DOCUMENT RESUME

ED 082 857 PS 006 949

AUTHOR Walker, Debbie Klein; And Others
TITLE The Quality of the Head Start Planned Variation Data. Volume II.
INSTITUTION Huron Inst., Cambridge, Mass.
SPONS AGENCY Office of Child Development (DHEW), Washington, D.C.
REPORT NO OCD-H-1926
PUB DATE 30 Aug 73
NOTE 281p.; For Volume I, see PS 006 948

EDRS PRICE MF-$0.65 HC-$9.87
DESCRIPTORS Achievement; Affective Tests; Cognitive Measurement; Compensatory Education; Data Analysis; *Evaluation; Family Background; Intelligence Tests; *Measurement Instruments; Predictive Validity; *Preschool Programs; Tables (Data); Task Performance; *Test Reliability; *Test Validity
IDENTIFIERS Planned Variation; *Project Head Start

ABSTRACT

This publication continues the descriptions of the cognitive, psychomotor, and socioemotional measures used in all years of the Head Start Planned Variation Evaluation study. Included is a detailed examination of each measure, a discussion of the theory behind it, and a review of the available data on the measure's reliability, validity and other technical qualities. The last half of the document contains appendices relating to the procedures used in examining the quality of the data: (a) Test-Retest/Inter-Tester Reliability Study; (b) Eight-Block Sort Reliability Study; (c) Classroom Behavior Inventory Test-Retest Reliability Study; (d) Coding Reliability Study; (e) Classroom Information Form Reliability Study; (f) Parent Information Form Test-Retest Reliability Study; and (g) Quality of the Testing Procedure. Data tables are included. (SET)

THE QUALITY OF THE HEAD START PLANNED VARIATION DATA

DEBBIE KLEIN WALKER
MARY JO BANE
ANTHONY BRYK

VOLUME II
(pages 270-547)

August 30, 1973

HURON INSTITUTE
CAMBRIDGE, MASSACHUSETTS

This document was prepared for Grant # H 1926 from the Office of Child Development, Department of Health, Education and Welfare, U. S. Government. The conclusions and recommendations in this report are those of the grantee and do not necessarily reflect the views of any federal agency.

The project director was Marshall S. Smith.

Project staff included:

Barbara Behrendt
John Butler
Tom Cerva
David Cohen
Jane David
Richard Elmore
Helen Featherstone
Nathan Fox
David Gordon
Debbie Gordon
Sharon Hauck
Gregory Jackson
Carol Lukas
Robert McMeekin
Anne Monaghan
David Napior
Ann Taylor
Herbert Weisberg
Jack Wiggins
Cicero Wilson
Cynthia Wohlleb
Joy Wolfe
Stanley Yutkins
Diane Zipperman

TABLE OF CONTENTS

NYU Booklets 3D and 4A

Purpose

The NYU Booklets were designed to measure areas of preschool achievement. Booklet 3D is designed to measure achievement in pre-math, pre-science and linguistic concepts. Booklet 4A is designed to measure achievement in shape, numeral and alphabet names. These concepts are taught in many preschool programs in the Head Start Planned Variations Study.

Description

Booklet 3D is composed of the following three subtests:

1. Pre-math relational concepts. Seven items assess basic concepts of quantity and serial relation. Examples are "Point to the boy who has all the balloons" and "Point to the closed door."

2. Pre-science relational concepts. Seven items assess the basic concepts of "dry", "young", "short", "thin", "far away", "wide", and "old".

3. Linguistic concept of prepositions. Five items assess the understanding of the prepositional phrases of physical relation: "over", "behind", "down", "away", and "against".

Booklet 4A is composed of the following three subtests:

1. Alphabet names. Nine items assess the child's recognition of printed capital letters.

2. Numeral names. Six items assess the child's knowledge of numerals.

3. Shape names. Three items assess the child's recognition of shapes: heart, diamond and rectangle.

Each item correct in each booklet is scored "1". The total maximum score is 19 for Booklet 3D and 18 for Booklet 4A. Summary scores and a set of scores, obtained by considering the three subtests as criterion-referenced measures, were used in the data analysis (see Smith, 1973).

Development of Instrument

The subtests of Booklets 3D and 4A are shortened versions of six Early Childhood Inventories which are being developed under the joint directorship of A. Coller and J. Victor at the Institute for Developmental Studies at the New York University School of Education. The Early Childhood Inventories have been developed to be easily administered, easily scored, and appropriate for disadvantaged children. At the present time there are 17 inventories available in experimental forms. In addition to the six being used in the HSPV Study there are inventories to measure body parts' names, color names, classroom objects' names, quantity matching, set matching, same and different relationships, lower case alphabet letters, comparatives concepts and superlatives concepts (Coller and Victor, 1971). Since these inventories are still in experimental stages, there is no technical information available.

Norms.

There is no original norming sample for the NYU Booklets since they were in experimental form before being adapted for use in the Head Start Planned Variation Study.

Norms for raw scores are available for the Fall 1970 HSPV sample for Booklet 3D (Table 1 - 7) and for Booklet 4A (Tables 8 - 14). Norm tables based on three month age divisions (ten groupings from 42-44 months to 69-71 months) give the number of children, the mean score and the standard deviation at each age level for the following groupings in the HSPV sample: total (3D - Table 1, 4A - Table 8), females (3D - Table 2, 4A - Table 9), males (3D - Table 3, 4A - Table 10), children with no previous preschool experience (3D - Table 4, 4A - Table 11), children with previous preschool experience (3D - Table 5, 4A - Table 12), black children (3D - Table 6, 4A - Table 13), and white children (3D - Table 7, 4A - Table 14). The mean Booklet 3D score for the total Fall 1970 sample was 11.849 (S.D. = 3.277, N = 2161), while the mean Booklet 4A score was 5.645 (S.D. = 3.273, N = 2150). In general, scores on both booklets increase with age, are higher for white children than for black children, and are higher for children with previous preschool experience than for children with no previous preschool experience.

TABLE 1

DISTRIBUTION OF NYU BOOKLET 3D SCORES FOR ALL CHILDREN IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	10	8.700	3.348
45-47	61	10.607	3.413
48-50	207	9.957	3.085
51-53	314	10.965	3.204
54-56	355	11.487	3.202
57-59	350	11.860	3.034
60-62	274	12.500	2.994
63-65	230	12.730	3.120
66-68	182	13.159	2.819
69-71	178	13.433	3.270
TOTAL	2161	11.849	3.277

[1]Includes all children not in Level I sites, Oraibi, or Fresno, who had adequate age information on sex, age, race, and preschool experience.

[2]Maximum score = 19

TABLE 2

DISTRIBUTION OF NYU BOOKLET 3D SCORES FOR FEMALES IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	5	7.000	3.521
45-47	32	10.969	2.995
48-50	102	10.392	2.958
51-53	151	11.040	3.092
54-56	179	11.665	3.232
57-59	162	12.228	2.855
60-62	140	12.157	2.824
63-65	116	13.043	3.182
66-68	89	13.393	2.807
69-71	86	13.360	2.965
TOTAL	1062	11.994	3.171

[1]Includes all children not in Level I sites, Oraibi, or Fresno, who had adequate age information on sex, age, race, and preschool experience.

[2]Maximum score = 19.

TABLE 3

DISTRIBUTION OF NYU BOOKLET 3D SCORES FOR MALES IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	5	10.400	2.059
45-47	29	10.207	3.782
48-50	105	9.533	3.147
51-53	163	10.896	3.302
54-56	176	11.307	3.162
57-59	188	11.543	3.146
60-62	134	12.858	3.122
63-65	114	12.412	3.023
66-68	93	12.935	2.812
69-71	92	13.500	3.531
TOTAL	1099	11.709	3.370

[1] Includes all children not in Level I sites, Oraibi, or Fresno, who had adequate age information on sex, age, race, and preschool experience.

[2] Maximum score = 19.

TABLE 4

DISTRIBUTION OF NYU BOOKLET 3D SCORES FOR ALL CHILDREN WITH NO PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	10	8.700	3.348
45-47	55	10.527	3.274
48-50	185	9.908	3.124
51-53	285	10.979	3.184
54-56	293	11.372	3.279
57-59	290	11.707	3.094
60-62	204	12.299	2.936
63-65	162	12.660	3.069
66-68	135	12.978	2.835
69-71	130	12.977	3.209
TOTAL	1749	11.638	3.263

[1] Includes all children not in Level I sites, Oraibi, or Fresno, who had adequate age information on sex, age, race, and preschool experience.

[2] Maximum score = 19.

TABLE 5

DISTRIBUTION OF NYU BOOKLET 3D SCORES FOR ALL CHILDREN WITH PREVIOUS RRESCHOOL EXPERIENCE IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	---	-----	-----
45-47	1	11.000	-----
48-50	15	9.933	2.695
51-53	16	11.063	3.749
54-56	42	11.738	2.769
57-59	44	12.477	2.659
60-62	59	12.864	3.132
63-65	66	13.015	3.188
66-68	42	13.833	2.590
69-71	45	14.756	3.156
TOTAL	330	12.855	3.207

[1]Includes all children not in Level I sites, Oraibi, or Fresno, who had adequate age information on sex, age, race, and preschool experience.

[2]Maximum score = 19.

TABLE 6

DISTRIBUTION OF NYU BOOKLET 3D SCORES FOR BLACK CHILDREN IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	3	12.000	0.816
45-47	38	10.737	3.160
48-50	127	9.929	2.927
51-53	171	10.585	3.149
54-56	188	11.277	3.274
57-59	189	11.646	3.157
60-62	107	12.019	2.845
63-65	107	12.822	2.912
66-68	84	12.786	2.695
69-71	105	12.848	3.411
TOTAL	1119	11.543	3.234

[1] Includes all children not in Level I sites, Oraibi, or Fresno, who had adequate age information on sex, age, race, and preschool experience.

[2] Maximum score = 19.

TABLE 7

DISTRIBUTION OF NYU BOOKLET 3D SCORES FOR WHITE CHILDREN IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	6	7.500	3.202
45-47	22	10.818	3.284
48-50	69	10.116	3.264
51-53	114	11.491	3.264
54-56	124	11.734	3.228
57-59	116	12.224	2.983
60-62	110	13.173	3.048
63-65	86	13.523	2.823
66-68	61	14.344	2.828
69-71	51	14.961	2.737
TOTAL	759	12.403	3.366

[1]Includes all children not in Level I sites, Oraibi, or Fresno, who had adequate age information on sex, age, race, and preschool experience.

[2]Maximum score = 19.

TABLE 8

DISTRIBUTION OF NYU BOOKLET 4A SCORES FOR ALL CHILDREN IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	10	3.200	2.561
45-47	61	4.869	2.670
48-50	205	4.805	2.568
51-53	309	5.107	2.603
54-56	353	5.742	3.341
57-59	348	5.807	3.291
60-62	276	6.062	3.708
63-65	230	5.530	3.275
66-68	181	6.127	3.367
69-71	177	6.452	3.764
TOTAL	2150	5.645	3.273

[1] Includes all children not in Level I sites, Oraibi, or Fresno, who had adequate age information on sex, age, race, and preschool experience.

[2] Maximum score = 18.

TABLE 9

DISTRIBUTION OF NYU BOOKLET 4A SCORES FOR FEMALES IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	5	3.200	3.187
45-47	32	4.688	1.895
48-50	101	5.188	2.379
51-53	149	5.114	2.445
54-56	179	5.492	2.966
57-59	161	5.851	3.221
60-62	141	6.000	3.696
63-65	116	5.440	3.249
66-68	89	6.416	3.496
69-71	86	6.535	3.669
TOTAL	1059	5.653	3.156

[1]Includes all children not in Level I sites, Oraibi, or Fresno, who had adequate age information on sex, age, race, and preschool experience.

[2]Maximum score = 18.

TABLE 10

DISTRIBUTION OF NYU BOOKLET 4A SCORES FOR MALES IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	5	3.200	1.720
45-47	29	5.060	3.311
48-50	104	4.433	2.688
51-53	160	5.100	2.741
54-56	174	6.000	3.669
57-59	187	5.770	3.350
60-62	135	6.126	3.719
63-65	114	5.623	3.299
66-68	92	5.848	3.213
69-71	91	6.374	3.851
TOTAL	1091	5.636	3.383

[1] Includes all children not in Level I sites, Oraibi, or Fresno, who had adequate age information on sex, age, race, and preschool experience.

[2] Maximum score = 18.

TABLE 11

DISTRIBUTION OF NYU BOOKLET 4A SCORES FOR ALL CHILDREN WITH NO PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	10	3.200	2.561
45-47	55	4.727	2.260
48-50	183	4.716	2.615
51-53	282	5.110	2.526
54-56	291	5.478	2.992
57-59	288	5.691	3.223
60-62	206	5.854	3.419
63-65	152	5.352	3.202
66-68	134	5.948	3.175
69-71	129	6.333	3.665
TOTAL	1740	5.469	3.083

[1]Includes all children not in Level I sites, Oraibi, or Fresno, who had adequate age information on sex, age, race, and preschool experience.

[2]Maximum score = 18.

