, and various patterns of behavior and, second, the relationship between school readiness, both actual and perceived, and the behavioral factors. Adhering to this dichotomization, the findings will be presented in two sections, the first dealing with the results pertaining to intelligence, and the second with the results pertaining to school readiness.

IV. RESULTS:

Part I: Intelligence:

Factor analyses were performed separately for Parts A and B of the observation schedule. The rotated, orthogonal factor loadings on the five Part A factors and the four Part B factors are shown below, along with the names assigned to each factor, in Tables 3A and 3B, respectively.

Table 3A. Observational Schedule Factor Loadings, Part A.

FACTOR	I	II	III	IV	V	h
I-Autonomous-Initiating Behavior						
Terminator - Self	.864	.104	.044	.375	.165	.927
Initiator of Act-Self	.858	.037	.216	.372	.035	.924
Who is involved-Self	.841	.048	.047	-.160	.332	.848
Orientation-External Manifest Goal Constructive	.841	.065	-.082	.031	.031	.720
Emotion-NoneApprop.	.742	.327	-.139	.112	.407	.855
Goal-Reached	.625	.400	.035	.568	.035	.875
II-Passive Responding Behavior						
Initiator of Act - Teachers/Adults	.076	.871	.047	-.104	.136	.796
Who is involved - Teachers/Adults	.244	.809	.081	.022	.011	.722
Orientation-Responding Constructive	.153	.773	.004	.19 1	.090	.668
Terminator-Other	-.025	.618	.234	.320	.275	.615
III-Social Destructiveness						
Orientation-External Manifest Goal Destructive	-.028	.050	.498	-.109	.046	.265
Orientation-Responding Destructive	.052	.163	.485	-.098	.003	.274
Goal-Not Reached	.308	.241	.448	.159	.022	.379
Orientation-Social Destructive	-.106	-.080	.408	.169	-.113	.225
Emotion-Positive Inappropriate	-.015	.002	.397	-.009	.025	.159
Emotion-Negative Appropriate	.057	.021	.342	.088	-.031	.130
IV-Social Constructiveness						
Who is involved-Peers	.058	.248	.113	.829	.248	.827
Orientation-Social Constructive	.225	-.056	.024	.725	-.169	.608
Initiator of Act-Peers	.175	.036	.007	.619	.436	.606
Emotion-Positive Appropriate	.035	.186	-.008	.525	-.156	.336
V-Non Purposive Behavior						
Goal-No Goal	.275	.096	-.037	-.014	.875	.852
Orientation-Responding Non-Purposive	.102	.199	-.154	-.101	.743	.636
Orientation-Non Purposive Random Act	.336	.022	.116	-.133	.487	.381

Table 3B. Observation Schedule Factor Loadings: Part B

FACTOR	I	II	III	IV	h
I - Task Oriented					
Investment of self in activity/greater-less great	.787	.132	.081	.158	.669
Goal Direct v. Random	.771	.221	.059	.151	.668
Long v. Short Attention Span	.737	-.047	.233	-.088	.608
Construct. v. Non-Constr. Play	.728	.041	.404	.053	.697
Success v. Failure	.644	.198	.222	.172	.533
Grace. v. Awkward	.399	.129	.311	.192	.309
II - Verbal Behavior and Reactivity					
Rich v. Sparse Verbal	.262	.763	.239	.092	.717
Verbal v. Non verbal	.194	.705	.349	.149	.679
Attention Seeking v. Autonomous	-.076	.656	-.284	.210	.561
Intelligible v. nonintelligible	.250	.629	.306	.100	.561
Intrusive v. Non Int.	.147	.614	-.218	.295	.564
Reacts v. non reacts to peripheral stim.	.085	.471	.065	.256	.298
III - SOCIALIZATION:					
Evoking Tchr Response Approp. v. inapprop.	.263	.203	.636	.046	.517
Response to Failure Constructive v. Non-constructive	.386	.044	.549	.054	.455
Evoking Peer Response Approp. v. Inapprop	.279	.234	.533	.230	.469
Tension Discharge: Frequent v. Seldom	-.205	.118	-.424	.059	.239
IV - AFFECTIVE INVOLVEMENT:					
Affect: happy v. not happy	.116	.173	.037	.831	.736
Affective Response: Overt v. Covert	.128	.246	-.075	.826	.764
Response to Sucess: Positive v. Indifferent	.130	.199	.013	.779	.663
Active v. Inactive	-.041	.230	-.259	.431	.344

Part B of the schedule was analyzed prior to Part A. At that point, it was not known whether the factorial structure of the interaction schedule would be the same across all of the study groups. Therefore, four separate factor analyses were completed on the Part B data: among the Head Start children, among the middle-class sample, among the upper middle-class sample, and for the total, aggregate data. Inspection of the resulting structure and item loadings indicated that the groups were approximately the same in terms of factorial structure. That is, the factors which emerged from each group were defined in terms of the same items, with approximately the same item-factor loadings. In view of this empirically-demonstrated similarity, no attempt was made to use the different factor analyses; instead, the data presented above in Table 3B are those representative of the analysis performed on the total study population.

Similarly, based upon the striking similarity among the study groups in terms of factorial structure, only one factor analysis of Part A of the schedule was undertaken, this time for the total study group, only.

As can be seen from an inspection of Table 3A, five factors emerged from the factor analysis performed on Part A of the observation schedule.

Factor I describes behavior which is essentially autonomous and self initiating. It describes an action initiated by the child in which he carries out the activity by himself, completes it without any particular display of affect, and terminates it by himself. If for instance, a child goes to

the shelf, gets out a puzzle, completes the puzzle himself, and then returns it to its place, this is the kind of behavior which would receive a high score on Factor I.

Factor II describes behavior which is essentially passive and responding in its orientation. It describes an action carried out by the child which is both initiated and terminated for him by an adult, and which is carried out in the company of, and at the direction of, that adult. If for instance, a teacher asks a child to clear the table after a snack, and he does so in conjunction with her and at her request, and then resumes his seat when told, this is the kind of behavior which would receive a high score on Factor II.

Factor III describes behavior which is essentially destructive and inappropriate. It describes an action executed by the child which is socially destructive in its goal, in response to an action initiated by someone else. The affect of the child is inappropriate, whether he looks angry or overjoyed as he carries out his destructive act. If for instance, a child is invited to play by another child, and he responds by destroying the other child's block-building while either looking angry or laughing uncontrollably, this is the kind of behavior which would receive a high score on Factor III.

Factor IV describes behavior which is characteristically social in its orientation. It describes an action initiated by a peer, executed in the company of at least one other child, with positive affect, and in a socially constructive manner. If for instance, a child is approached by another child who invites him to play house and if he joins in the activity and has a good time, this is the kind of behavior which receives a high score on Factor IV.

Factor V involves behavior which is essentially non-purposive and random. It describes an action which has no goal or purpose; it is non-directed behavior. If for instance, a child spends considerable time wandering around the classroom, making no contact either positive or negative with any object or person, this is the kind of behavior which would receive a high score on Factor V.

Turning now to an inspection of Table 3B, it can be seen that four factors emerged from the factor analysis of Part B of the observation schedule.

Factor I describes behavior which is essentially task oriented and which involves the competence skills of the child. A child who scores high on this factor over the five 20-minute observation periods, is a child who characteristically seeks out a task with which he becomes highly involved over a relatively long period of time, in a constructive manner. In addition, his skills are such that he has the coordination to execute the task successfully. If for instance, over a 20-minute period a child's main activity is to write the letters of the alphabet, which he is able to do competently, and with a great deal of concentration, this is the kind of behavior which would receive a high score on Factor I.

Factor II is a description of the child's style of communication and his reactivity to others. A child who scores high on Factor II is a child who relies heavily on verbal communication, and whose verbal communications tend to be very rich and full. In addition, he is likely to seek out others since he gains more satisfaction from verbal exchange than from autonomous play. When he is involved in autonomous play, he is highly

reactive to peripheral stimuli and tends to be quite instrusive into the affairs of others because of his reactivity and his need to communicate with them verbally. If for instance, a child is involved in doing a painting or a collage, and is clearly more involved in talking about what he is doing or in general conversation, than in the task itself, and if, in addition he leaves his place a number of times to comment on what others are doing and to see "what's happening," this is the kind of behavior which would receive a high score on Factor II.

Factor III describes behavior which involves the style of the child's interaction with others, his ability to ellicit a response from others appropriately, and his capacity to tolerate frustration. A child who scores high on Factor III is one who is quite capable of evoking a response appropriately from either his teachers or his peers. He is also a child whose socialization has reached a level where adequate frustration tolerance has been developed. He is able to respond to failure constructively, and in addition he does not discharge tension frequently, nor is tension discharge the major goal of his behavior. If for instance, a child over a 20-minute period is able to elicit responses either from his teacher or from his peers, a number of times, in an appropriate way, this is the kind of behavior which would receive a high score on Factor III.

Finally, Factor IV describes the affective behavior of the child. A child who scores high on Factor IV is one whose af-

fect tends to be quite overt and readily discernible. He looks distinctly happy and pleased most of the time, rather than neutral or unhappy. Similarly in response to success he shows joy and delight openly. He tends also to be a rather active child who in his affective exhuberance tends to become involved in a variety of activities. If for instance, a child participates in a number of activities with considerable enthusiasm and positive affect he would receive a high score on Factor IV.

The consistency of the factors and their intrinsic logic lend considerable validity to the observation schema. In addition, it is noteworthy that these factors held up for all groups, so that clearly the schema can be used in the observation of very different kinds of children. On the basis of this factor analysis the original schema has now undergone another revision for future use; those items not listed in Tables 3A or 3B will be deleted in future studies.

Following the completion of the factor analysis, the mean factor score for each child was then computed, i.e., each child's factor score represents the sum of his individual scores on those items subsumed under that particular factor, over the five occasions that he was observed, and divided by five. The mean factor score for each group is then the sum total of the mean scores on that factor for each of the children in that group, divided by the total number of children.

In order to permit comparisons to be drawn <u>among</u> factors, all factor scores were standardized, by subtracting the overall mean for each factor from each individual score and dividing by the standard deviation of the study population with regard

to that factor. The result of this routine standardization is that, taking the whole study group, the mean for each factor is zero, and the standard deviation is equal to 1. In order to avoid the negative individual scores which would arise from such a procedure, a constant 20 was added to each score, so that the resulting scores would give rise to a mean of 20, again with a standard deviation of 1.

Whereas the scores in Part B of the schedule represent summary ratings for each child at the end of a 20-minute period, and are therefore constant in number, the Part A means are related directly to the number of interactions observed. Thus, the Part A factor scores represent not only relative proportions of occurence among the categories, but also the total number of interactions taking place. In order to avoid the obvious bias reflective not of differences in behavior, but of differences in number of interactions, it was necessary to weight each subject's interaction factor scores in terms of the number of interactions manifested by each subject during each observation period. This was done as follows. The average number of interactions per observation period was calculated, for the entire study population. This average was then divided by the number of interactions occuring during each subject observation period; each factor score of each child then was multiplied by the resulting fraction, thus weighting the factor scores for the number of interactions observed. It is these weighted, standardized scores which are represented in the tables, and which served as the bases for analysis.

Table 4 represents the mean IQ scores on the Stanford-Binet and the PPVT for the Head Start and upper middle-class groups.

Table 4. Mean IQs and Standard Deviation on the Binet and PPVT for Head Start and Upper Middle-Class Groups.

	BINET			PPVT		
	N	Mean IQs	Standard Deviation	N	Mean IQs	Standard Deviation
Head Start	36	94.31	11.93	36	79.69	17.03
Upper Middle-Class	40	137.76	16.67	40	116.60	15.15

The r between the Binet IQ scores and the PPVT IQ scores among the Head Start sample is .73 which is significant at the .01 level. However, the r between the Binet and PPVT scores among the upper middle class sample is only .35 which is not statistically significant. This suggests that the PPVT is not a valid measure of IQ among this population. In other words, it seems that for children who score at the upper end of the intelligence scale, the PPVT is not a valid test. It is also important to note that the PPVT IQ is considerably lower than the Binet IQ for both samples, so that in Head Start studies where the PPVT is used alone and very low IQs are reported, this discrepancy between the tests should be taken into account. In any event, due to the apparent invalidity of the PPVT in the upper range of intelligence, and because the Stanford-Binet is a far more comprehensive measure, all further analyses involving IQ scores rely only on the Binet.

Table 5 represents the Spearman Rho correlation between actual Binet IQ and Perceived IQ. The actual IQs for each class of children were ranked according to score, i.e., the child with the highest IQ received a rank of 1, the next highest a rank of 2, etc. IQ scores within five points of each other were considered to be a tie. The perceived IQs were already represented as rankings, since that is exactly what the teachers were asked to do, i.e., each teacher ranked the children in her class according to her estimation of their intelligence.

Table 5. Spearman Rho Correlations for Actual Binet IQs and Teacher-Perceived IQs Among the Different Head Start Groups.

	N	r	t	P
Head Start				
Class 1 (East	9	.763	3.123	.01
Class 2 Tremont)	10	.752	3.226	.01
Class 3 (Coney Island)	17	.401	1.694	NS

As can be seen from inspection of Table 5, the first prediction, i.e., that the correlation between the Head Start teacher's ratings of intelligence and actual intelligence would be low, was not borne out for the 19 children located in East Tremont, but was borne out for the 17 children located in Coney Island. The significant correlations found among the East Tremont group well may be seriously inflated for the following reason. Due to the difficulties mentioned earlier, i.e., that a number of the children in each class were Spanish-speaking or had been in the program only a relatively very

short period of time, only nine out of 30 children in one class, and 10 out of 30 children in the second class, were used in the study. Therefore, the teachers were asked to rate only a truncated sample of the entire class; seemingly, it would be much easier to evaluate validly the relative intelligence of 9 or 10 children, than it would be to rate that of 30 children.

In any event, since the correlations are positive and significant, the East Tremont samples cannot be used to test predictions 2 and 3. Hence, only the group of 17 children at Coney Island was used for this portion of the analyses. However, it is important to note in terms of the generalizability of these findings to the rest of our Head Start population, that there were no significant differences between the Coney Island and East Tremont samples, along any dimension of observed behavior. Tables 6A and 6B show the mean factor scores on Parts A and B of the observation schedule for the two Head Start groups, t tests between the means, and P values.

Table 6A. Mean factor scores on Part A of the observation schedule for the Coney Island and East Tremont Head Start samples, t tests, and P values.

Part A

Factor	CONEY ISLAND			EAST TREMONT			t	P
	N	Mean	S.D.	N	Mean	S.D.		
I-(auton-init.	17	20.92	.656	19	19.31	.925	.059	NS
II-(Pass.-respond.)	17	20.03	.775	19	19.68	.632	.012	NS
III-(Soc. destruc.)	17	19.45	.483	19	19.99	1.236	.019	NS
IV-(Soc. construc.)	17	20.36	.162	19	20.43	.382	.001	NS
V-(non-purposive)	17	20.78	.851	19	20.20	1.032	.023	NS

Table 6B. Mean factor scores on Part B of the observation schedule for the Coney Island and East Tremont Head Start samples, t tests, and P values.

Part B

Factor	CONEY ISLAND			EAST TREMONT			t	P
	N	Mean	S.D.	N	Mean	S.D.		
I-(Task orient.)	17	16.22	7.378	19	14.74	5.12	.149	NS
II-(Verbal & react.)	17	16.13	6.676	19	8.77	5.63	.761	NS
III-(Socialization)	17	13.52	5.188	19	10.92	7.08	.593	NS
IV-(Affect)	17	15.75	4.693	19	16.66	6.36	.085	NS

In order to test prediction 2 the data were treated in the following manner. The observational data on the eight children at Coney Island who were rated as being the most bright by the teacher were compared with the data on those eight children who were rated as being the least bright. In order to have equal Ns in each of these sub-groups, the middle case was dropped.

F. tests for heterogeneity of variance were performed for the data on Parts A and B of the schedule. Since the Fs were not significant, homegeneity of variance has been assumed between that group which is rated above the median and that group which is rated below the median in intelligence.

Table 7 shows the mean factor scores on Part A of the observation schema, the individual t tests, and p values.

Table 7A. Mean factor scores on Part A of the observation schedule for those Coney Island Head Start children ranked by the teacher as above or below the median in intelligence, t tests between the means, and P values.