TABLE 12

DISTRIBUTION OF NYU BOOKLET 4A SCORES FOR ALL CHILDREN WITH PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	---	-----	-----
45-47	1	5.000	-----
48-50	15	5.533	2.276
51-53	14	5.643	3.772
54-56	42	6.095	4.017
57-59	44	5.386	2.357
60-62	59	6.220	4.423
63-65	66	6.015	3.436
66-68	42	6.619	3.909
69-71	45	6.644	3.854
TOTAL	328	6.101	3.704

[1] Includes all children not in Level I sites, Oraibi, or Fresno, who had adequate age information on sex, age, race, and preschool experience.

[2] Maximum score = 18.

TABLE 13

DISTRIBUTION OF NYU BOOKLET 4A SCORES FOR BLACK CHILDREN IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score	S.D.
42-44	3	4.333	2.055
45-47	38	5.158	2.434
48-50	126	4.849	2.520
51-53	170	4.882	2.459
54-56	188	5.564	3.336
57-59	189	5.497	3.230
60-62	108	5.259	3.348
63-65	107	5.187	3.383
66-68	84	5.512	3.393
69-71	104	5.962	3.838
TOTAL	1117	5.319	3.165

[1]Includes all children not in Level I sites, Oraibi, or Fresno, who had adequate age information on sex, age, race, and preschool experience.

[2]Maximum score = 18.

TABLE 14

DISTRIBUTION OF NYU BOOKLET 4A SCORES FOR WHITE CHILDREN IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	6	2.000	2.082
45-47	22	4.591	2.871
48-50	68	4.515	2.361
51-53	112	5.402	2.846
54-56	124	5.847	3.391
57-59	115	6.522	3.569
60-62	110	6.745	4.069
63-65	86	6.256	3.275
66-68	60	6.950	3.481
69-71	51	7.294	3.696
TOTAL	754	6.060	3.478

[1]Includes all children not in Level I sites, Oraibi, or Fresno, who had adequate age information on sex, age, race, and preschool experience.

[2]Maximum score = 18.

Reliability

Booklet 3D and 4A reliability estimates (KR-20's) for the total sample and subsamples of the fall 1969 and fall 1970 HSPV samples are listed in Tables 15-18. In general, the reliability is low (in the .60's) for both booklets. The range of the fifty-five coefficients calculated for the fall 1969 Booklet 3D sample (Table 15) was from .341 (n = 17) for northern older blacks with no previous preschool experience to .778 (n = 178) for northern children with previous preschool experience. Only 24% of all the coefficients were below .60. For the fall 1970 Booklet 3D sample (Table 16) the coefficients ranged from .361 (n = 44) for southern young blacks with previous preschool experience to .687 (n = 932) for white children. Only 7% of all the estimates were below .60. For the fall 1969 Booklet 4A sample (Table 17) the coefficients ranged from .039 (n = 16) for northern older white children with previous preschool experience to .803 (n=16) for nothern older blacks with no previous preschool. 35% of these coefficients were below .60. Coefficients for the fall 1970 Booklet 4A sample (Table 18) ranged from .391 (n=153) for southern young blacks with no previous preschool to .839 (n = 40) for northern young blacks with previous preschool. Only 15% of these KR-20's were below .60.

Item Characteristics

A factor analysis of Booklets 3D and 4A together demonstrated that the tests should remain separate for analyses. A factor analysis of Booklet 3D alone revealed there

TABLE 15

KR-20 RELIABILITIES FOR FALL 1969 HSPV BOOKLET 3D SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	1692	12.698	3.321	.694
Black	1176	12.411	3.323	.686
White	516	13.353	3.228	.699
Male	821	12.557	3.309	.685
Female	865	12.822	3.331	.702
Young[3]	808	11.346	3.055	.595
Old	884	13.933	3.065	.683
Previous Preschool	542	13.908	3.176	.709
No Previous Preschool	1150	12.128	3.236	.658
North	649	12.057	3.484	.710
South	1043	13.097	3.153	.671

[1]Sample includes all blacks and whites between 35 and 77 months at October 1, 1969, who had a fall test score and data on the previous preschool experience question.

[2]Maximum score is 19.

[3]Young is less than 60 months; old is greater than 59 months.

TABLE 16

KR-20 RELIABILITIES FOR FALL 1970 HSPV BOOKLET 3D SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	2581	11.845	3.253	.657
Black	1309	11.548	3.202	.641
White	932	12.401	3.327	.687
Male	1332	11.700	3.347	.674
Female	1249	11.999	3.143	.635
Young[3]	1143	10.907	3.202	.629
Old	1437	12.592	3.097	.637
Previous Preschool	474	12.705	3.187	.661
No Previous Preschool	2100	11.645	3.239	.650
North	1495	11.683	3.163	.634
South	1086	12.067	3.360	.684

[1]Sample includes all children who were not in a Level I site or Oraibi between 32 and 79 months at October 1, 1970.

Note: 98% of the children were between 41 and 71 months. Spanish-speaking children are included in the total sample.

[2]Maximum score is 19.

[3]Young is less than 57 months; old is greater than 56 months.

TABLE 17

KR-20 RELIABILITIES FOR FALL 1969 HSPV BOOKLET 4A SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	1676	6.039	3.279	.676
Black	1162	5.998	3.290	.677
White	514	6.130	3.255	.680
Male	807	5.886	3.231	.668
Female	863	6.172	3.322	.684
Young[3]	803	5.277	2.552	.475
Old	873	6.740	3.692	.742
Previous Preschool	538	6.846	3.587	.723
No Previous Preschool	1138	5.657	3.051	.631
North	642	5.947	3.159	.651
South	1034	6.096	3.351	.691

[1]Sample includes all blacks and whites between 35 and 77 months at October 1, 1969, who had a fall test score and data on the previous preschool experience question.

[2]Maximum score is 18.

[3]Young is less than 60 months; old is greater than 59 months.

TABLE 18

KR-20 RELIABILITIES FOR FALL 1970 HSPV BOOKLET 4A SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	2568	5.661	3.238	.686
Black	1307	5.347	3.144	.670
White	927	6.042	3.448	.725
Male	1324	5.603	3.287	.696
Female	1244	5.723	3.186	.674
Young[3]	1132	5.274	2.913	.618
Old	1435	5.965	3.445	.719
Previous Preschool	472	6.413	3.820	.772
No Previous Preschool	2089	5.483	3.062	.649
North	1484	5.830	3.209	.676
South	1084	5.430	3.265	.697

[1]Sample includes all children who were not in a Level I site or Oraibi between 32 and 79 months at October 1, 1970.

Note: 98% of the children were between 41 and 71 months. Spanish-speaking children are included in the total sample.

[2]Maximum score is 18.

[3]Young is less than 57 months; old is greater than 56 months.

is only one stable, interpretable factor.

Close investigation of the frequency distributions of scores of Booklet 3D for four subsamples of children with previous preschool experience (young whites, old whites, young blacks, old blacks) for fall 1969 and spring 1970 reveals that there is a ceiling effect in the spring. In all these groups at both times scores were negatively skewed. Ceiling effects are most prominent in the spring with the older children, especially the older white children with previous preschool experience (see Tables 19 and 20). In spring 1970, 78% of the older white children were at the three top scores (score 17--27%; score 18--22%; score 19--29%).

Investigation of the frequency distributions of scores of Booklet 4A for the same four subsamples for fall 1969 and spring 1970 reveals that there are no floor or ceiling effects. In all of the groups the scores were positively skewed in the fall and more normally distributed in the spring.

Correlations with Other Tests

Correlations of Booklets 3D and 4A with the CPSCS, the MI Subtests and the 64-item PSI are listed in Table 21 for the total fall 1970 HSPV sample and several subsamples (blacks, whites, young, old, previous preschool experience, no previous preschool experience). Correlations with the PSI are estimates of the concurrent validity of the NYU Booklets. In every case, the correlations of Booklet 3D with the 64-item PSI are higher than the correlations of Booklet 4A with the PSI. This is probably explained by the greater similarity

TABLE 19

DISTRIBUTION OF SPRING 1970 BOOKLET 3D SCORES

FOR YOUNG WHITES AND YOUNG BLACKS

WITH PREVIOUS PRESCHOOL EXPERIENCE[1]

Young Whites

Total Score	Number of Children	% of Children	Cum. % of Children	
0	0	0.0	0.0	*
1	0	0.0	0.0	*
2	0	0.0	0.0	*
3	0	0.0	0.0	*
4	0	0.0	0.0	*
5	0	0.0	0.0	*
6	1	0.00	0.00	**
7	0	0.0	0.00	*
8	2	0.01	0.01	***
9	3	0.01	0.03	****
10	7	0.03	0.06	********
11	16	0.07	0.12	******************
12	12	0.05	0.18	************
13	26	0.11	0.29	*****************************
14	27	0.12	0.41	******************************
15	39	0.17	0.57	**
16	32	0.14	0.71	**********************************
17	36	0.16	0.87	**
18	22	0.09	0.96	************************
19	9	0.04	1.00	**********
	232			

[1] Young is less than 60 months.

TABLE 19

(Con't)

Young Blacks

Total Score	Number of Children	% of children	Cum. % of children	
0	0	0.0	0.0	*
1	0	0.0	0.0	*
2	0	0.0	0.0	*
3	0	0.0	0.0	*
4	0	0.0	0.0	*
5	0	0.0	0.0	*
6	0	0.0	0.0	*
7	1	0.01	0.01	**
8	2	0.02	0.04	***
9	2	0.02	0.06	***
10	12	0.15	0.21	**********
11	4	0.05	0.26	*****
12	5	0.06	0.32	******
13	4	0.05	0.37	*****
14	13	0.16	0.54	**********
15	13	0.16	0.70	**********
16	10	0.13	0.82	**********
17	9	0.11	0.94	**********
18	4	0.05	0.99	*****
19	1	0.01	1.00	**
	80			

TABLE 20

DISTRIBUTION OF SPRING 1970 BOOKLET 3D SCORES

FOR OLD WHITES AND OLD BLACKS

WITH PREVIOUS PRESCHOOL EXPERIENCE[1]

Old Whites

Total Score	Number of Children	% of Children	Cum. % of Children	
0	0	0.0	0.0	*
1	0	0.0	0.0	*
2	0	0.0	0.0	*
3	0	0.0	0.0	*
4	0	0.0	0.0	*
5	0	0.0	0.0	*
6	0	0.0	0.0	*
7	0	0.0	0.0	*
8	0	0.0	0.0	*
9	0	0.0	0.0	*
10	0	0.0	0.0	*
11	0	0.0	0.0	*
12	0	0.0	0.0	*
13	1	0.02	0.02	**
14	0	0.0	0.02	*
15	4	0.10	0.12	*****
16	4	0.10	0.22	*****
17	11	0.27	0.49	**********
18	9	0.22	0.71	*********
19	12	0.29	1.00	**********
	41			

[1]Old is greater than 59 months.

TABLE 20

(Cont)

Old Blacks

Total Score	Number of Children	% of Children	Cum. % of Children	
0	0	0.0	0.0	*
1	0	0.0	0.0	*
2	0	0.0	0.0	*
3	0	0.0	0.0	*
4	0	0.0	0.0	*
5	0	0.0	0.0	*
6	0	0.0	0.0	*
7	2	0.02	0.02	***
8	0	0.0	0.02	*
9	2	0.02	0.04	***
10	1	0.01	0.05	**
11	4	0.04	0.10	*****
12	4	0.04	0.14	*****
13	10	0.11	0.25	**********
14	14	0.15	0.40	**********
15	8	0.09	0.49	*********
16	13	0.14	0.63	**********
17	14	0.15	0.78	**********
18	14	0.15	0.93	**********
19	6	0.07	1.00	*******
	92			

TABLE 21

CORRELATIONS OF BOOKLET 3D AND 4A SCORES WITH THE CPSCS, MI SUBTESTS, THE 64-ITEM PSI,[1] AND THE STANFORD-BINET IQ AND MA[1] FOR FALL 1970 HSPV DATA

	NYU 3D	PSI 64-item	CPSCS	MI[2] walk	MI draw	MI truck	IQ[4]	MA
Total Sample								
NYU 4A	.429 (2125)	.467 (2117)	.240 (2045)	.142 (1072)	.142 (1077)	.106 (1065)	.365 (749)	.435 (750)
NYU 3D		.696 (2127)	.297 (2057)	.275 (1073)	.298 (1078)	.136 (1065)	.427 (753)	.640 (754)
Blacks								
NYU 4A	.480 (759)	.513 (752)	.269 (723)	.180 (440)	.148 (440)	.030 (440)	.442 (294)	.534 (294)
NYU 3D		.710 (756)	.303 (728)	.303 (440)	.297 (440)	.120 (440)	.434 (296)	.696 (296)
Whites								
NYU 4A	.467 (1082)	.492 (1076)	.214 (1043)	.150 (545)	.126 (546)	.121 (542)	.394 (370)	.446 (371)
NYU 3D		.699 (1054)	.270 (1052)	.278 (547)	.326 (548)	.153 (543)	.436 (374)	.658 (375)
Young[3]								
NYU 4A	.407 (1012)	.451 (1007)	.222 (979)	.075 (408)	.046 (412)	.160 (412)	.256 (313)	.317 (313)
NYU 3D		.652 (1017)	.314 (990)	.253 (408)	.200 (412)	.205 (412)	.453 (316)	.528 (316)

[1]Sample size for each correlation is in parentheses. Children included were not in Level I sites, Oraibi, or Fresno; between 43 and 74 months; and in school for the full year.