Part A

	Above median in ranked intell.			Below median in ranked intell.			t	P
	N	Mean	S.D.	N	Mean	S.D.		
I-(Autono-init.)	8	20.72	.65	8	21.10	.61	1.206	NS
II-(Pass.-respond.)	8	20.30	.83	8	18.76	.61	4.230	.005
III-(Soc. destruc.)	8	19.43	.44	8	19.46	.52	.125	NS
IV-(Soc. construc.)	8	20.39	.10	8	20.32	.20	.889	NS
V-(Non-purposive	8	20.34	.78	8	21.20	.69	2.336	.025

As can be seen from inspection of Table 7A, there are significant differences between those Head Start children who are perceived by their teachers as being above the median in intelligence, and those perceived as being below the median on two of the factors. Those children who are perceived as being above the median in intelligence score higher on Factor II (Passive responding) than do those children who are perceived as being below the median. In other words, those Head Start children who are rated as relatively more intelligent are more likely to carry out an activity which is initiated or terminated for them by the teacher, are more likely to follow directions, and are more likely to be cooperative in their interactions with the teacher, than are those children perceived by the teacher as being relatively less bright.

With regard to Factor V (Non-purposive behavior), those children perceived by the teacher as being less bright manifest more random non-purposive or non-directed behavior than do those children who are ranked as relatively more bright.

It seems, on the basis of the analysis of Part A of the schedule that prediction 2 is at least partially borne out, i.e., there is a significant difference in behavior between those Head Start children rated as being the most intelligent and those children rated as being the least intelligent. The relevant behaviors seem to be those which involve the willingness to respond to the teacher's direction, and the relatively lower incidence of aimless non-purposive behavior.

Table 7B shows the mean factor scores on Part B of the observation schedule, the individual t tests and the P values.

Table 7B. Mean factor scores on Part B of the observation schedule for those Coney Island Head Start children ranked by the teacher as above and below the median in intelligence, t tests between the means, and P values.

Part B

Factor	Above median in ranked intell.			Below median in ranked intell.			t	P
	N	Mean	S.D.	N	Mean	S.D.		
I-(Task orient.)	8	17.26	6.59	8	15.16	7.95	.575	NS
II-(Verbal & react.)	8	19.86	6.52	8	12.39	4.58	2.651	.01
III-(Socialization)	8	13.70	6.01	8	13.33	4.20	.143	NS
IV-(Affect)	8	18.29	4.97	8	13.21	2.54	2.574	.02

As can be seen from inspection of Table 7B, there are significant differences among those children who are ranked by the teacher as being above the median in intelligence and those ranked below for Factors II (Verbal Rehavior and Reactivity) and IV (Affective involvement).

A child who is rated as relatively "more intelligent" is more likely to be verbal, to seek attention, and to be more highly aware of what goes on around him when contrasted with his peers who are rated as less intelligent. In addition, he looks happy most of the time, shows joy openly, and is an active exhuberant participator.

It seems then, that the analysis of Part B of the schedule supports the analysis of Part A in terms of the second prediction, in that there are behavioral differences between those children who are perceived as being above or below the median by their teacher.

Since it has been determined that in the Coney Island Head Start group the teacher was making invalid estimates of intelligence, on the basis of her observations of the children's behavior it became important to determine whether or not there are any differences in behavior between those children who are above the

median in actual intelligence and those children who are below the median in actual intelligence.

Tables 8A and 8B show the mean factor scores, standard deviations, t tests and P values obtained for those children who are above and those who are below the median in actual intelligence. In addition, Table 8B shows the measured mean IQ for each group.

Table 8A. Mean factor scores on Part A of the observation schedule for those Coney Island Head Start children above and those below the median in actual intelligence, t tests, and P values.

Part A

Factor	Above median in actual intell.			Below median in actual intell.			t	P
	N	Mean	S.D.	N	Mean	S.D.		
I-(Auton-init.)	8	21.14	.40	8	20.69	.78	1.451	NS
II-(Pass.-respond.)	8	19.81	.55	8	20.25	.90	1.180	NS
III-(Soc. destruc.)	8	19.31	.32	8	19.58	.57	1.169	NS
IV-(Soc. construc.)	8	20.34	.11	8	20.37	.20	.372	NS
V-(Non-purposive	8	20.60	.73	8	20.95	.93	.838	NS

Table 8B. Mean factor scores on Part B of the observation schedule for those Coney Island Head Start children above and those below the median in actual intelligence, t tests, and P values. Mean Binet IQs for the two groups.

Part B

Factor	Above median in actual intell.			Below median in actual intell.			t	P
	N	Mean	S.D.	N	Mean	S.D.		
I-(Task orient.)	8	18.49	6.09	8	13.94	7.84	1.296	NS
II-Verbal & react.)	8	18.71	6.53	8	13.54	5.95	1.655	NS
III-(Socialization)	8	13.70	4.91	8	13.33	5.44	.143	NS
IV-(Affect)	8	17.06	4.75	8	14.44	4.25	1.331	NS
Binet IQ	8	104.50	4.09	8	86.51	10.54	4.503	.001

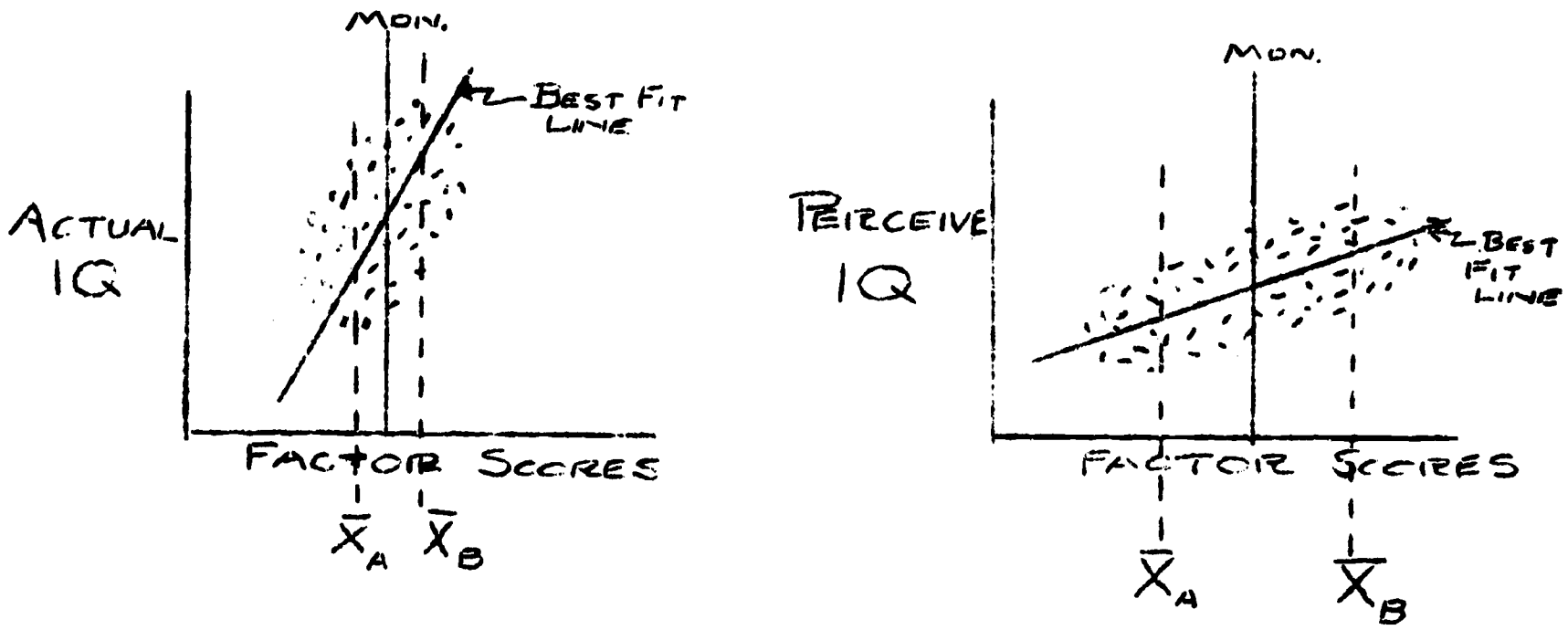
As indicated by an examination of the data presented in Tables 8A and 8B, the two groups are different in terms of Binet IQ as expected, but there were no differences between

children whose actual intelligence was above the median in this particular group, and those whose actual intelligence was below the median, in terms of any of the factor scores, in either Part A or Part B. That is, while there were differences in certain of the average factor scores between those perceived as more intelligent than those perceived as less intelligent, there is no actual such difference, again in terms of mean factor scores, between those who, as a group, are actually more intelligent that those who, again as a group, are actually less intelligent.