[2]MI scores are log transformations of the "slow" times; MI scores were used only if the child had passed two out of four practice items.

[3]Under 58 months.

[4]Pinneau IQ calculations used.

TABLE 21

(Cont)

	NYU 3D	PSI 64	CPSCS	MI walk	MI draw	MI truck	IQ	MA
Old[5]								
NYU 4A	.426 (1113)	.469 (1110)	.235 (1066)	.160 (664)	.167 (665)	.085 (653)	.470 (436)	.476 (437)
NYU 3D		.676 (1110)	.245 (1067)	.259 (665)	.303 (666)	.113 (653)	.541 (437)	.641 (438)
Previous Preschool								
NYU 4A	.443 (409)	.485 (408)	.191 (398)	.147 (203)	.094 (202)	.082 (198)	.441 (147)	.451 (147)
NYU 3D		.717 (410)	.295 (400)	.378 (203)	.330 (202)	.155 (198)	.389 (147)	.638 (147)
No Previous Preschool								
NYU 4A	.417 (1716)	.452 (1709)	.244 (1647)	.140 (869)	.142 (875)	.124 (867)	.338 (602)	.413 (607)
NYU 3D		.684 (1717)	.285 (1657)	.249 (870)	.280 (876)	.141 (867)	.422 (606)	.624 (607)

[5]Over 58 months.

of items between the PSI and Booklet 3D. There are very few items of recognizing letters, numbers and shapes (contents of Booklet 4A) in the PSI. If the correlations are corrected for unreliability[1], the estimated correlation between the true score components of the 64-item PSI and Booklet 3D is .90 ($.70 \div \sqrt{(.66)(.92)}$). In general, correlations between the NYU Booklets and other tests are higher for whites, older children, and children with previous preschool experience.

Remarks

Neither Booklet 3D or 4A is an adequate achievement estimate alone since they both have low internal reliability and the 3D has definite floor and ceiling effects. Interpretations of summary scores are sometimes difficult to make. This is less true of Booklet 3D since its true correlation with the 64-item PSI is very high, indicating that they are measuring the same cognitive domain.

These booklets are best used as a set of criterion-referenced measures. Using this concept, the percentages of children in various sites and models who obtain either a perfect score or only one item incorrect on each subtest and who fail to get more than one item correct on each subtest are reported.

---

[1] Using formula $\frac{r_{1 \cdot 2}}{\sqrt{t_1 \cdot t_2}}$, where $r_{1 \cdot 2}$ is the correlation between tests and t_1, t_2 are estimates of test reliability.

Reference

Coller, A. and J. Victor. Early Childhood Inventories Project. New York City: Institute for Developmental Studies, New York University School of Education, 1971.

Smith, M.S. Some short term effects of Project Head Start: A preliminary report on the second year of planned variation: 1970-71. Prepared under Grant # H 1926 for the Office of Child Development. Cambridge, Mass.: Huron Institute, 1973.

Parent Information Form

Purpose

The Parent Information Forms were designed to obtain a variety of background information about the parents and children in the Head Start Planned Variation sample. The data can be used as independent and dependent variables in the analysis. This information enables investigators to assess what a child brings with him from his background to the Head Start experience and what changes, if any, the Head Start experience have on a child and his family.

Description

A Parent Information Form (PIF) was administered to mothers in the HSPV sample who were given the Eight-Block Sort Task. The content of the Parent Information Form changed from year to year, as questions were added, deleted and modified. In general, the PIF included questions about the following areas:

1. Demographic. These questions served primarily as a check on the data gathered on the Classroom Information Form. This information was included only in Fall 1971.

2. Non-demographic family background. The parents were asked for such information as how often they read to their children, how often they go on trips, and what the child does at home. The information tapped by these questions can be interpreted as measures of the background

which the child brings to Head Start, and thus as independent variables in the analysis. Changes can also be interpreted as possible effects of Head Start on parents, and thus as dependent variables.

3. Parent and child attitudes. On some forms of the PIF parents were asked a series of attitude questions designed to measure the parents' sense of control over the environment. On all forms parents were asked about their own and their child's feelings toward the Head Start program. Like the non-demographic family background measures, these can be interpreted both as dependent and independent variables.

4. Parent participation. The parents were asked about their own involvement in the Head Start programs, Community Action Program (CAP), membership, volunteer work, etc. Since parent involvement is an important goal of the Head Start program, these questions measure an important desired outcome. Parent involvement may also be a contributing factor to child success.

Reliability

A test-retest reliability study was done on the rather short form of the PIF which was given in the fall of 1971. The study is reported in Appendix F. This form did not include any questions on parental sense of control. Nor did it include, since it was given in the fall, questions about participation or feelings about Head Start. Thus nothing can be said about the reliability of these items.

The findings using a very small sample indicate that the PIF is a reliable instrument for gathering demographic data of the sort on the short form of Fall 1971. Questions of a simple yes/no format are especially reliable. Because of the low response rates for many questions and the findings of moderate consistency on the educational aspirations and expectancies questions, it is doubtful that adequate attitude data can be collected on such a form.

Remarks

The data on the non-attitudinal, simple format questions used in the PIF appears to have adequate reliability, based on a very small sample. Reliability estimates for these items using a larger sample and for attitude and participation items using any sample need to be determined before such a form is used in other studies.

Peabody Picture Vocabulary Test

Purpose

The Peabody Picture Vocabulary Test (PPVT) is designed to measure a child's verbal intelligence by measuring his receptive vocabulary--the number of words which he knows when he hears them. Vocabulary is a major component of general intelligence measures. Vocabulary subparts of both the Stanford-Binet and the WPPSI correlate in the low .70's with their respective total "intelligence" scores (McNemar, 1942; Wechsler, 1967).

Description

The PPVT is an untimed individual test consisting of a booklet with three practice items and 150 test plates each with four numbered pictures. The version used in the HSPV Study is Form A, modified by SRI and ETS to include pictures of blacks. For each item the stimulus word (a noun or verb form) is presented orally and the child is required to indicate the picture corresponding to the word, either by pointing or by giving the number of the appropriate picture. Items increase in difficulty and are presented to a child until six errors are made out of eight consecutive

responses or the test is completed. A complete list of the words appears in Table 1.

The maximum number of words given was 100 in Fall 1971 and 150 in Spring 1972. In the fall all children began the test at item 1 and continued until the ceiling was reached or 100 words had been given. In the spring every child began at item 25 and a basal level was established:

1. If the child got items 25 through 32 correct.
2. If the child missed any item from 25 to 32, the tester gave the items backwards from item 24 until
   a. the child got eight correct in a row, or
   b. the child went through items 24 to 1 without getting eight correct in a row.

In the first case, the test was continued from item 32 until the ceiling or end of the test was reached. In the second case, the test was continued from the first item missed by the child until the ceiling or the end of the test was reached or the test was discontinued if the child had already missed six out of eight items.

Each item was scored as correct, incorrect, child refused or indeterminate. This is a "tailored test," meaning that there is not a fixed number of items given to each child. The test is also Guttman-scaled; in other words, it is assumed a child will get all items correct below any specific item on the test.

TABLE 1

WORDS INCLUDED IN THE PEABODY PICTURE VOCABULARY TEST

1. car
2. cow
3. baby
4. girl
5. ball
6. block
7. clown
8. key
9. can
10. chicken
11. blowing
12. fan
13. digging
14. skirt
15. catching
16. drum
17. leaf
18. tying
19. fence
20. bat
21. bee
22. bush
23. pouring
24. sewing
25. wiener
26. teacher
27. building
28. arrow
29. kangaroo
30. accident
31. nest
32. caboose
33. envelope
34. picking
35. badge
36. goggles
37. peacock
38. queen
39. coach
40. whip
41. net
42. freckle
43. eagle
44. twist
45. shining
46. dial
47. yawning
48. bumble
49. signal
50. capsule
51. submarine
52. thermos
53. projector
54. group
55. cackling
56. transportation
57. counter
58. ceremony
59. pod
60. bronco
61. directing
62. funnel
63. delight
64. lecturer
65. communication
66. archer
67. stadium
68. excavate
69. assaulting
70. stunt
71. meringue
72. appliance
73. chemist
74. arctic
75. destruction

TABLE 1 (CON'T)

76.	porter	101.	graduated	126.	dormer
77.	coast	102.	hieroglyphic	127.	coniferous
78.	hoisting	103.	orate	128.	consternation
79.	wailing	104.	cascade	129.	obese
80.	coil	105.	illumination	130.	gauntlet
81.	kayak	106.	nape	131.	inclement
82.	sentry	107.	genealogist	132.	cupola
83.	furrow	108.	embossed	133.	obliterate
84.	beam	109.	mercantile	134.	burnishing
85.	fragment	110.	encumbered	135.	bovine
86.	hovering	111.	entice	136.	eminence
87.	bereavement	112.	concentric	137.	legume
88.	crag	113.	vitreous	138.	senile
89.	tantrum	114.	sibling	139.	deleterious
90.	submerge	115.	machete	140.	raze
91.	descend	116.	waif	141.	ambulation
92.	hassock	117.	cornice	142.	cravat
93.	canine	118.	timorous	143.	impale
94.	probing	119.	fettered	144.	marsupial
95.	angling	120.	tartan	145.	predatory
96.	appraising	121.	sulky	146.	incertitude
97.	confining	122.	obelisk	147.	imbibe
98.	precipitation	123.	eclipse	148.	homunculus
99.	gable	124.	entomology	149.	cryptogam
100.	amphibian	125.	bumptious	150.	pensile

Development of Instrument

In 1959 Lloyd M. Dunn developed the PPVT in two parallel Forms A and B as a measure of receptive vocabulary for ages two-and-one-half to eighteen. In selecting the final stimulus words for the test Dunn had several groups of subjects of all ages selectively sort from an original pile of 2,055 line drawings of illustrable nouns and verbs (Buros, 1965). Since its creation the PPVT has been widely used in studies with children, especially mentally retarded and handicapped children (see references in Buros, 1965; 1972). Several investigators have used the original PPVT version with disadvantaged preschoolers (Costello & Ali, 1971; Datta, 1967; DiLorenzo & Brady, 1968; Milgram & Ozer, 1967; Rieber & Womack, 1968; Shipman et al., 1971). The first 75 items of the standard Dunn test were used in the first year of the ETS Longitudinal Study (Shipman, 1972). For the second year of the study, a modified version of 60 items which contained redrawings of a number of human pictures to include blacks and adults in a variety of roles, was used. This modified 60 item version was extended to all 150 pictures by Shipman and Tanaka in 1971 for use in the 1971-72 HSPV study and Follow Through evaluation.

Standardization

The original PPVT standardization sample was based entirely on 4,012 white children in and around Nashville, Tennessee. Children ranging in number from 92 to 354 and representing 19 different age levels from 2.5 to 18 years were tested on both forms. Only children under nine were given the test individually. The standard scores (PPVT IQs) were derived by assigning an IQ of 100 (S.D. = 15) to the mean raw score for each distribution of subjects arranged in six month age intervals. There is a problem in using these norms with younger children since two children with the same raw score one month apart will get widely discrepant IQ scores. For example, a 44-month-old child with a raw score of 28 would be assigned an IQ of 89 while a child one month older with the same raw score would be assigned an IQ of 76 (a 13 point difference). Because of the inconsistencies in the norm tables, DiLorenzo and Brady (1968, p. 247) concluded that "the use of PPVT IQ data in the evaluation of preschool programs could produce invalid results and thus lead to spurious conclusions regarding program effectiveness." In several previous studies with disadvantaged preschool populations (Datta, 1967; DiLorenzo & Brady, 1968; Milgram & Ozer, 1967; Rieber & Womack, 1968) PPVT scores have been substantially lower than the normalization sample.

Norms (mean, S.D., percentiles) for the raw scores of the ETS Head Start Longitudinal sample are available for children in Year 1 (in three month age intervals from 42-59 months) and for children in Year 2 (in three month age intervals from 51-69 months) (see Shipman, 1972). The mean for the total sample in Year 1 was 26.3 (S.D. = 12.85, N = 1198); the mean for the total sample in Year 2 was 41.6 (S.D. = 9.75, N = 1309).

Norms for raw scores of the HSPV version of the PPVT are available in Tables 1 - 8. Based on 15 three month age intervals from 36-38 months to 78-80 months, these tables give the number of children, the mean score and the standard deviation at each age level for the following groupings of the HSPV sample: total sample (Table 1), males (Table 2), females (Table 3), children with previous preschool experience (Table 4), children with no previous preschool experience (Table 5), black children (Table 6), white children (Table 7) and Mexican-American children (Table 8). The mean score for the HSPV sample was 31.525 (S.D. = 13.258, N = 2996). A developmental age trend can be evidenced in all of the norm tables.