This finding is particularly important because it lends considerable support to a major assumption of the study, i.e., that those teachers who rank children incorrectly are responding to differences in the behavior patterns manifested by different children, rather than to actual intelligence or to behavior factors which are positively related to actual intelligence. This is not to say, however, that there is no relation between behavior and intelligence; as will be shown later, there very decidely is such a relationship, when expressed in correlational terms. However, when comparing the group means, there is no actual difference between the aggregate group with a higher IQ as contrasted with the aggregate group of relatively lower IQ. The following, hypothetical, simplified diagram may help to explain this seeming contradiction, i.e., that while a strong correlational relationship may exist between variables, there might be no difference in the average scores of one group and a second group, formed from a

larger group split at the median. As will be noted from this example, the mean of Group A (The more Intelligent) differs from that of Group B (The less Intelligent) only very slightly, in a context of considerable overlapping variance; however, the general shape of the distribution is such as to suggest that there is, in fact, a high positive correlational relationship between the two variables. Moreover, given such a strong relationship, it would seem that the inaccurate rater is, in effect, altering the slope of the regression line, or more simply the weight or "importance" accorded each of these factors, so that the mean in terms of perceived intelligence of the "A" group is considerably different from that of the "B" group, sufficiently different to give rise to an acceptable level of statistical significance when the means are compared.

Figure 1. Schematic representation of group scores.



In the aggregate, this particular finding, taken in conjunction with the preceeding diagram, suggests that those

teachers who rate their children inaccurately are giving improper weight to the various factors. This suggestion is borne out by the regression analyses conducted on these data and reported on later in this section.

Since prediction 2 was borne out, i.e., it seems that error in teachers' ratings is related to children's behavior, one can ask whether the behavior of the group which is rated above the median in intelligence is more similar to middle class behavior, than is the behavior of the group rated below the median. This, in fact, is prediction 3.

Tables 9A and 9B present the mean factor scores and standard deviations for Parts A and B of the observation schedule for the middle class comparison group, as well as for those Head Start children above and below the median in teacher ranked intelligence. The Head Start data are the same as those already presented in Tables 7A and 7B. They are re-presented here for the purposes of a more readily convenient comparison with the middle-class norms.

Table 9A. Mean Factor Scores of the Middle-Class Comparison Group and for those Coney Island Head Start Children Rated Above and Below the Median in Intelligence, on Part A of the Observational Schema.

PART A

	N	Mean	S.D.
Factor I: Autonomous-Initiating			
Middle-Class	30	19.80	.90
Head Start - Above mdn.	8	20.72	.65
Head Start - Below mdn.	8	21.10	.61
Factor II: Passive-Responding			
Middle Class	30	19.84	.69
Head Start - Above mdn.	8	20.30	.83
Head Start - Below mdn.	8	18.76	.61
Factor III: Socially destructive			
Middle Class	30	20.38	1.09
Head Start - Above mdn.	8	19.43	.44
Head Start - Below mdn	8	19.46	.52
Factor IV: Socially Constructive			
Middle Class	30	20.91	.31
Head Start - Above mdn.	8	20.39	.10
Head Start - Below mdn.	8	20.32	.20
Factor V: Non-Purposive			
Middle Class	30	19.52	.83
Head Start - Above mdn.	8	20.34	.78
Head Start - Below mdn.	8	21.20	.69

Table 9B. Mean Factor Scores for the Middle-Class Comparison Group and for those Coney Island Head Start Children Rated Above and Below the Median in Intelligence on Part B of the Observation Schema.

	N	Mean	S.D.
Factor I: Task-Oriented			
Middle Class	30	20.77	5.71
Head Start - Above mdn.	8	17.26	6.59
Head Start - Below mdn.	8	15.16	7.95
Factor II: Verbal Behavio:			
Middle Class	30	21.60	5.52
Head Start - Above mdn.	8	19.86	6.52
Head Start - Below mdn.	8	12.39	4.58
Factor III: Socialization			
Middle Class	30	16.81	7.73
Head Start - Above mdn.	8	13.70	6.01
Head Start - Below mdn.	8	13.33	4.20
Factor IV: Affect			
Middle Class	30	23.48	5.34
Head Start - Above mdn.	8	18.29	4.97
Head Start - Below mdn.	8	13.21	2.54

In Tables 10A and 10B are presented the results of the analysis of variance performed to determined the significance of the differences between the mean factor scores. A separate analysis of variance was performed for each of the factors, between the normative middle-class group and those Head Start children rated above the median on the one hand, and between the normative Middle-Class group and those Head Start childred rated below the median on the other.

Table 10. Anova for Each Factor of Part A of the Observation Schema Between: a) the Middle-Class Comparison Group and those Head Start Children Rated Above the Median in Intelligence and b) the Middle-Class Comparison Group and those Head Start Children Rated Below the Median.

Middle-Class vs. Head Start Children Rated Above Median Intelligence.

PART A

FACTORS	ss	df	ms	F	P
I: Auto-Inititiating					
Total	33.20	37			
Between	5.38	1	5.381	6.97	<.05
Within	27.82	36	.772		
II: Pass-Responding					
Total	21.107	37			
Between	1.323	1	1.323	2.407	NS
Within	19.784	36	.550		
III: Soc-Destructiveness					
Total	42.677	37			
Between	5.66	1	5.666	5.506	<.05
Within	37.017	36	1.028		
IV: Soc-Constructiveness					
Total	4.702	37			
Between	1.721	1	1.722	20.786	<.01
With	2.981	36	.083		
V: Non-Purposive					
Total	30.012	37			
Between	4.273	1	4.273	5.977	<.05
Within	25.739	36	.715		

Middle-class vs. Head Start children rated below median intelligence.

	ss	df	ms	F	P
FACTOR I: Autonomous-Initiating					
Total	38.114	37			
Between	10.700	1	10.700	14.051	.01
Within	27.414	36	.762		
FACTOR II: Passive-Responding					
Total	17.336	37			
Between	.043	1	.043	1	NS
Within	17.293	36	.480		
FACTOR III: Social Destructivness					
Total	43.038	37			
Between	5.381	1	5.381	5.144	.05
Within	37.657	30	1.046		
FACTOR IV: Social Constructiveness					
Total	5.474	37			
Between	2.232	1	2.232	24.802	.01
Within	3.242	36	.090		
FACTOR V: Non-purposive					
Total	42.549	37			
Between	17.879	1	17.879	26.093	.01
Within	24.670	36	.685		

It seems from inspection of Tables 9A and 10A that Prediction 3, as measured by Part A of the observation schema is not borne out. The differences between the middle-class normative group and the Head Start children rated above the median on intelligence, are as great as the differences between the middle-class normative group and those Head Start children who are rated below the median. However, it is most interesting to note that in the case of three of the

five factors, the differentials are very considerably higher when contrasting the below median Head Start children with the middle-class children. That is, the between group difference turns out to be far more significant among these three groups in comparing the middle-class children with the Head Start children rated as below the median in intelligence.

Table 10B. Anova for Each Factor on Part B of the Observation Schema Between: a) the Middle-Class Comparison Group and Those Head Start Children Rated Above the Median in Intelligence and b) the Middle-Class Comparison Group and Those Head Start Children Rated Below the Median.

Middle-Class vs. Head Start Children Rated Above Median Intelligence.

PART B

	ss	df	ms	F	P
FACTOR I: Task Oriented					
Total		37			
Between	77.848	1	77.848	1	NS
Within	3041.638	36	84.490		
FACTOR II: Verbal Reactivity					
Total	1276.292	37			
Between	18.994	1	18.994	1	NS
Within	1257.298	36	34.925		
FACTOR III: Socializaticn					
Total	2126.923	37			
Between	75.029	1	75.029	1.316	NS
Within	2051.895	36	56.997		
FACTOR IV: Affect					
Total	1222.604	37			
Between	170.287	1	170.287	5.826	.05
Within	1052.317	36	29.231		

Middle-Class vs. Head Start Children Rated Below Median Intelligence.

	ss	df	ms	F	P
FACTOR I: Task Oriented					
Total	1682.401	37			
Between	198.830	1	198.830	4.825	.05
Within	1483.571	36	41.403		
FACTOR II Verbal & React.					
Total	1620.973	37			
Between	535.635	1	535.635	17.767	.01
Within	1085.338	36	30.148		
FACTOR III: Socialization					
Total	2996.773	37			
Between	92.243	1	92.243	1.1433	NS
Within	2904.530	36	80.681		
FACTOR IV: Affect					
Total	1571.597	37			
Between	674.464	1	674.464	27.065	.01
Within	897.133	36	24.921		

As can be seen from an inspection of Tables 9B and 10B, Prediction 3 is essentially borne out in terms of Part B of the observation schedule. The comparison between the middle-class normative group and those Head Start children who are rated above the median, reveals that on three of the four factors of Part B there are no differences. The only significant difference between these groups emerges on Factor IV, or the affective component of behavior. It is clear from the mean factor scores and from the testing of Prediction 2 that the Head Start teacher rated those children who are more open in their affective expression as

being more intelligent. However, even these Head Start children do not approach the behavior of the normative middle-class group in this respect. In other words, the teacher has clearly used overtness of affective expression in her assessment of intelligence, but there is no group of Head Start children which is as openly expressive as the normative middle-class group.

While those children rated above the median in intelligence differ only in the overt expression of affect, those children rated below the median, differ significantly on three of the factors of Part B from the normative middle-class sample. The children in the below the median in rated intelligence group manifest less task-oriented competent behavior (Factor I), rely less on verbal communication and are less sensitive and reactive to their environment, (Factor II) and they are less overt in their affective expression (Factor IV). They very closely resemble their peers who are perceived as being above the median in their level of socialization. Neither group is significantly different from the middle class group, but the differences are in the predicted direction. In any event, it is clear that for at least this one teacher those children who are less task-oriented, less verbal, and less overtly expressive in terms of affect than a middle-class normative group, are those whom she perceives as being below the class median in intelligence.

The question finally arises whether, since behavior can be misleading when it influences the evaluation of intelligence, there is any relationship between actual intelligence and observed behavior and whether it is possible to predict intelligence from behavior. That this is possible is in fact the substance of Prediction 4. One of the major aspects of this investigation has been to develop a multiple correlation between actual intelligence on the one hand and the various dimensions of observed bheavior as expressed by the nine factor scores on the other.

Prior to the presentation of the actual multiple correlations which have been developed, it might be useful to present the $\bar{X}$ factor scores for all of the Head Start groups combined and all of the upper middle class groups combined since the multiple correlation coefficient between actual intelligence and behavior was developed for these two samples separately. As will be seen, the differences in behavior between the two groups are extensive and readily apparent. The mean factor scores for both groups are presented in Tables 11A and 11B. In addition, the Mean Binet IQ score for the Head Start and for the upper middle-class group originally presented in Table 4 are re-presented here for the readers' convenience.

Table 11A. Mean Factor Scores on Part A of the Observation Schedule and Mean Binet IQs for Total Head Start and Total Upper Middle-Class Samples.

FACTOR	HEAD START			UPPER MIDDLE-CLASS			t	P
	N	Mean	S.D.	N	Mean	S.D.		
I - Autonomous Initiating								
3	36	20.064	1.137	40	20.125	.674	.276	NS
II - Passive Responding								
	36	19.846	.724	40	20.250	.750	2.376	.01
III - Social Destruct.								
	36	19.732	.997	40	19.987	.858	1.180	NS
IV - Social Construct.								
	36	20.355	.405	40	20.673	.243	4.141	.005
V- Non-purposive								
	36	20.469	.993	40	19.959	.936	2.286	.025
BINET IQ	36	94.31	11.93	40	137.76	16.67	13.162	.001

Table 11B. Mean Factor Scores on Part B of the Observation Schedule for Total Head Start and Total Upper Middle-Class Samples.

FACTOR	HEAD START			UPPER MIDDLE-CLASS			t	P
	N	Mean	S.D.	N	Mean	S.D.		
I - Task Oriented								
	36	15.435	6.327	40	22.482	5.379	5.113	.005
II - Verbal & React.								
	36	12.232	7.193	40	23.039	5.229	7.282	.005
III - Socialization								
	36	12.141	6.390	40	19.883	5.818	5.413	.005
IV - Affect								
	36	16.231	5.077	40	21.006	6.044	3.695	.005

As can be seen from an inspection of Tables 11A and 11B, differences between Head Start and the upper middle-class private school sample are significant not only in terms of

intelligence, but also in almost every area of behavior. On Part A of the observation schedule, there is a significant difference in Factors II, IV, and V.

The private school children show a higher frequency of actions which are responding and adult initiated in their orientation (Part A, Factor II). This suggests that they are more responsive to and dependent on, the teacher for direction. This is supported by the findings pertaining to Part B, Factor II reported on below. It is also reflective of the very different expectations and programs of the Head Start versus the private school classroom. In the latter, the emphasis is far more on actual teaching of various skills and the teachers do initiate more for their children. In general, their orientation is to provide more structure and direction.

The private school children also show a higher frequency of socially constructive behavior (Part A, Factor IV). Their social behavior is more mature in terms, not only, of a greater number of peer contacts which occur, but in terms, also, of the quality of these contacts. The kind of genuine conversations or extensive exchanges of information which occurred among the private school children never were seen to occur among the Head Start children.

Perhaps in part as a function of the greater degree of structure provided by the enviornment, the private school children also engage in fewer instances of aimless non-purposive behavior (Part A, Factor V). They are less apt

to wander around the classroom without any goals or object in mind. This finding supports the overall impression that many of the Head Start children appear rather vague and as if they don't really have any sense of their own place in the classroom or of what they are doing there.

Part A of the schedule, being highly quantitative, leaves little room for qualitative judgments and impressions. In view of this, it is important to point out that while there is no quantitative difference between the Head Start and private school children in terms of Factor I, (Autonomous Initiating behavior) significant qualitative differences (as measured in Part B) were observed. For example, the difficulty level of tasks executed by the two groups was entirely different. If, for instance, a Head Start child does an eight piece puzzle, such an activity receives the same score on Factor I as does the activity of the private school child who may be engaged in working out a multiplication table.

Table 11B shows that on Part B of the observation schedule, which is more sensitive to qualitative differences, the differences between the two groups are significant on all factors. In terms of Factor I, private school children are more likely to seek out a task in which they become highly involved over a relatively long period of time and which they can perform with considerable competence. In general, the level of their skills is much greater than the level of skills of the Head Start children. In the three classes observed at the private school, virtually all of the

children can write, some of them can read with varying degrees of competence, and they all have great familiarity with numbers. Many of the four year olds can add, subtract, and multiply and they freely use equipment which is designed to teach them these concepts. It is clear that much more is expected of these children: the equipment is far more complex and challenging, and they are far more accustomed to becoming involved in school work.

In terms of Factor II, the private school children rely more heavily on verbal behavior than do the Head Start children: their verbalizations are richer, fuller, and more intelligible. They are also more reactive to the environment. In other words, they have learned to respond to the environment, rather than to shut it out as a form of irrelevant stimulation. However, their relative lack of autonomy is striking and worthy of comment. These children are far more likely to perceive the teacher and other children as a source of information and stimulation, hence they seek others out more and in this sense they are more dependent than are the Head Start children. For instance, it was observed frequently that a Head Start child would make a collage without ever consulting or showing his work to anyone at any time during or after the termination of his work; when he was finished he would simply put his work away, without comment, and seemingly with no expectation of, or interest in, praise. Among the private school children this kind of behavior was not observed.

These children elicit responses and opinions from the teacher and from the other children around them; they comment freely and with great interest on the productions of the other children around them. They function less on their own and demand more in terms of evaluation by others.

In terms of Factor III, it is apparent that the private school children have developed a more sophisticated repertoire of social skills. They are more capable of eliciting a response both from each other and from their teachers, in an appropriate way. Their frustration tolerance is also more adequately developed.

In terms of Factor IV, the private school children are far more open in their affective expression. Their participation in an activity is more enthusiastic and intense. It was observed in the Head Start classrooms that the children frequently showed no affect, regardless of what happened to them. They showed no pleasure in their own accomplishments and they seemed generally subdued and unresponsive. In group activities it was often difficult for the teacher to elicit any affective expression or active participation. When an activity was suggested, most of the children did what they were told but without any discernible facial expression. In marked contrast, the private school children showed a far greater capacity to express joy or sadness, to respond positively to their own sense of successful accomplishment, and the tenor of the classroom always seemed more lively and intense.

As is apparent from the foregoing discussion, the differences in behavior between these two groups are extensive and manifold. The differences in intellectual functioning, as measured by the Stanford-Binet also are very great. It remains to be seen whether the differences in intelligence and in behavior are related to each other and whether, as has been predicted, it is possible to predict intelligence from behavioral observations.

Tables 12A and 12B show the correlations between Binet IQ and the behavior factors for the Head Start and private school groups.

Table 12A. Pearson Product r Between Total Binet Scores and the Factor Scores on Part A of the Observation Schedule for Head Start and Private School Children.