TABLE 1

DISTRIBUTION OF PPVT SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	5	20.400	13.185
39-41	4	23.500	3.354
42-44	22	19.864	9.251
45-47	76	21.697	10.885
48-50	263	24.711	10.930
51-53	474	26.633	11.841
54-56	476	29.626	11.938
57-59	468	31.630	13.021
60-62	381	34.554	13.360
63-65	259	35.216	12.611
66-68	261	36.659	12.346
69-71	211	39.336	12.205
72-74	89	42.000	12.437
75-77	4	36.750	23.424
78-80	3	33.000	9.416
TOTAL	2996	31.525	13.258

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 150.

TABLE 2

DISTRIBUTION OF PPVT SCORES FOR MALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	1	22.000	-----
42-44	8	18.250	5.449
45-47	47	21.447	10.536
48-50	130	25.815	11.291
51-53	252	27.433	12.429
54-56	235	30.055	12.310
57-59	239	31.594	13.455
60-62	201	35.552	13.421
63-65	120	35.725	12.930
66-68	138	37.391	12.790
69-71	98	41.061	12.034
72-74	47	42.936	12.542
75-77	2	60.000	2.000
78-80	2	35.000	11.000
TOTAL	1520	32.161	13.620

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 150.

TABLE 3

DISTRIBUTION OF PPVT SCORES FOR FEMALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	5	20.400	13.185
39-41	3	24.000	3.742
42-44	14	20.786	10.732
45-47	29	22.103	11.415
48-50	133	23.632	10.453
51-53	222	25.725	11.066
54-56	241	29.207	11.549
57-59	229	31.668	12.551
60-62	180	33.439	13.202
63-65	139	34.777	12.313
66-68	123	35.837	11.775
69-71	113	37.841	12.155
72-74	42	40.952	12.234
75-77	2	13.500	3.500
78-80	1	29.000	------
TOTAL	1476	30.870	12.843

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 150.

TABLE 4

DISTRIBUTION OF PPVT SCORES FOR ALL CHILDREN WITH PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	------	------
39-41	---	------	------
42-44	2	32.000	7.000
45-47	8	35.000	13.257
48-50	35	25.857	11.736
51-53	66	30.318	11.455
54-56	58	29.345	11.897
57-59	76	33.882	12.190
60-62	118	36.610	11.854
63-65	98	34.939	11.758
66-68	94	37.160	10.522
69-71	96	39.135	11.656
72-74	38	40.474	10.351
75-77	2	60.000	2.000
78-80	2	37.500	8.500
TOTAL	693	34.999	12.216

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 150.

TABLE 5

DISTRIBUTION OF PPVT SCORES FOR ALL CHILDREN WITH NO PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	5	20.400	13.185
39-41	4	23.500	3.354
42-44	19	18.316	8.639
45-47	67	20.060	9.457
48-50	222	24.595	10.868
51-53	391	26.097	11.900
54-56	404	29.921	11.931
57-59	379	31.158	13.097
60-62	253	33.648	13.972
63-65	153	36.275	12.683
66-68	161	36.901	12.729
69-71	112	40.036	12.154
72-74	51	43.137	13.677
75-77	2	13.500	3.500
78-80	1	24.000	------
TOTAL	2224	30.660	13.381

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 150.

TABLE 6

DISTRIBUTION OF PPVT SCORES FOR BLACK CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	5	20.400	13.185
39-41	1	23.000	------
42-44	14	20.571	8.650
45-47	43	19.953	8.221
48-50	142	21.542	9.462
51-53	213	23.005	9.432
54-56	207	25.324	10.061
57-59	194	27.918	11.129
60-62	143	29.063	10.945
63-65	111	31.369	10.265
66-68	107	33.523	10.066
69-71	107	35.916	10.652
72-74	35	34.914	10.283
75-77	2	13.500	3.500
78-80	1	24.000	------
TOTAL	1325	27.343	11.211

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 150.

TABLE 7

DISTRIBUTION OF PPVT SCORES FOR WHITE CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	------	------
39-41	3	23.667	3.859
42-44	8	18.625	10.099
45-47	33	23.970	13.254
48-50	93	29.903	10.751
51-53	183	32.290	12.114
54-56	205	34.415	11.984
57-59	188	36.920	12.656
60-62	160	40.506	12.462
63-65	95	40.516	10.719
66-68	100	41.500	11.236
69-71	81	45.963	9.735
72-74	52	47.231	10.321
75-77	2	60.000	2.000
78-80	2	37.500	8.500
TOTAL	1205	36.972	12.970

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 150.

TABLE 8

DISTRIBUTION OF PPVT SCORES FOR MEXICAN-AMERICAN CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	------	------
39-41	---	------	------
42-44	---	------	------
45-47	---	------	------
48-50	22	27.500	10.352
51-53	68	24.044	11.314
54-56	53	27.868	11.646
57-59	71	29.296	13.580
60-62	69	32.246	14.057
63-65	51	33.098	16.236
66-68	51	33.431	15.281
69-71	19	29.211	15.182
72-74	---	------	------
75-77	---	------	------
78-80	---	------	------
TOTAL	404	29.629	13.978

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 150.

Reliability

In general the reliability estimates for the PPVT are quite good. Dunn (1965) reports that the parallel forms reliability estimate with children between three and six is about .72. The internal reliability estimates (Alpha coefficients) computed for the ETS Longitudinal study were .96 in Year 1 for the 70 item original version and .91 in Year 2 for the 60 item revised version (Shipman, 1972). Since the Year 1 version was cut-off at 70 items and the Year 2 version had only 60 items, these estimates are probably inflated.

Costello and Ali (1971) report a test-retest (two week interval) coefficient of .77 between PPVT raw scores for 36 black preschoolers (ages 4-1 to 5-0). Milgram and Ozer (1967) found that the test-retest coefficient after four weeks was .69 for the PPVT MA's of 65 Head Start children (ages 4 1/2 to 6). The test-retest coefficient after 10 months (from ages 3-1 to 3-11) was .80 for PPVT MA's of 51 disadvantaged preschoolers. The .69 coefficient of stability (correlation of form A administered in Year 1 and Year 2) for the PPVT was the highest of any in the ETS Longitudinal Study test battery (Shipman, 1972).

No internal consistency estimates were calculated for the HSPV sample since, according to Stanley (1971), these estimates are inappropriate for a "tailored test" like the PPVT. A "tailored test" is one where there is not a fixed number of items given to each child; instead, the test is "tailored" to the child's level and needs.

Validity

Congruent or concurrent validity estimates include comparisons with intelligence, language and achievement tests. Predictive validity estimates include comparisons with future school success. In general, there are many more concurrent validity estimates available for the PPVT: most of these are correlations with intelligence measures.

Comparisons with the Stanford-Binet. Dunn (1965) reports that PPVT scores correlate .83 with SB IQ scores and .64 with WPSSI scores. Several studies with disadvantaged preschoolers have used both the Stanford-Binet and the PPVT. Even though the two tests are highly correlated, the PPVT IQ scores found in these studies have been consistently lower than the SB IQ's. Using a predominantly (85%) disadvantaged preschool sample (n = 563) in seven New York communities, DiLorenzo and Brady (1968) found that PPVT IQ's were consistently about nine points lower than the SB mean IQ's. These differences, ranging in magnitude from 6.33 to 12.32, existed for every 10-point interval on the SB IQ scale from 50 to 130. The difference for the entire sample was 8.83 (SB mean IQ = 93.68; PPVT IQ = 84.85). The correlation between the PPVT and SB was .79. DiLorenzo and Brady attributed the discrepancies between the two tests to the inadequate standardization norms.

Milgram and Ozer (1967) found that the PPVT MA scores of two disadvantaged populations were consistently

lower than the SB MA's. For example, the SB MA of 65 Head Start children (CA = 5-6) enrolled in a summer program was 4-8 in June, compared with PPVT MA scores of 3-6 in June and 4-0 in August. The authors felt that the PPVT scores were susceptible to a decelerating trend while SB scores were not. The correlation between the SB IQ and the PPVT IQ at age 5 was .65 for 51 preschoolers. Milgram and Ozer concluded that the PPVT is more susceptible to environmental impoverishment than the SB.

In a study of 36 black preschoolers (ages 4-1 to 5-0) Costello and Ali (1971) found that the PPVT raw score correlated .43 with the SB MA and .28 with the SB IQ. They hypothesized that the lower PPVT scores were attributable to either environmental variables or to examiner and situation variables.

Finally, Bruinicks and Lucker (1970) found that the SB IQ calculated at the beginning of the first grade was a better predictor of the reading subtest of the Metropolitan Achievement Test (correlation at end of first grade = .32; correlation at the end of fourth grade = .60) than was the PPVT IQ (correlation at the end of first grade = .18; correlation at the end of fourth grade = .45). The correlation between the two test IQ's for 36 lower class first grade children was .71.

Correlations with the PSI. Datta found that correlations of the PPVT with the original PSI were high for a sample of 956 Head Start children in 72 centers. Correlations of the PPVT faw score with PSI were .73 for the total, .69 for age 4, .62 for age 5, and .80 for age 6, (Datta, 1967).

In the ETS Longitudinal Study, Shipman (1972) found that the highest correlation of the PPVT with another test in the battery was .58 with the 64-item PSI in Year 1 and .66 with the 64-item PSI in Year 2.

In the third year of the HSPV Study, the highest correlation between the PPVT and another test in the battery was .665 for the 32-item PSI (See Table 9).

Correlations with the ITPA. Using a sample of lower class Australian children, Teasdale (1969) found that the PPVT raw scores correlated .45 with the Verbal Expression Subtest and .74 with the total ITPA score. Costello and Ali (1971) found a correlation of .28 with the Verbal Expression Subtest.

In the third year of the HSPV study, the correlation between the PPVT and the ITPA-Verbal Expression Subtest was .487 (See Table 9).

Correlations with the MI. In the ETS Longitudinal Study, the PPVT and the average

TABLE 9

INTERCORRELATIONS OF FALL 1971 SCORES FROM THE PPVT, WRAT SUBTESTS, 32-ITEM PSI, ITPA VERBAL EXPRESSION SUBTEST, ETS ENUMERATION SUBTESTS, BROWN, MI-TRUCK SUBTEST, AND EIGHT-BLOCK SORT SUCCESS SCORES[1]

	PPVT	WRAT-COPY MARKS	WRAT-RECOG. LETTERS	WRAT-NAME LETTERS	WRAT-READ #'s	WRAT-DOT COUNT.	PSI 32-ITEM	ITPA-VERBAL EXPRESS.	ETS. ENUM. TOTAL	ETS. ENUM. COUNT.	ETS ENUM. TOUCH.	ETS ENUM. SAME # MATCH.	BROWN UNADJ.	BROWN ADJ.	MI - TRUCK	EIGHT-BLOCK PLACE.	EIGHT-BLOCK REASON
WRAT-COPY MARKS	.413 (2881)																
WRAT-RECOG,LETTERS	.537 (2881)	.375 (2995)															
WRAT-NAME LETTERS	.346 (2881)	.358 (2995)	.302 (2995)														
WRAT-READ NUMBERS	.407 (2881)	.412 (2995)	.325 (2995)	.600 (2995)													
WRAT-DOT COUNTING	.453 (2881)	.466 (2995)	.419 (2995)	.344 (2995)	.451 (2995)												
PSI (32-item)	.665 (2855)	.551 (2860)	.481 (2860)	.414 (2860)	.508 (2860)	.589 (2860)											
ITPA-VERBAL EXPRESSION	.487 (1147)	.339 (1172)	.371 (1172)	.276 (1172)	.341 (1172)	.368 (1172)	.506 (1138)										
ETS ENUMERATION TOTAL	.475 (1075)	.508 (1097)	.427 (1097)	.307 (1097)	.446 (1097)	.542 (1097)	.584 (1073)	.459 (1115)									
ETS ENUMERATION COUNTING	.492 (1075)	.504 (1097)	.422 (1097)	.359 (1097)	.500 (1097)	.620 (1097)	.625 (1073)	.384 (1115)	.781 (1135)								
ETS ENUMERATION TOUCHING	.282 (1075)	.358 (1097)	.293 (1097)	.196 (1097)	.271 (1097)	.383 (1097)	.382 (1073)	.308 (1115)	.721 (1135)	.390 (1135)							
ETS ENUMERATION SAME # MATCHING	.237 (1075)	.225 (1097)	.199 (1097)	.095 (1097)	.176 (1097)	.148 (1097)	.232 (1073)	.298 (1115)	.664 (1135)	.257 (1135)	.202 (135)						
BROWN-UNADJUSTED	.322 (2689)	.162 (2753)	.243 (2753)	.145 (2753)	.173 (2753)	.270 (2753)	.323 (2689)	.261 (1145)	.228 (1073)	.271 (1073)	.160 (1073)	.054 (1073)					
BROWN-ADJUSTED	.239 (2689)	.127 (2753)	.166 (2753)	.100 (2753)	.124 (2753)	.194 (2753)	.259 (2689)	.215 (1145)	.159 (1073)	.172 (1073)	.134 (1073)	.034 (1073)	.637 (2879)				
MI-TRUCK[3]	.174 (607)	.061 (625)	.048 (625)	.083 (625)	.121 (625)	.006 (625)	.164 (608)	.052 (637)	.136 (597)	.135 (597)	.047 (597)	.107 (597)	.118 (610)	.109 (610)			
EIGHT-BLOCK PLACEMENT	.304 (1119)	.222 (1148)	.271 (1148)	.145 (1148)	.207 (1148)	.304 (1148)	.305 (1090)	.305 (1096)	.322 (1032)	.413 (1032)	.200 (1032)	.180 (1032)	.212 (1113)	.183 (1113)	.005 (573)		
EIGHT-BLOCK REASON	.445 (1119)	.364 (1148)	.333 (1148)	.286 (1148)	.372 (1148)	.350 (1148)	.443 (1090)	.418 (1096)	.405 (1032)	.402 (1032)	.258 (1032)	.211 (1032)	.178 (1113)	.168 (1113)	.063 (573)	.520 (1211)	
EIGHT-BLOCK SUCCESS TOTAL	.439 (1119)	.346 (1148)	.351 (1148)	.257 (1148)	.344 (1148)	.404 (1148)	.440 (1090)	.422 (1096)	.422 (1032)	.416 (1032)	.266 (1032)	.226 (1032)	.220 (1113)	.200 (1113)	.046 (573)	.839 (1211)	.901 (1211)

[1] Sample size for each correlation is included in parenthesis. Children in sample are those with adequate information not in Level I sites.

[2] ETS ENUMERATION Score= sum of counting, touching and same number matching subtest scores.

[3] MI scores are log transformations of slow times.

slow time of the Drawing and Walking subtests of the MI correlated .36 in Year 1 and .34 in Year 2 (Shipman, 1972).

In the HSPV Study, the PPVT correlated .174 with the slow time of the MI Truck subtest (See Table 9).

Correlations with the Eight Block Sort Task. In the ETS Longitudinal Study, the PPVT and total success score from the Eight-Block Sort Task correlated .39 in Year 1 and .53 in Year 2 (Shipman, 1972).

In the HSPV study the PPVT correlated .439 with the total success score, .445 with the reason success score, and .304 with the placement success score (see Table 9).

Other Correlations. Other PPVT correlations of interest from the ETS Longitudinal Study were those with the TAMA General Knowledge Test (.52 in Year 1; .63 in Year 2), with the Children's Auditory Discrimination Inventory: Nonsense Words (.52 in Year 1; .47 in Year 2), with the Matching Familiar Figures: mean errors per valid item (-.45 in Year 1; -.50 in Year 2) and with the Seguin Form Board: Log fastest time for correct placement (-.40 in Year 1; -.46 in Year 2).

Correlations with the PPVT and tests in the Fall 1971 HSPV battery can be found in Table 9. Correlations over .40 that have not already been cited are .413 (WRAT-Copying Marks), .537 (WRAT-Recognizing Letters), .407 (WRAT-Reading Numbers), .475 (ETS Enumeration Total) and .492 (ETS Enumeration:

Counting Subtest).

Remarks

The PPVT may be susceptible to practice effects and to unintended gestural or verbal hinting by the examiner. This could introduce systematic biases which have not fully been examined in previous analyses.

There are some problems about how to record changes in children's answers. The test manual says that when a child changes his choice, his last response should be recorded. It is possible that sometimes this change would be missed if a child pointed to a different picture while the tester was recording his first response. This problem may produce systematic effects on both the reliability and validity of test scores, especially those with young children.

Some items have a low probability of occurring in the natural environments of the children being tested. For instance, "weiner" is a label few children know. "Capsule" is most probably known in relationship to space rather than as a synonym for a pill. It is hard to know if such items lower everyone's score equally, or introduce systematic biases.

Another area which needs to be further explored is the effect of switching between nouns and verb forms throughout the test. John and Goldstein (1964) found that black preschoolers had more trouble with verb forms than noun forms of the original PPVT. Jeruchimowicz, Costello and Bagur (1971) found that

lower SES black preschoolers had a significantly higher proportion of errors on the action words (verbs) than the object words (nouns) of the PPVT, while middle SES black preschoolers showed no difference between verb and noun errors. It is also unclear what effects result from omitting articles before nouns (i.e., "Point to cat" rather than "point to the cat" or "point to a cat"). Articles were apparently eliminated to preserve symmetry of presentation for nouns and verbs.

There may be a confounding of a child's increased attention span and increased vocabulary knowledge. This is further complicated in the fall data since all children start at the beginning rather than at an appropriate floor for him (as is done in the spring). Greater variation in the number of items presented to each child is paralleled by greater differences in the demand on the child's attention span. Thus it would be hard to know if an improved score between two times reflected either an increase in receptive vocabulary or an increase in attention, neither, or both.

Some of the above mentioned problems have been eliminated in the Modified Peabody Picture Vocabulary Test used by Ali and Costello (1971). The modified version consists of 70 items randomized for difficulty levels specified stimulus instructions and controlled schedules of reinforcement. Both the test-retest coefficient and overall scores of black preschool

children were higher for the modified version than the standard version of the PPVT. Further use and development of this modified version as well as the ETS short (60 item) modified version, is encouraged in future studies with preschoolers.

Even though the PPVT has correlated fairly high with other intelligence and language measures and has loaded highest on the "g" factor (general information -- processing ability) in factor analyses of the ETS Longitudinal data (Shipman, 1971), it is recommended that it be used only as a measure of receptive vocabulary at this time. As Costello and Ali state: "While Form A of the Peabody could be used as first approximation in a continuing assessment program, scores cannot be considered alone for either intellectual or language evaluation (1971, p. 755)."

References

Ali, F., & Costello, J. Modification of the Peabody Picture Vocabulary Test. Developmental Psychology, 1971, 5, 86-91.

Bruininks, R. H., & Lucker, W. G. Change and stability in correlations between intelligence and reading test scores among disadvantaged children. Journal of Reading Behavior, 1970, 2, 295-305.

Buros, O. K. (Ed.) The sixth mental measurements yearbook. Highland Park, N.J.: Gryphon, 1965.

Buros, O. K. (Ed.) The seventh mental measurements yearbook. Highland Park, N. J.: Gryphon, 1972. 2 vols.

Costello, J., & Ali, F. Reliability and validity of Peabody Picture Vocabulary Test scores of preschool disadvantaged children. Psychological Reports, 1971, 28, 755-60.

Datta, L. Draw-A-Person as a measure of intelligence in preschool children from very low income families. Journal of Consulting Psychology, 1967, 31, 626-30.

DiLorenzo, L. T., & Brady, J. J. Use of the Peabody Picture Vocabulary Test with preschool children. Psychological Reports, 1968, 22, 247-51. (Also in Training School Bulletin, 1969, 65, 117-21.)

DiLorenzo, L. T., & Salter, R. An evaluative study of prekindergarten programs for educationally disadvantaged children: follow-up and replication. Exceptional Children, 1968, 35, 111-20.

Dunn, L. M. Expanded manual for the Peabody Picture Vocabulary Test. Circle Pines, Minnesota: American Guidance Service, 1965.

John, V. P., & Goldstein, L. S. The social context of language acquisition. Merrill-Palmer Quarterly, 1964, 100, 265-75.

Jeruchimowicz, R., Costello, J., & Bagur, J. Knowledge of action and object words: A comparison of lower and middle-class negro preschoolers. Child Development, 1971, 42, 455-64.

Milgram, N. A., & Ozer, M. N. Peabody Picture Vocabulary Test scores of preschool children. Psychological Reports, 1967, 20, 779-84.

McNemar, Q. The revision of the Stanford-Binet scale. Boston: Houghton-Mifflin, 1942.

Rieber, M., & Womack, M. The intelligence of preschool children as related to ethnic and demographic variables. Exceptional Children, 1968, 34, 609-16.

Shipman, V. C., Barone, J., Beaton, A., Emmerich, W., & Ward, W. Disadvantaged children and their first school experiences: Structure and development of cognitive competencies and styles prior to school entry. Princeton, N. G.: Educational Testing Service, 1971.

Shipman, V. C. (Ed.) Disadvantaged children and their first school experiences: Technical report series. Prepared for the Office of Child Development, Department of Health, Education & Welfare Grant #H-8526. Princeton, N.J.: Educational Testing Service, 1972.

Stanley, J. C. Reliability. In Thorndike, R. L. (Ed.), Educational Measurement. 2nd Ed. Washington, D.C.: American Council on Education, 1971.

Teasdale, G. R. Validity of the PPVT as a test of language ability with lower SES children. Psychological Reports, 1969, 25, 746.

Wechsler, D. Wechsler Preschool and Primary Scale of Intelligence. New York: Psychological Corporation, 1967.

Preschool Inventory

Purpose

The Preschool Inventory (PSI)

> "...was developed to give a measure of achievement in areas regarded as necessary for success in school. The Inventory is by no means culture free; in fact, one aim in its development was to provide educators with an instrument that would permit them to highlight the degree of disadvantage which a child from a deprived background has at the time of entering school so that any observed deficits might be reduced or eliminated. Another goal was to develop an instrument that was sensitive to experience and could thus be used to demonstrate changes associated with educational intervention." (Cooperative Tests and Services, 1970, p. 4.)

Description

Two versions of the PSI have been used in the HSPV Study. A 64-item version (Revised Edition - 1970) was used in the first two years of the study. The 64 items include 21 on general knowledge, two on listening and work meanings, ten on listening and comprehension, four on writing and form copying, 24 on quantitative concepts, and three on speaking and labeling. In the manual published by Cooperative Tests and Services (1970) the items are divided into four main areas: Personal-Social Responsiveness (18 items), Vocabulary (12 items), Concept Activation-Numerical (15 items), and Concept Activation-Sensory (19 items). Most (60%) of the items require a verbal response from the child while the rest require him to follow directions such as "Point to the middle checker" or "Color the triangle orange."

Only 32 items, all of which were in the Revised Edition-1970, were used in the third year of the HSPV Study with the Hertzig-Birch coding. Representative item examples include:

1. Pointing to and naming body parts.
2. Prepositional concepts such as "on", "behind", and "under".
3. General knowledge such as "What is your first name?" and "What does a dentist do?"
4. Numerical concepts such as "middle", "last", and "more".
5. Copying forms.
6. Recognizing colors.
7. Motoric reproductions such as "how a ferris wheel goes".

The total PSI score for either version was defined as the total number of correct items. Separate analyses were done in the third year on the Hertzig-Birch codes. (See Hertzig-Birch section of this report for a complete explanation.)

Development of Instrument

The Preschool Inventory was originally constructed by B. Caldwell in the summer of 1965 to provide Project Head Start with a practical measure of preschool achievement. The test was intended to measure educational achievement for three to six year olds on skills traditionally expected of middle-class kindergarten children. It was designed as a practical measure, more a criterion-reference classroom test to be used by teachers for diagnostic purposes than a test of psychometric intelligence of generalized cognitive ability. In addition, since the instrument was intended for

use in the field, there was a concern that it require minimal training or special expertise to administer, and simple equipment (Cooperative Tests and Services, 1970).

Caldwell was responsible for an initial list of 161 questions, from which 85 were selected on the basis of a preliminary study and clinical item-analysis. The original list of questions was designed to measure performance in seven basic areas:

1. Basic information and vocabulary.
2. Number concepts and ordination.
3. Concepts of size, shape, motion and color.
4. Concepts of time, object class and social functions.
5. Visual-motor performance.
6. Following instructions.
7. Independence and self-help.

In the statistical analysis which led to the formulation of the original 85-item test, a principal components factor analysis revealed that the 161 original questions involved four factors: concept activation (numerical and sensory), independent action, personal-social responsiveness, and associative vocabulary. It was decided to remove altogether the questions contributing to the "independent action" factor, and to weigh the "concept activation" factor doubly in the final 85-item test since it clearly accounted for the greatest number of shared variance. In the 85-item test there also was a preference for questions highly correlated with total test score, questions which varied in difficulty, and questions which were interesting to the children taking the test (Caldwell, 1967; Cooperative Tests and Services, 1970).

Since the original formulation of the test, the inventory has been reduced in length: first in 1968-69 when it was cut to 64 items (Revised Edition) and then again in 1970-71 when the HSPV 32-item version and a Follow Through 29-item version were created. Each of the revisions involved eliminating some of the original 85 items without adding new ones to the test. The 64-item version was also used in the 1968-69 Head Start national evaluation (Research Triangle, 1972), and in the first two years of the ETS longitudinal study (Shipman, 1972). The 29-item version was used in a 1971 Follow Through pilot study (Emrick, 1972). The three items of the 32-item PSI dropped in the 29-item version are "What is your first name" and "Color the triangle orange" (counts as 2 items). The 32-item version is also being used in the Home Start study (Hi/Scope, 1973).

Standardization

The original standardization of the 64-item PSI is based on the responses of 1531 children tested in fall 1969 in over 150 Head Start classes throughout the United States. This sample includes only children tested in English. Some regional data, based on 107 to 248 subjects per region per age level, are available. Percentile ranks for each age group and some of the regions are given in the manual (Cooperative Tests and Services, 1970). The number of children, mean scores and standard deviations for each of the age groups are summarized in Table 1.

TABLE 1

DISTRIBUTION OF 64-ITEM PSI SCORES FOR THE TEST STANDARDIZATION SAMPLE OF HEAD START CHILDREN IN FALL 1969[1]

Age Group	Number	Mean Score	S.D.
3-0 to 3-11	158	25.6	9.8
4-0 to 4-5	528	30.0	10.1
4-6 to 4-11	438	33.9	10.5
5-0 to 5-5	259	38.4	10.1
5-6 to 6.5	148	42.4	11.0

[1] Reported in Preschool Inventory Revised Edition - 1970: Handbook by Cooperative Tests and Services, 1970.