PART A FACTOR	HEAD START		PRIVATE SCHOOL	
	r with Binet	P	r with Binet	P
I - (Autonomous initiating)	.265	NS	.278	NS
II - (Passive-responding)	.107	NS	-.139	NS
III - (Socially destruct)	.032	NS	-.116	NS
IVIV - (Socially Constructive)	.505	.01	-.087	NS
V - (Non-purposive)	-.198	NS	.016	NS

Table 12B. Pearson Product r Between Total Binet Scores and the Factor Scores on Part B of the Observation Schedule for Head Start and Private School Children.

PART B

FACTOR	HEAD START		PRIVATE SCHOOL	
	r with Binet	P	r with Binet	P
I - (Task Oriented)	.347	.05	.018	NS
II - (Verbal & React.)	.442	.01	.070	NS
III - (Socialization)	.268	NS	-.005	NS
IV - (Affect)	.418	.01	.029	NS

As shown by Tables 12A and 12B, few of the relationships between the Binet and individual factors are significant; it remains to be seen whether a multiple regression analysis can raise the correlation to a higher level of statistical significance.

Table 13 shows the multiple r for Parts A and B of the observation schedule with Stanford-Binet IQ for the Head Start and private school sample.

Table 13. Multiple r Between Actual IQ and Behavior Scores on Parts A and B of the Observation Schedule for Head Start, and Private School Samples.

	PART A			PART B		
	r 1.23456*	N	P	r 1.2345**	N	P
Head Start	.5994	36	.01	.5611	36	.01
Private School	.3107	40	.05	.0769	40	NS

* 1 = IQ; 2, 3, 4, 5, 6 = the 5 factors of Part A.
** 1 = IQ; 2, 3, 4, 5, = the 4 factors of Part B..

As can be seen from Table 13, the multiple r between IQ and the factor scores of the observation schedule is significant in all instances except for the private school group on Part B. As might be expected, the lack of any significant multiple r for this group reflects the lack of any individual significant correlation between any of the factors and the criterion (IQ) measure, accompanied by some fairly high factor intercorrelations, among this group only. For the Head Start sample it appears that with a knowledge of the behavior patterns of the children it is possible to predict IQ with a very fair degree of accuracy, once the regression equation is established. These equations, i.e., the Beta weights for each factor and the constant which is to be added to or subtracted from each equation, are presented below in Table 14. The Beta weights for the private school group were not developed for Part B of the schedule since the multiple r is not significant.

Table 14. Beta Weights for Multiple r Between Actual IQ and Behavior Scores on Parts A and B of the Observation Schedule for the Head Start and Private School Samples.

PART A FACTOR	HEAD START	PRIVATE SCHOOL
I - (Autonomous-Initiating)	+ 3.873	+ 5.811
II - (Passive-Responding)	+ 1.013	- 2.133
III - (Socially destructive)	+ .737	- 1.225
IV - (Socially constructive)	+15.889	- .351
V - (Non-Purposive)	- 3.769	- .573
Constant	-797.600	-62.896

PART A FACTOR	HEAD START	PRIVATE SCHOOL
I - (Task Oriented)	.529	-
II - (Verbal & React.)	.441	-
III - (Socialization)	.019	-
IV - (Affective)	.615	-
Constant	+22.138	-

It is clear that there is no simple linear relationship between intelligence and behavior which holds across the very different groups of children. As can be seen from Table 14, the various factors are given very different weight depending on the population under consideration. This means that behavior observed in one population simply does not has the same relative implications for intelligence, as does behavior observed in another population. For instance, among the Head Start population it is clear that socially constructive behavior (Part A, Factor IV) is given the greatest weight in predicting intelligence. This is not the case among the private school sample. This actually makes very good common sense. Among the Head Start population, where social interaction is on a far less sophisticated and less mature level, the ability of a child to socialize is apparently indicative of relatively greater intellectual capacity. Among the private school population the importance of this factor is virtually negligible, probably because it already has been achieved by all of the children and does not therefore discriminate among them.

Autonomous-initiating behavior (Part A, Factor I) apparently is quite important in the assessment of intelligence in both groups, although it is even more important for the private school group. It is interesting to note that among the Head Start sample, random non-purposive behavior contributes as much to intelligence scores in the

negative direction as autonomous-initiating behavior does in the positive direction.

For the Head Start group, it is noteworthy that responding behavior and socially destructive behavior contribute only minimally to an assessment of intelligence. This suggests that, within the context of Head Start, while passive responding behavior and lack of aggressiveness may make life easier for the teacher, she should not make the error of perceiving such children as being more or less intelligent. The data suggest that these factors have very little to do with intelligence for this population.

Turning now to the Beta weights for Part B of the observation schedule, it must be stated that the failure of Part B in predicting intelligence among the private school sample is not surprising. As can be seen from the mean IQ of 137 obtained for these children, this is an exceptionally bright group of children. It is more likely that variations in behavior in this population are a function of personality rather than intellectual differences. It seems that is is more difficult to predict intelligence on the basis of behavior among children, most of whom function at the "Very Superior" level of intelligence, than among children who function within "Dull Normal" or "Average" limits.

In terms of the Head Start sample, it is clear that affective behavior seems to be the most critical factor. It is consistent with clinical experience to find that the

child who is affectively expressive and unconstricted, is the same child who is attuned to the environment and available to information from it.

The presence of constructive task-oriented behavior, and the quality of the verbal behavior of the child and his reactivity to the environment must also be taken into account by the Head Start teacher who wishes to correctly assess intellectual status.

It is clear that what is needed is a replication study on a new sample in which Head Start children are observed using the observation schedule and in which the assigned Beta weights are used to predict intelligence. Following the testing of each of the children it would then be possible to determine whether the multiple r found in the present study holds for any urban Head Start group.

The implications for teacher training are of course, many. If the multiple r found in this study can be replicated, Head Start teachers could be trained to make a more realistic assessment of intelligence. They would know, for instance, that the advanced social behavior of a particular child, and the overtness of his affective expression, indicate that he is functioning at an intellectually higher level than his peers. Similarly, teachers would know that a child who is inappropriate in the way he seeks a response may be annoying and he may be a "pest", but that this behavior is not related to intelligence.

Before turning to Part II of this report and the results

bearing on school readiness, it might be useful to present a summary of the findings reported on in this section which bear on the relationship between intelligence and observed behavior.

1. A non-significant relationship between intelligence as rated by the teacher and intelligence as measured by the Stanford-Binet occurs with respect to one of the three Head Start teachers in the study. Hence, even though the other ratings were done in a "rarefied" climate as noted previously (rating only 9 or 10 of the 30 children, for example) a cautionary note has been introduced. The kind of errors made by this teacher may not be similar to those made by other teachers. This particular teacher was apparently influenced by the child's willingness to respond to directions, (Part A, Factor II); the relative absence of random non-purposive behavior, (Part A, Factor V); the child's verbal skills and awareness of his environment (Part B, Factor II); and by the overtness of his affective expression, (Part B, Factor IV). Since the Beta weights for Part A, for instance, indicate that Part A, Factor II (passive-responding) contributes relatively little and that other aspects of behavior such as social maturity (Part A, Factor IV) contribute greatly to an assessment of intelligence, it is easy to see why her assessment was incorrect.

2. Examination of the Head Start group in which the teacher's assessment of intelligence was invalid, disclosed clear differences in observed behavior between those children who were perceived by the teacher as being above the median in intelligence and those perceived as being below this median.
3. As compared to a normative middle-class sample, the Head Start children who are perceived by their teacher as being above the median in intelligence are different on fewer factors than those children who are below the median in perceived intelligence. In other words, the behavior of the children who are percevied as being brighter more nearly approaches the behavior of the normative middle-class sample.
4. There are no differences in behavior between those children who are above and those who are below the median in actual intelligence. The relationship between behavior and intelligence is a very complex one and if intelligence is to be correctly predicted on the basis of behavior, then the observed behavior must be weighted in a particular manner.

5. There is evidence to suggest that the prediction of intelligence on the basis of a qualitative description of behavior cannot be made among children of "Very Superior" intelligence. In other words, it seems

that the question of how "superior" is this child of "Very Superior" intelligence cannot be answered on the basis of behavioral observations, unless they are strictly quantitative.

6. In attempting to predict intelligence on the basis of behavior it is necessary to know exactly what population is being evaluated. Among the Head Start population the prediction of intelligence on the basis of behavior seems quite feasible; this finding needs to be replicated in a new study.