The Research Triangle Institute (1972) reports the scores on the 64-item PSI for the 1968-69 Head Start national evaluation sample. Mean scores of a subsample of 1162 children ranging in age from 2-7 to 6-0 years were slightly above those of the original standardization sample.

The 64-item PSI scores (mean and S.D.) for Year 1 and Year 2 of the ETS Longitudinal Study are presented in Table 2 (Shipman, et al., 1971; Shipman, 1972). The mean score for the year 1 sample (42-59 months) was 27.9 (S.D. = 11.9, N = 1974); the mean score for the year 2 sample (51-69 months) was 38.1 (S.D. = 12.3, N = 1311). In both years girls, who had a mean score of 29.1 (S.D. = 11.8) in Year 1 and a mean score of 40.0 (S.D. = 11.7) in Year 2, obtained significantly higher scores than boys, who had a mean score of 26.8 (S.D. = 11.9) in Year 1 and a mean score of 36.5 (S.D. = 12.6) in Year 2.

PSI (64-item) normative data for the Fall 1970 HSPV sample is reported in Tables 3-9. Norm tables based on three month age divisions (ten groupings from 42-44 months to 69-71 months) give the number of children, the mean score and the standard deviations at each age level for the following groupings in the HSPV sample: total (Table 3), children with previous preschool experience (Table 4), children with no previous preschool experience (Table 5), black children (Table 6), white children (Table 7), females (Table 8), and males (Table 9). The mean score for the total sample (N = 2134) was 35.188 (S.D. = 12.216). Children with previous preschool experience

TABLE 2

DISTRIBUTION OF 64-ITEM PSI SCORES FOR THE ETS HEAD START LONGITUDINAL SAMPLE[1]

Year 1

Age Group	Number	Mean Score[2]	S.D.
3-6 to 3-8	89	22.3	11.4
3-9 to 3-11	317	25.0	10.9
4-0 to 4-2	348	26.4	11.5
4-3 to 4-5	392	29.0	11.6
4-6 to 4-8	270	32.1	12.0
4-9 to 4-11	58	35.3	12.6
TOTAL	1974	27.9	11.9

Year 2

Age Group	Number	Mean Score[2]	S.D.
4-3 to 4-5	82	35.2	11.2
4-6 to 4-8	309	35.8	12.3
4-9 to 4-11	306	37.3	12.0
5-0 to 5-2	351	39.0	11.7
5-3 to 5-5	247	41.6	13.2
5-6 to 5-8	16	42.1	10.4
TOTAL	1311	38.1	12.3

[1] Reported in Shipman, 1972.

[2] Maximum score = 64.

had a higher mean (n = 407, mean = 40.4, S.D. = 11.7) than children with no previous preschool experience (n = 1727, mean = 34.0, S.D. = 12.0) (Tables 4 and 5). While children had a higher mean score (n = 759, mean = 37.5, S.D. = 12.8) than the black children (n = 1122, mean = 33.7, S.D. = 11.6) (Tables 6 and 7). Unlike the scores reported in the ETS Longitudinal Study by Shipman (1972), the mean scores for males and females were within one point of each other (Tables 8 and 9).

Norms for the 32-item PSI for the Fall 1971 HSPV sample are available in Tables 10-17. Based on 15 three-month age intervals from 36-38 months to 78-80 months, these tables give the number of children, the mean score, and the standard deviation at each age level for the following subgroups of the HSPV sample: total sample (Table 10), males (Table 11), females (Table 12), children with previous preschool experience (Table 13), children with no previous preschool experience (Table 14), white children (Table 15), black children (Table 16), and Mexican-American children (Table 17). The mean score for the total sample (N = 2972) was 14.585 (S.D. = 6.163). Scores in all the tables increased with age. The difference in mean scores for males (14.189, S.D. = 6.177) and females (14.995; S.D. = 6.121) was less than one point. There was a large difference (3 points or one-half of a standard deviation) between means for children with previous preschool experience (17.131, S.D. = 6.308) and children with no previous preschool

TABLE 3

DISTRIBUTION OF 64-ITEM PSI SCORES FOR ALL CHILDREN IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	8	20.500	11.533
45-47	63	24.667	10.788
48-50	204	27.108	9.955
51-53	316	30.203	10.914
54-56	341	33.179	11.605
57-59	348	35.589	11.047
60-62	270	38.867	11.468
63-65	228	39.978	11.123
66-68	180	41.006	10.905
69-71	176	43.244	10.918
TOTAL	2134	35.188	12.216

[1]Includes all children; not in Level I sites, Oraibi or Fresno; who had adequate information on sex, age, race and preschool experience.

[2]Maximum score = 64.

TABLE 4

DISTRIBUTION OF 64-ITEM PSI SCORES FOR ALL CHILDREN WITH PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	0	---	---
45-47	6	32.500	14.683
48-50	22	30.727	10.639
51-53	28	31.643	11.539
54-56	59	37.559	10.890
57-59	61	37.902	9.220
60-62	69	42.870	11.647
63-65	68	43.382	10.637
66-68	47	42.936	11.358
69-71	47	47.298	9.516
TOTAL	407	40.378	11.737

[1] Includes children; not in Level I sites, Oraibi or Fresno; who had adequate information on sex, race, age, and preschool experience.

[2] Maximum score = 64.

TABLE 5

DISTRIBUTION OF 64-ITEM PSI SCORES FOR ALL CHILDREN WITH NO PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1970 HSVP SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	8	20.500	11.533
45-47	57	23.849	9.940
48-50	182	26.670	9.779
51-53	288	30.063	10.841
54-56	282	32.262	11.540
57-59	287	35.098	11.338
60-62	201	37.493	11.078
63-65	160	38.531	11.009
66-68	133	40.323	10.657
69-71	129	41.767	11.022
TOTAL	1727	33.965	12.004

[1] Includes children; not in Level I sites, Oraibi or Fresno; who had adequate information on sex, race, age, and preschool experience.

[2] Maximum score = 64.

TABLE 6

DISTRIBUTION OF 64-ITEM PSI SCORES FOR BLACK CHILDREN IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	3	25.333	11.441
45-47	39	24.667	9.444
48-50	128	26.813	9.257
51-53	171	29.427	10.447
54-56	187	32.059	11.530
57-59	190	34.463	10.710
60-62	108	36.169	10.413
63-65	107	39.617	9.973
66-68	84	39.548	10.208
69-71	105	41.114	11.420
TOTAL	1122	33.774	11.622

[1]Includes children; not in Level I sites, Oraibi or Fresno; who had adequate information on sex, age, race, and preschool experience.

[2]Maximum score = 64.

TABLE 7

DISTRIBUTION OF 64-ITEM PSI SCORES FOR WHITE CHILDREN IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	5	17.600	10.575
45-47	24	24.667	12.671
48-50	67	28.209	11.433
51-53	118	31.492	11.568
54-56	124	35.548	11.658
57-59	117	38.410	10.957
60-62	108	41.407	11.546
63-65	86	42.965	10.870
66-68	60	45.583	10.185
69-71	50	47.600	8.911
TOTAL	759	37.510	12.802

[1]Includes children; not in Level I sites, Oraibi or Fresno; who had adequate information on sex, age, race, and preschool experience.

[2]Maximum score = 64.

TABLE 8

DISTRIBUTION OF 64-ITEM PSI SCORES FOR FEMALES IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	4	19.750	13.141
45-47	33	23.879	8.943
48-50	102	29.382	10.186
51-53	150	30.567	11.342
54-56	176	33.778	11.535
57-59	160	36.256	10.613
60-62	136	38.051	11.366
63-65	116	39.784	11.439
66-68	89	42.011	11.268
69-71	84	43.583	10.864
TOTAL	1050	35.605	12.127

[1] Includes children; not in Level I sites, Oraibi or Fresno; who had adequate information on sex, race, age, and preschool experience.

[2] Maximum score = 64.

TABLE 9

DISTRIBUTION OF 64-ITEM PSI SCORES FOR MALES IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	4	21.250	9.601
45-47	30	25.533	12.449
48-50	102	24.833	9.172
51-53	166	29.873	10.502
54-56	165	32.539	11.644
57-59	188	35.021	11.374
60-62	134	39.694	11.512
63-65	112	40.199	10.782
66-68	91	40.022	10.444
69-71	92	42.935	10.958
TOTAL	1084	34.785	12.289

[1] Includes children; not in Level I sites, Oraibi or Fresno; who had adequate information on sex, age, race, and preschool experience.

[2] Maximum score = 64.

TABLE 10

DISTRIBUTION OF 32-ITEM PSI SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	4	7.750	4.815
39-41	4	6.750	1.479
42-44	19	8.316	3.948
45-47	70	10.486	4.628
48-50	248	10.835	4.733
51-53	451	11.410	5.061
54-56	468	12.571	5.076
57-59	461	13.733	5.260
60-62	389	16.470	5.632
63-65	269	17.156	5.520
66-68	267	18.311	5.494
69-71	222	20.144	5.761
72-74	92	20.054	6.030
75-77	5	16.800	9.704
78-80	3	12.667	6.944
TOTAL	2972	14.585	6.163

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 32.

TABLE 11

DISTRIBUTION OF 32-ITEM PSI SCORES FOR MALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	------	-----
39-41	1	7.000	-----
42-44	6	8.000	3.464
45-47	43	10.605	5.297
48-50	123	10.927	4.811
51-53	235	10.821	4.951
54-56	241	12.166	5.138
57-59	233	13.129	5.188
60-62	207	16.295	5.538
63-65	124	16.726	5.669
66-68	142	17.669	5.524
69-71	103	19.670	6.307
72-74	49	19.898	6.149
75-77	3	24.667	1.247
78-80	2	17.000	4.000
TOTAL	1512	14.189	6.17

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 32.

TABLE 12

DISTRIBUTION OF 32-ITEM PSI SCORES FOR FEMALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	4	7.750	4.815
39-41	3	6.667	1.700
42-44	13	8.462	4.144
45-47	27	10.296	3.287
48-50	125	10.744	4.654
51-53	216	12.051	5.102
54-56	227	13.000	4.975
57-59	228	14.351	5.262
60-62	182	16.670	5.729
63-65	145	17.524	5.363
66-68	125	19.040	5.367
69-71	119	20.555	5.208
72-74	43	20.233	5.886
75-77	2	5.000	1.000
78-80	1	4.000	----
TOTAL	1460	14.995	6.121

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 32.

TABLE 13

DISTRIBUTION OF 32-ITEM PSI SCORES FOR ALL CHILDREN WITH PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	---	-----	-----
42-44	2	12.000	3.000
45-47	6	13.000	3.266
48-50	35	12.143	5.861
51-53	63	13.048	5.311
54-56	57	13.193	5.602
57-59	81	14.975	5.589
60-62	121	17.694	5.614
63-65	99	17.192	5.810
66-68	96	19.677	5.090
69-71	99	20.576	5.822
72-74	1	22.150	5.213
75-77	3	24.667	1.247
78-80	2	8.500	4.500
TOTAL	704	17.131	6.308

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 32.

TABLE 14

DISTRIBUTION OF 32-ITEM PSI SCORES FOR ALL CHILDREN WITH NO PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	4	7.750	4.815
39-41	4	6.750	1.479
42-44	16	7.625	3.789
45-47	63	10.159	4.647
48-50	207	10.628	4.512
51-53	374	11.176	4.932
54-56	397	12.542	5.024
57-59	368	13.380	5.112
60-62	257	15.887	5.556
63-65	162	17.019	5.395
66-68	165	17.442	5.586
69-71	119	19.924	5.545
72-74	52	18.442	6.119
75-77	2	5.000	1.000
78-80	1	21.000	---
TOTAL	2191	13.775	5.888

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 32.

TABLE 15

DISTRIBUTION OF 32-ITEM PSI SCORES FOR WHITE CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	------	-----
39-41	3	7.333	1.247
42-44	7	8.714	4.832
45-47	31	11.097	5.526
48-50	91	12.341	5.263
51-53	177	13.119	5.564
54-56	203	14.079	5.464
57-59	187	15.176	5.137
60-62	164	17.762	5.484
63-65	95	18.147	5.113
66-68	99	19.364	5.221
69-71	82	22.012	5.339
72-74	52	20.962	5.170
75-77	3	24.667	1.247
78-80	2	8.500	4.500
TOTAL	1196	15.977	6.172

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 32.

TABLE 16

DISTRIBUTION OF 32-ITEM PSI SCORES FOR BLACK CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	4	7.750	4.815
39-41	1	5.000	-----
42-44	12	8.083	3.303
45-47	39	10.000	3.693
48-50	135	9.830	4.130
51-53	202	10.124	4.313
54-56	203	11.148	4.338
57-59	194	12.711	5.153
60-62	149	15.101	5.311
63-65	123	16.041	5.235
66-68	115	17.913	5.320
69-71	117	18.838	5.756
72-74	38	19.026	6.819
75-77	2	5.000	1.000
78-80	1	21.000	-----
TOTAL	1335	13.382	5.918

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 32.