PART II - SCHOOL READINESS

It must be stated at the outset, that a question has been raised (by Caldwell herself, among others) as to whether or not the Caldwell Inventory represents a test of school readiness. It is clear that the test measures achievement or preparedness in the sense that it measures what a child has already mastered. As Caldwell (1965) has put it: "if this 'preparedness' represents 'readiness', then indeed the anticipated inventory was intended to measure school readiness." (p. 3). For the purposes of the present study, the Caldwell Inventory has been used as a measure of school readiness, in the expectation that some information about its behavioral correlates would be useful not only in the prediction of readiness, but in providing more knowledge about the Inventory itself.

Table 15 shows the means and standard deviations for the Caldwell Inventory scores for both the Head Start and private school groups.

Table 15. Mean Caldwell Inventory Scores and Standard Deviations for Head Start and Upper Middle-Class Groups.

Caldwell	Max. Poss. Score	HEAD START N	Mean	S.D.	PRIVATE SCHOOL N	Mean	S.D.	t	P
Comp.	18	36	12.80	3.05	40	17.23	1.45	7.92.	.001
Non-Verb.	19	36	11.94	3.24	40	17.10	1.51	8.73	.001
Numbers	19	36	6.61	2.60	40	14.00	3.15	11.17	.001
Verbal	24	36	10.61	3.23	40	19.10	3.38	11.17	.001
Total	80	36	41.70	10.10	40	67.40	7.22	12.69	.005

As is apparent, and hardly surprising, the upper middle-class group does significantly better than does the Head Start sample in every area measured by the Caldwell. In fact, our impression is that the test does not discriminate among children in our private school sample because it is far too easy and thus there is very little variation or spread among scores. In spite of this, it is interesting to note that the Pearson product moment correlation between the Binet and the Caldwell score is about the same for both the private school and Head Start groups. In the first instance $r = .587$; in the second instance $r = .534$. Both of these are significant at the .005 level.

Table 16 represents the Spearman Rho correlation between actual Caldwell Scores and perceived school readiness as rated by the teachers in the Head Start sample. Just as with the IQ scores described in the previous section, the actual

Caldwell scores were ranked according to score, i.e., the child with the highest Caldwell score received a rank of 1, the next highest a rank of 2, etc. The perceived readiness scores already were represented as rankings, since that is exactly what the teachers were asked to do, i.e., each teacher ranked the children in her class according to her estimation of their readiness for formal schooling.

Table 16. Spearman Rho Correlations for Actual Caldwell Scores and Teacher-Perceived School Readiness Among the Head Start Groups.

	N	r	t	p
HEAD START				
Class 1 (East Tremont)	9	-.125	.332	NS
Class 2 (East Tremont)	10	.952	8.797	.01
Class 3 (Coney Island)	17	.555	2.584	.02

As can be seen from an inspection of Table 16, Prediction 5, that the correlation between the Head Start teacher's ratings of readiness and actual readiness, as measured by the Caldwell, would be low, was borne out only for Class 1 at East Tremont. This group has only nine children. Since the testing of Predictions 6 and 7 requires a breakdown of those children who are rated as being above and below the median on readiness, this would mean that all the statistical tests would have to be performed with the highly unrespectable N of 4 in each sub-group. As such analyses would make no sense whatsoever, we had to veer from our original plan and to ask instead the following question. Considering across all Head

Start groups those children who score above the median on the Caldwell vs. those who score below the median, are there any differences in behavior, as measured by the observation schedule? In other words, the emphasis of the school readiness aspects of the study has shifted from the original attempt to understand why certain teachers are mistaken in their assessment of readiness, to an attempt to understand something about the correlates in behavior of a high or low school readiness score.

Tables 17A and 17B represent the mean factor scores on Parts A and B of the observation schedule for those Head Start children who score above and below the median on the Caldwell Inventory.

Table 17A. Mean factor scores on Part A of the Observation Schedule for Head Start children who score above and below the median on the Caldwell.

PART A

FACTOR	ABOVE MEDIAN			BELOW MEDIAN			t	P
	N	Mean	S.D.	N	Mean	S.D.		
I - (Auto-Initiating)	18	20.279	.871	18	19.804	1.336	1.192	NS
II - (Pass-Responding)	18	20.009	.696	18	19.726	.666	1.176	NS
III - (Soc-Destructive)	18	19.819	1.181	18	19.643	.837	.487	NS
IV - (Soc-Constructive)	18	20.379	.359	18	20.306	.251	.670	NS
V - (Non-Purposive)	18	20.126	.760	18	20.755	.969	2.044	NS

Table 17B. Mean factor scores on Part B of the Observation Schedule for Head Start children who score above and below the median on the Caldwell.

PART B

FACTOR	ABOVE MEDIAN			BELOW MEDIAN			t	P
	N	Mean	S.D.	N	Mean	S.D.		
I (Task Oriented)	18	16.686	6.657	18	14.06	5.859	1.183	NS
II (Verbal & Reactivity)	18	16.406	6.901	18	8.781	6.553	3.205	.01
III (Socialization	18	13.963	6.765	18	10.219	5.716	1.691	NS
IV (Affect)	18	19.234	4.451	18	13.851	4.037	3.583	.01

As can be seen from Tables 17A and 17B, the only significant difference, among Head Start children, between those who score relatively high on the Caldwell and those who score relatively low, is in verbal expressiveness and reactivity (Part B, Factor II) and overtness of affective expression (Part B, Factor IV). In other words, the Head Start child who is more verbal and who is more openly engaged, in terms of his emotional reactions, in the environment, is one who is apt to do better on the Caldwell.

It seems that it is relatively difficult to predict school readiness from behavior, without knowing what weights to assign to the various dimensions of behavior. Thus, it is relevant to ask what the weighted behavioral correlates of school readiness might be, and whether it is possible to predict school readiness on the basis of observed behavior, using weighted scores. This in fact is Prediction 8. The correlations between the Caldwell and the behavioral factors are presented in Tables 18A and 18B below for the Head Start and private school groups.

Table 18A. Pearson Product r between total Caldwell scores and the factor scores on Part A of the Observation Schedule for Head Start and private school children.

PART A

FACTOR	HEAD START r with Caldwell	P	PRIVATE SCHOOL r with Caldwell	P
I - (Auto-Initiating)	.271	NS	.054	NS
II - (Pass-Responding)	.044	NS	-.423	.005
III - (Soc-Destructive)	-.124	NS	.371	.025
IV - (Soc-Constructive)	.276	NS	.289	.05
V - (Non-Purposive)	-.290	NS	-.178	NS

Table 18B. Pearson Product r between total Caldwell scores and the factor scores on Part B of the Observation Schedule for Head Start and private school children.

PART B

FACTOR	HEAD START		PRIVATE SCHOOL	
	r with Caldwell	P	r with Caldwell	P
I - (Task-Oriented)	.407	.025	.245	NS
II - (Verbal & React)	.333	.050	.486	.005
III - (Socialization)	.355	.025	.053	NS
IV - (Affective)	.247	NS	.301	.050

As can be seen from an inspection of Tables 18A and 18B a number of relationships between the Caldwell and individual factors are significant; it remains to be seen whether a multiple regression analysis can raise the correlation to a higher level of statistical significance.

Table 19 shows the multiple r for Parts A and B of the observation schedule with total Caldwell score for the Head Start and private school samples.

Table 19. Multiple r between actual school readiness behavior scores on Parts A and B of the Observation Schedule for Head Start, and Private School Groups.

	PART A			PART B		
	r 1.23456*	N	P	r 1.23456**	N	P
Head Start	.586	36	.01	.456	36	.01
Private School	.572	40	.01	.5608	40	.01

* 1 = Caldwell score; 2, 3, 4, 5, 6 = the 5 factors of Part A.
** 1 = Caldwell score; 2, 3, 4, 5 = the 4 factors of Part B.

Examination of Table 19 shows that multiple r for Part A of the observation schedule among the Head Start group is significant. As can readily be seen from an inspection of Table

18A, none of the individual factors on Part A was significantly related to school readiness among the Head Start group. Hence, while no individual factor is powerful enough to predict school readiness by itself, the combined scores of all the factors can do the job of prediction. In terms of Part B of the schedule, for this population, although multiple r is significant, it is not substantially greater than the r between school readiness and task-oriented behavior.

Multiple r for Part A of the observation schedule among the private school group is also significant, and is considerably greater than the r between school readiness and any single behavior factor. Similarly, with Part B, multiple r is significant and greater than the correlation with any single behavioral factor.

It appears then, that with a knowledge of the behavior patterns of a child it is possible to predict the child's readiness for school with a fair degree of accuracy. In order to do this, it is necessary to know what weight to assign to each factor. The Beta weights for each factor as well as the constant which is to added to each equation, are presented below in Tables 20A and 20B.