TABLE 17

DISTRIBUTION OF 32-ITEM PSI SCORES FOR MEXICAN AMERICAN CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	------	-----
39-41	---	------	-----
42-44	---	------	-----
45-47	---	------	-----
48-50	19	11.684	3.742
51-53	67	10.716	4.428
54-56	51	12.235	4.676
57-59	66	13.348	4.845
60-62	67	16.075	5.503
63-65	49	17.837	6.428
66-68	50	17.160	6.130
69-71	19	20.158	5.304
72-74	---	------	-----
75-77	---	------	-----
78-80	---	------	-----
TOTAL	388	14.528	5.973

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 32.

experience (13.775, S.D. = 5.888). The mean score for white children (15.977, S.D. = 6.172) was higher than the mean score for Mexican-American children (14.528, S.D. = 5.973) and for black children (13.382, S.D. = 5.918).

Reliability

64-item PSI. In general, the reliability estimates for the 64-item PSI are high. Two kinds of reliability estimates -- KR-20's and split-half (odd-even) coefficients, corrected for length by the Spearman-Brown formula -- are listed in Table 18 for each of the age groups in the standardization sample (Cooperative Tests and Services, 1970, p. 21).

The alpha coefficient for the total ETS sample was .92 in Year 1 (n = 1467) and .93 in Year 2 (n = 1311). The correlation between Year 1 and Year 2 scores was .66, one of the highest stability coefficients in the ETS study (Shipman, 1972).

Reliability estimates (KR-20's) for the total sample and subsamples in the Fall 1969 and Fall 1970 HSPV samples are listed in Tables 19 and 20. The KR-20 for the Fall 1969 sample was .925. The range of the 55 coefficients calculated for the Fall 1969 scores (Table 19) was from .825 (n = 15) for northern old white children with previous preschool experience to .938 (n = 175) for northern children with previous preschool experience. Only two coefficients out of 55 were below

TABLE 18

INTERNAL RELIABILITY ESTIMATES FOR THE 64-ITEM PSI TEST BASED ON THE HEAD START STANDARDIZATION SAMPLE IN FALL 1969.[1]

Age Group	n	KR-20	Corrected Split-Half
3-0 to 3-11	158	.88	.84
4-0 to 4-5	528	.88	.89
4-6 to 4-11	438	.86	.90
5-0 to 5-5	259	.89	.90
5-6 to 6.5	148	.92	.93
TOTAL	1531	.91	.92

[1] Reported by Cooperative Tests and Services, 1970, in Preschool Inventory Revised Edition - 1970: Handbook.

TABLE 19

KR-20 RELIABILITIES FOR FALL 1969 HSPV 64-ITEM PSI SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	1674	38.550	12.116	.925
Black	1163	37.017	12.194	.925
White	511	42.039	11.192	.915
Male	811	37.629	12.226	.925
Female	857	39.473	11.919	.924
Young[3]	799	33.229	11.076	.906
Old	875	43.409	10.931	.913
Previous Preschool	541	43.996	11.561	.925
No Previous Preschool	1133	35.950	11.501	.913
North	636	36.030	12.914	.935
South	1038	40.094	11.333	.914

[1]Sample includes all blacks and whites between 35 and 77 months at October 1 1969, who had a fall test score and data on the previous preschool experience question.

[2]Maximum score is 64.

[3]Young is less than 60 months old is greater than 59 months.

TABLE 20

KR-20 RELIABILITIES FOR FALL 1970 HSPV 64-ITEM PSI SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	2591	35.185	12.184	.924
Black	1314	33.808	11.515	.915
White	935	37.440	12.798	.933
Male	1337	34.632	12.214	.924
Female	1254	35.774	12.130	.924
Young[3]	1151	30.387	11.382	.911
Old	1439	39.035	11.413	.915
Previous Preschool	476	40.245	11.771	.923
No Previous Preschool	2108	34.003	11.983	.921
North	1503	34.239	12.081	.922
South	1088	36.491	12.211	.926

[1]Sample includes all children who were not in a Level I site or Oraibi between 32 and 79 months at October 1, 1970.

Note: 98% of the children were between 41 and 71 months. Spanish-speaking children are included in the total sample.

[2]Maximum score is 64.

[3]Young is less than 57 months; old is greater than 56 months.

.88. The KR-20 for the Fall 1970 sample was .924. The range of coefficients for the Fall 1970 PSI scores (Table 20) was from .832 (n = 45) for southern young blacks with previous preschool experience to .947 (n = 45) for young whites with previous preschool experience. Only four KR-20's were under .86.

32-item PSI. In Fall 1971 the 32-item version of the PSI was included in a test-retest/inter-tester reliability study conducted by the Huron Institute and SRI. Details of this study using two sites in the third year HSPV sample are reported in Appendix A. In general, the test-retest reliabilities were high and there were no significant tester effects. The range of test-retest coefficients for a sample of approximately 20 children after a two week interval was from .833 (paraprofessional B - paraprofessional B) to .952 (paraprofessional A - paraprofessional A). Internal consistency estimates (KR-20's) were high, considering the sample was small. The KR-20 was .84 for the test condition (n = 152) and .84 for the retest condition (n = 142).

Internal consistency coefficients (KR-20's) for the Fall 1971 HSPV total sample and main subsamples are listed in Table 21. The KR-20 for the total sample (n = 3176) was .824. The KR-20's for 92 subsamples with a size greater than 20 ranged from .681 for young black males with no previous preschool experience (n = 241) to .905 for Mexican-American females with previous preschool experience (n = 21). About two-thirds (67%) of these KR-20's were greater than .80 while only 4% were

TABLE 21

KR-20 RELIABILITIES FOR FALL 1971 32-ITEM PSI SCORES

	n	Mean[2]	S.D.	KR-20
Total[1]	3176	14.449	6.158	.824
Black	1415	13.224	5.912	.815
White	1277	15.876	6.173	.825
Mexican-American	425	14.337	6.004	.813
Male	1574	14.111	6.172	.826
Female	1526	14.896	6.156	.823
Young[3]	1338	11.565	5.082	.765
Old	1741	16.752	5.981	.811
Previous Preschool	760	16.896	6.400	.837
No Previous Preschool	2336	13.652	5.877	.810

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 32.

[3] Young is less than 57 months; old is greater than 56 months.

greater than .85. If the Spearman-Brown formula[1] is applied to the 32-item PSI reliability estimate (.824), the estimated reliability for a test double in length is .904. This estimate is almost identical to the KR-20's calculated for the Fall 1969 (KR-20 = .925) and Fall 1970 (KR-20 = .924) 64-item PSI scores.

The internal consistency reliability (alpha coefficient) for the Fall 1972 Home Start sample (n not given) was .83 (Hi/Scope, 1973).

29-item PSI. In the Fall 1971 Follow Through evaluation (Emrick, 1972), the 29-item version of the PSI was included in a supplementary battery given to kindergarten and entering first grade children in 17 projects. The measures of internal consistency were adequate for the test and retest given two to three weeks later. The range of KR-20 coefficients was .673 to .964 (average .834) for the test condition and .562 to .933 (average .839) for the retest condition. The test-retest coefficient for the entire sample (n = 597) after a 2-3 week interval was .845.

Item and Score Characteristics

64-item PSI. In the standardization data test difficulty is measured by expressing mean raw scores as a percentage of

---

[1] $r^* = \frac{2r}{1+r}$, where r^* = estimated reliability for test double and r = reliability of test.

the total number of questions on the test. Three-year-olds got an average of forty percent of test questions correct while children five-and-a-half to six-and-a-half got an average of about sixty-six percent correct. The standardization sample did not indicate ceiling effects, although further subgroup analyses might reveal such effects for certain older groups. Mean biserial correlations between each item and total score of children increased with age, ranging from .45 (3-0 to 3-1) to .56 (5- to 6-5). (Cooperative Tests and Services, 1970)

Close analyses of the frequency distributions for fall 1969 and spring 1970 PSI scores of four HSPV subsamples (young white, young black, old white, old black) reveal there is a ceiling effect in the spring scores of older white children (see Table 22). In spring 1970 twenty percent of the older white children were at the three top scores (score 62--6%; score 63--10%; score 64--4%).

Factor analyses done by Shipman et al. (1971) on the ETS Longitudinal sample and by the Huron Institute on the fall 1969 HSPV sample do not find the four factors which were found in the original study. The factor analysis done on one subgroup in the 1969-70 HSPV data (older blacks in Fall 1969) revealed the existence of only one factor. The first three eigenvalues obtained were 9.30, 2.54, and 2.15. The first value accounted for 14.5% of the total variance.

TABLE 22

PARTIAL FREQUENCY DISTRIBUTION OF PSI SCORES FOR OLD[1] WHITE CHILDREN IN SPRING 1970.

Score	Number of Children	Percent of Children	Cum. Percent of Children	
50	3	1	20	XXX
51	8	3	23	XXXXXXXX
52	9	3	26	XXXXXXXXX
53	11	4	30	XXXXXXXXXXX
54	11	4	34	XXXXXXXXXXX
55	11	4	38	XXXXXXXXXXX
56	10	4	42	XXXXXXXXXX
57	19	7	49	XXXXXXXXXXXXXXXXXXX
58	28	10	59	XXXXXXXXXXXXXXXXXXXXXXXXXXX
59	16	6	65	XXXXXXXXXXXXXXX
60	19	7	72	XXXXXXXXXXXXXXXXXX
61	23	8	81	XXXXXXXXXXXXXXXXXXXXX
62	16	6	87	XXXXXXXXXXXXXXX
63	26	10	96	XXXXXXXXXXXXXXXXXXXXXXXXX
64	10	4	100	XXXXXXXXX

Total n = 271

[1] Old is greater than 59 months.

32-item PSI. The distribution of 32-item PSI scores for the Fall 1971 and Spring 1972 HSPV samples were normally-shaped. There was no indication of a ceiling effect for the total sample or any possible subsamples in either fall or spring.

A principal components analysis followed by a varimax rotation of the Fall 1971 HSPV 32-item PSI scores substantiated the existence of one general factor. The analysis produced 9 eigenvalues greater than 1.0; the sum of 9 factors accounted for 47% of the variance. The largest eigenvalue was 5.973 which accounted for 19% of the total variance. The next eight eigenvalues ranged from 1.452 to 1.002. A similar factor analysis done on the Home Start Fall 1972 data confirms the HSPV finding of one factor. In the Home Start analysis, the first factor accounted for 18% of the total variance (Hi/Scope, 1973).

The percent passing each item of the 32-item PSI for five age groups (3-1/2, 4, 4-1/2, 5, 5-1/2 years) of the Fall 1971 HSPV sample are listed in Table 23. The most difficult items across all the age groups were items #17 (How many toes do you have?), #24 (Which of 2 groups has more checkers?), #7 (Put 2 cars behind the middle box), #10 (Where would you look for a lion?), and #22 (Point to the second checker). The easiest item for all age groups was item #1 (What is your first name?). Other items which were relatively easy for all age groups were #2 (Show me your shoulder), #25 (Point to the drawing most like

TABLE 23

32-ITEM PRESCHOOL INVENTORY: PERCENT PASSING EACH ITEM

ITEM	Ages[1] 3-1/2	4	4-1/2	5	5-1/2
1	91	93	89	90	93
2	69	65	69	78	87
3	51	60	65	75	83
4	30	38	45	60	67
5	20	36	43	52	61
6	08	21	22	34	40
7	06	08	10	17	25
8	51	49	56	64	71
9	24	31	36	49	57
10	06	16	24	34	36
11	26	25	38	53	60
12	12	32	35	44	48
13	08	21	24	33	40
14	32	48	52	63	71
15	42	47	54	63	73
16	08	19	24	38	54
17	00	02	03	07	14
18	34	53	56	61	66
19	20	36	44	56	72
20	18	30	31	41	48
21	14	20	25	33	47
22	08	15	17	20	33
23	22	37	36	44	42
24	00	04	05	12	16
25	59	66	68	78	81
26	02	14	22	38	57
27	02	06	08	20	35
28	59	58	58	68	75
29	18	33	39	44	50
30	30	31	33	42	49
31	22	37	41	49	60
32	22	50	54	67	77
n =	49	501	912	835	521

[1] Intervals include 2 months before and 4 months after indicated age (e.g., 4-year-old category includes children from 46 to 51 months).

a tent), and #28 (Which one [of the crayons] is the color of night?). For the remaining items, the percent passing increased with age.

The percent passing each item of the 32-item PSI were also computed for the Home Start pilot data (Hi/Scope, 1973). In general, the findings were very similar to the HSPV results. The most difficult items for all ages were #7, #17, #24, #26 (Make one like this [point to square].) and #12 (Which way does a phonograph record go?). The easiest item for all ages was item #1.

Item intercorrelations and item-total correlations for the HSPV Fall 1971 data are listed in Table 24. In general all of the item intercorrelations were low. None were negative and the few highest were in the .40's. The item intercorrelations computed for the Home Start data (Hi/Scope, 1973) were also low; a few of these correlations were negative.

The item-total correlations (not corrected for overlap) for the HSPV data ranged from .14(item #1 - What is your first name?) to .59 (item #19 - Point to the middle checker). The mean item total correlations was .42. Seven of the correlations were greater than .50, and two (item #14 and item #23 - Which of these two groups has less checkers in it?) were less than .20.