Table 20A. Beta Weights for Multiple r Between Actual School Readiness and Behavior Scores on Parts A and B of the Observation Schedule for the Head Start and Private School Groups.

PART A

FACTOR	HEAD START	PRIVATE SCHOOL
I - (Auton-Initiating)	+ 3.208	+ 1.675
II - (Pass-Responding)	+ .222	- 1.485
III - (Soc-Destructive)	- 1.437	+ 3.564
IV - (Soc-Constructive)	+ 2.804	+ 5.177
V - (Non-Purposive)	- 6.388	- .672
Constant	+124.173	-348.100

Table 20B. Beta Weights for Multiple r between Actual school Readiness and Behavior Scores on Parts A and B of the Observation Schedule for the Head Start and Private School Groups.

PART B

FACTOR	HEAD START	PRIVATE SCHOOL
I - (Task-Oriented)	+ .348	- .261
II - (Verbal & React)	+ .266	+ .726
III - (Socialization)	+ .076	+ .333
IV - (Affective)	+ .403	+ .156
Constant	+ 6.719	+ 17.708

Examination of the Beta weights suggests that among the Head Start children non-purposive random behavior contributes the most, in a negative way, to the prediction of school readiness. In other words, those children who exhibit relatively more non-purposive aimless behavior are those children who are relatively less ready for school. The next most important factors are autonomous-initiating behavior and socially constructive behavior. Those children who show relatively more autonomy and initiative, and who are socially constructive are relatively more ready for school. The presence of socially destructive behavior is negatively related to school readiness, but this is not a major contributor. Similarly, any consideration of passive-responding behavior is relatively meaningless in the assessment of school readiness.

In terms of Part B, those Head Start children who are affectively expressive, who engage in task-oriented behavior, and who are verbally expressive and reactive are more apt to be ready for school, than children who evidence fewer of these behaviors.

Among upper middle-class children, the most important variable seems to be social behavior, in both its constructive and destructive forms. This relationship between both socially constructive and destructive behavior on the one hand and school readiness on the other, is striking in light of the finding reported in Part I that these behaviors were unrelated to intelligence, among this population. Apparently, social behavior in its positive and negative forms is positively related to school readiness, but not to intelligence among upper middle-class children. The presence of autonomous-initiating behavior contributes about as much in a positive direction to school readiness as does passive-responding behavior in a negative direction. The contribution of random non-purposive behavior, which is so important among the Head Start children, is virtually negligible here.

In terms of Part B of the schedule, by far the most important variable is the verbal expressiveness and reactivity of the child. It is again noteworthy, that while there is no relationship in this population between intelligence and verbal expressiveness, there is this very high correlation between verbal expressiveness and reactivity and school readiness. One can speculate that in this sample of highly "superior" children, verbal expressiveness is not relevant to intelligence because there may be a number of highly intelligent children who are, however, quiet, reflective, and directed more by their inner processes than by their reactivity to the environment. While it is true that those children

may be very intelligent, it also seems apparent that in spite of their intelligence, their relative lack or responsiveness to the environment lowers their Caldwell scores. Hence, while among Head Start children, silence and the relative absence of verbal communication is related to relatively lower intelligence; among upper middle-class children such behavior seems to bear no relationship to intelligence.

It may be helpful at this point to summarize the findings with respect to school readiness.

1. A non-significant relationship between school readiness as rated by the teacher, and school readiness as measured by the Stanford-Binet, occurs with respect to one of the three Head Start teachers in this study. Since the number of children in that group was too small for any statistical analysis, it was not possible to determine the kinds of bias which may have influenced her ratings.
2. Among the entire Head Start sample, there are very few differences in behavior between those Head Start children who are above and below the median in terms of actual school readiness. The differences which do emerge suggest that the child who is actually more school-ready is more

verbally expressive and reactive to his environment; he is also more affectively reactive and communicative.

3. It is possible to predict school readiness from behavior, if the relative weights to be assigned to various forms of behavior are used.

4. The prediction of school readiness does not involve the same weighting of behavior as the prediction of intelligence. In other words, it is quite clear that independent judgments of these two variables must be made, and that the behavior which is predictive of intelligence is not necessarily predictive of school readiness.

5. As was the case with the prediction of intelligence, when it comes to predicting readiness for school, one population is not like another and behavior in one group does not mean the same and cannot be given the same weight as behavior in another group. Among the Head Start children the most important behavioral variables involve the absence of random non-purposive behavior, and the presence of autonomous-initiating, socially constructive, and affectively expressive behavior. Among the upper middle

class children the most important behavioral variables seem to involve the presence of social behavior of any kind, either constructive or destructive, and verbal expressiveness and reactivity to the environment. These findings and the particular beta weights need to be replicated in a new study.

V - CONCLUSIONS AND IMPLICATIONS

1. Perhaps the most important finding is the one which shows that behavior which is important or adaptive among one group of children is not necessarily as important in another group. For instance, among Head Start children it seems that it is very important to look for a relative absence of aimless non-directed behavior when the child's readiness for school is under consideration. However, among an upper middle class group this aspect of behavior is relatively unimportant. Hence, a teacher has to look for different cues; what she looks for, and how she weights what she sees, has to vary according to the population being assessed.

While these findings are tentative and must await replication, it does seem that patterns of behavior are quite different in different populations. Hence, the Head Start teacher who would accurately assess school readiness must look for a relative lack of non-purposive aimlessness. Similarly, she must note the presence of autonomous-initiating behavior. In other words she has to observe carefully

to see whether this is a child who starts activities for himself and carries them out independently. She must also observe for expressions of socially constructive and affiliative behavior, and for the presence of affectively expressive behavior. Thus, the Head Start child who is relatively ready for school is one who knows what he is doing in the classroom, who can relate effectively to other children, and whose affective participation in the environment is sufficiently open to allow new learning and development to take place.

On the other hand, the teacher of highly advantaged children must look for other kinds of evidence in her attempts to assess school readiness. She must look for evidence of sociability and relatedness to others. She can also place considerable emphasis in her evaluation on the child's ability to express himself verbally; on his intelligibility and on the richness of his communication. As part of his verbal expressiveness she also needs to focus on the extent of his reactivity to, and intrusiveness in, the environment.

2. Another important finding is that just as behavior which is particularly relevant to an assessment in one group is not relevant to assessment in another, so behavior which is relevant to an assessment of one variable such as school readiness is not relevant to the assessment of another, such as intelligence.

The Head Start teacher who wishes to assess the relative intelligence of her students apparently must give enormous weight to the presence of socially constructive behavior.

Similarly, she has to take into account the presence of affectively expressive behavior, of successful task-oriented behavior, and of good verbal skills. The teacher of highly intelligent advantaged children must look for the presence of autonomous-initiating behavior. However it seems, at least on the basis of this first attempt, that it is very difficult to predict intelligence from behavior among this population.

3. Finally, it has been shown tentatively that school readiness and intelligence can be predicted from behavior with varying degrees of success. Therefore, it should be possible for teachers in most instances, given proper training, (our observyrs were trained for considerably less than a week), to make accurate assessments of their pupils' intellectual capacity and level of achievement. This is critical for their functioning as educators who then can address themselves to each child at his particular level of ability and understanding. In the face of accurate assessment, there would be far fewer instances of children who are sent to first grade routinely, but before they actually are ready to participate in formal education. There are children who need an extra year of experience in social skills and in verbal expressiveness before they can master what is expected of them in first grade. It is far better that they should have that extra year, than that they should be doomed to begin their formal education unprepared. All too frequently, this lack of readiness results only in a

child's failure, which then leads to the familiar spiral of frustration--->constricted withdrawal or aggressive acting out--->dropping out of school.

Similarly, the accurate assessment of the child's level of intellectual functioning is very important. All too readily a child may be given tasks which are beyond his comprehension at the time, or he may be given tasks which are too easy and not sufficiently stimulating. There are many Head Start children who are functioning at a level which is high enough to begin the early rudiments of formal education. In spite of their disadvantaged status, they are ready to be intellectually stimulated and the "head start" for them can be of a very different order than the "head start" given to children who are less intellectually and socially mature. Not every child in the Head Start classroom is at the same level of intellectual functioning, and this too must be taken into account by their teachers.

It is clear, as has been stated many times, that considerable further research needs to be done in order to validate these findings. Once this is done, it will be important to make a start in the direction of tailoring the Head Start programs to the needs of individual children, with widely discrepant skills, levels of functioning, and of achievement.