The item-total correlations (corrected for overlap) for the Home Start data (Hi/Scope, 1973) ranged from .03 (item #22 - Point to the second checker) to .54 (item #6 - Put the blue car under the green box, and item #19 - Point to the

TABLE 24

32-ITEM PRESCHOOL INVENTORY TOTAL/ITEM INTERCORRELATIONS

	Total	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	.14																															
2	.39	.07																														
3	.43	.07	.24																													
4	.53	.03	.25	.40																												
5	.52	.04	.20	.20	.27																											
6	.53	.04	.18	.19	.28	.40																										
7	.43	.01	.12	.10	.18	.24	.33																									
8	.38	.06	.15	.19	.20	.14	.14	.10																								
9	.42	.07	.15	.17	.19	.18	.17	.19	.17																							
10	.47	.02	.14	.17	.27	.16	.20	.17	.20	.20																						
11	.51	.06	.16	.21	.29	.21	.24	.18	.25	.25	.29																					
12	.42	.04	.13	.14	.16	.16	.16	.12	.16	.16	.24	.23																				
13	.43	.02	.12	.12	.17	.15	.16	.15	.15	.16	.24	.21	.43																			
14	.45	.04	.16	.16	.19	.20	.22	.19	.15	.14	.22	.21	.16	.16																		
15	.38	.04	.15	.16	.17	.16	.16	.09	.12	.13	.13	.17	.15	.17	.27																	
16	.48	.05	.13	.15	.23	.23	.25	.18	.13	.18	.20	.21	.16	.18	.19	.14																
17	.36	.03	.11	.12	.16	.18	.18	.14	.10	.13	.12	.14	.12	.15	.13	.11	.24															
18	.27	.02	.09	.10	.10	.10	.09	.09	.08	.07	.11	.14	.09	.09	.11	.08	.12	.09														
19	.59	.04	.22	.23	.33	.30	.31	.24	.18	.18	.24	.26	.18	.20	.26	.19	.26	.17	.10													
20	.45	.06	.13	.12	.17	.21	.22	.20	.10	.15	.15	.19	.11	.14	.16	.11	.20	.15	.10	.37												
21	.49	.04	.14	.17	.22	.21	.25	.21	.13	.15	.17	.19	.13	.15	.17	.13	.22	.20	.09	.44	.24											
22	.38	.02	.08	.10	.14	.13	.18	.15	.08	.12	.12	.11	.10	.12	.13	.10	.17	.18	.06	.27	.18	.44										
23	.18	.01	.05	.01	.04	.10	.04	.02	.04	.06	.05	.02	.01	.02	.05	.05	.05	.05	.01	.06	.07	.05	.02									
24	.28	.05	.08	.08	.08	.12	.15	.15	.03	.08	.04	.06	.02	.09	.10	.07	.15	.15	.07	.11	.15	.12	.15	.12								
25	.30	.02	.11	.08	.11	.11	.12	.12	.10	.08	.12	.14	.12	.12	.09	.06	.08	.06	.05	.14	.13	.11	.08	.09	.08							
26	.52	.05	.12	.17	.22	.22	.22	.18	.15	.17	.23	.23	.20	.21	.19	.17	.26	.19	.11	.30	.23	.22	.17	.02	.13	.09						
27	.43	.02	.13	.15	.18	.17	.17	.19	.11	.14	.15	.17	.13	.12	.13	.13	.24	.18	.06	.21	.17	.19	.16	.05	.16	.09	.46					
28	.36	.06	.13	.10	.14	.17	.16	.13	.12	.13	.15	.13	.13	.13	.13	.10	.13	.09	.05	.18	.13	.15	.10	.03	.09	.11	.16	.14				
29	.44	.02	.10	.08	.19	.18	.19	.21	.12	.16	.20	.18	.16	.14	.14	.10	.18	.12	.08	.16	.15	.15	.14	.05	.12	.09	.21	.17	.12			
30	.39	.03	.10	.14	.18	.20	.21	.14	.10	.13	.13	.14	.11	.10	.13	.06	.16	.16	.04	.17	.12	.18	.12	.05	.12	.09	.18	.17	.14	.14		
31	.49	.05	.14	.16	.23	.22	.23	.18	.13	.19	.23	.22	.16	.17	.19	.13	.18	.12	.11	.22	.18	.17	.16	.03	.13	.12	.26	.20	.13	.45	.16	
32	.53	.04	.18	.22	.27	.28	.29	.19	.16	.18	.22	.25	.17	.16	.21	.17	.21	.15	.09	.33	.18	.21	.13	.04	.11	.14	.27	.22	.15	.22	.28	.26

n = 3176 (total HSPV sample)

middle checker). The mean item-total correlation was .34. Three of the correlations were greater than .50, and five were less than .20.

29-item PSI. The mean score for the overall Follow Through fall 1971 supplementary battery sample (n = 651) for the test condition was 16.7 (about 58% correct) with a standard deviation of 5.71. Initial test mean scores for the 17 projects ranged from 12.6 to 20.0. The overall difficulty levels, which ranged generally from 30% to 70%, seemed appropriate for this sample. The few items which were the easiest or most difficult on the pretest were also the easiest or most difficult on the posttest. The easiest items are at the beginning of the test. There did not seem to be any ceiling or floor effects in the scores for this sample (Emrick, 1972).

Correlations with Other Tests

64-item PSI. Correlations of the 64-item PSI scores and Stanford-Binet IQs are available for the standardization sample (Cooperative Tests and Services, 1971). The correlations ranged from .39 (ages 3-0 to 3-11) to .65 (ages 5-0 to 5-5).

Correlations of the 64-item PSI scores with the NYU Booklets 3D and 4A, the Motor Inhibition Subtests, the CPSCS, the Eight-Block Sort success scores, and the Stanford-Binet

IQ and MA for the fall 1970 HSPV sample are listed in Table 25. Correlations with the NYU Booklet 3D (.696) and the Stanford-Binet MA (.756) are the highest. The former correlation is a good concurrent validity estimate for the PSI, since the 3D Booklet is an achievement test which measures many similar relational concepts. The lower correlation (.467) between the 4A Booklet and the PSI is not surprising since the PSI only includes a few items of recognizing numbers, letters, and shapes. If the correlations with the NYU Booklets are corrected for unreliability[1], the estimated correlations between the true score components of the tests are .90 for Booklet 3D and .59 for Booklet 4A.

The correlations of the PSI with the other tests in the ETS Longitudinal Study (Shipman, 1972) support its use as an achievement measure. Correlations with other cognitive-perceptual tests were the highest. The correlation with the PPVT was .58 in Year 1 and .66 in Year 2. Other 64-item PSI correlations of interest in the ETS Study are .47 in Year 1 and .53 in Year 2 with the Eight-Block Sort Total Success Score, .30 in Year 1 with ETS Enumeration I (pointing items), and .58 in Year 2 with ETS Enumeration II (counting items). In the factor analyses of Year 1 and Year 2 data, the PSI had the

---

[1] Using formula $\frac{r_{1\cdot 2}}{\sqrt{t_1 \cdot t_2}} =$ where $r_{1\cdot 2}$ is the correlation between the two tests and t_1, t_2 are the reliability estimates for the two tests.

TABLE 25

INTERCORRELATIONS OF FALL 1970 SCORES FROM THE CPSCS, NYU BOOKLETS 3D AND 4A, 64-ITEM PSI, MI SUBTESTS, EIGHT-BLOCK SORT SUCCESS SCORES, AND THE STANFORD-BINET IQ AND MA[1]

	CPSCS	NYU 3D	NYU 4A	64-ITEM PSI	MI[2] WALK	MI DRAW	MI TRUCK	MI WALK+ DRAW	EB Placement	EB Reason	EB Success Total	IQ[3]
NYU 3D	.297 (2057)											
NYU 4A	.240 (2045)	.429 (2125)										
64-ITEM PSI	.390 (2064)	.696 (2127)	.467 (2117)									
MI WALK[2]	.054 (1024)	.275 (1073)	.142 (1072)	.279 (1074)								
MI DRAW	.078 (1028)	.298 (1078)	.142 (1077)	.356 (1079)	.459 (1073)							
MI TRUCK	.056 (1015)	.136 (1065)	.106 (1065)	.165 (1066)	.255 (1058)	.223 (1063)						
MI (WALK+ DRAW)	.080 (1022)	.326 (1071)	.158 (1070)	.370 (1072)	.772 (1073)	.901 (1073)	.280 (1056)					
EB Placement	.114 (547)	.171 (554)	.116 (552)	.254 (556)	.088 (276)	.207 (278)	.067 (273)	.170 (276)				
EB Reason	.115 (547)	.260 (554)	.159 (552)	.333 (556)	.098 (276)	.166 (278)	.093 (273)	.158 (276)	.424 (576)			
EB Success Total	.134 (547)	.266 (554)	.168 (552)	.356 (556)	.110 (276)	.212 (278)	.097 (273)	.189 (276)	.739 (576)	.924 (576)		
IQ	.321 (752)	.427 (753)	.365 (749)	.510 (752)	.152 (352)	.229 (352)	.120 (348)	.203 (350)	.361 (44)	.109 (44)	.221 (44)	
MA	.373 (753)	.640 (754)	.435 (750)	.756 (753)	.259 (352)	.436 (352)	.032 (348)	.396 (350)	.297 (44)	.378 (44)	.387 (44)	.719 (773)

1 Sample size for each correlation is included in the parenthesis. Children included in the sample were those not in Level I sites, Oraibi, or Fresno; who had adequate information on age, sex, race, and preschool experience. Only children between 43 and 74 months who attended preschool for the full year were included. Only completed tests with valid codes were used.

2 MI scores are log transformations of the "slow" times: A child's MI scores were used if he had passed two out of the four pretests.

3 From Pinneau's revised IQ tables (see Terman and Merrill, 1960).

highest loading of any measure on the "g" factor (general information-processing skills).

32-item PSI. Correlations of the 32-item PSI with other tests in the Fall 1971 HSPV battery are presented in Table 26. The PSI had the highest correlations of any test in the battery. The PSI correlated highest with the PPVT (.665), the ETS Enumeration-Counting Subtest (.625) and the ETS Enumeration Total Score (.584). Correlations with the Fall WRAT subtests were in the .40-.50 range. The PSI correlations with the Eight-Block Sort scores were .305 (Placement), .443 (Reason), and .440 (Total). Correlations with the Brown Self-Concept Test, the MI-Truck Subtest, and the Touching and Same Number Matching Subtests of the ETS Enumeration Test were low.

29-item PSI. Correlations of the 29-item PSI with the Brown Self-Concept Test, the ITPA-Verbal Expression Subtest, and Faces Test were calculated for the fall 1971 Follow Through supplemental battery sample (Emrick, 1972). Correlations with the Brown were .293 (test) and .378 (retest); with Faces, .315 (test) and .334 (retest); and with the ITPA-Verbal Expression Subtest, .556 (test) and .517 (retest).

Remarks

The PSI is one of the best tests in the HSPV battery. It is unpretentious about what it is trying to measure, and because it assesses concrete attainments and verges on being a criterion-reference measure, it can claim a face-validity

TABLE 26 INTERCORRELATIONS OF FALL 1971 SCORES FROM THE PPVT, WRAT SUBTESTS, 32-ITEM PSI, ITPA VERBAL EXPRESSION SUBTEST, ETS ENUMERATION SUBTESTS, BROWN, MI-TRUCK SUBTEST, AND EIGHT-BLOCK SORT SUCCESS SCORES[1]

	PPVT	WRAT-COPY MARKS	WRAT-RECOG. LETTERS	WRAT-NAME LETTERS	WRAT-READ #'s	WRAT-DOT COUNT.	PSI 32-ITEM	ITPA-VERBAL EXPRESS.	ETS. ENUM. TOTAL	ETS. ENUM. COUNT.	ETS ENUM. TOUCH.	ETS ENUM. SAME # MATCH.	BROWN UNADJ.	BROWN ADJ.	MI-TRUCK	EIGHT-BLOCK PLACE.	EIGHT-BLOCK REASON
WRAT-COPY MARKS	.413 (2881)																
WRAT-RECOG. LETTERS	.537 (2881)	.375 (2995)															
WRAT-NAME LETTERS	.346 (2881)	.358 (2995)	.302 (2995)														
WRAT-READ NUMBERS	.407 (2881)	.412 (2995)	.325 (2995)	.600 (2995)													
WRAT-DOT COUNTING	.453 (2881)	.463 (2995)	.419 (2995)	.344 (2995)	.451 (2995)												
PSI (32-item)	.665 (2855)	.551 (2860)	.481 (2860)	.414 (2860)	.508 (2860)	.589 (2860)											
ITPA-VERBAL EXPRESSION	.487 (1147)	.339 (1172)	.371 (1172)	.276 (1172)	.341 (1172)	.388 (1172)	.506 (1138)										
ETS ENUMERATION TOTAL[2]	.475 (1075)	.508 (1097)	.427 (1097)	.307 (1097)	.446 (1097)	.542 (1097)	.584 (1073)	.459 (1115)									
ETS ENUMERATION COUNTING	.492 (1075)	.504 (1097)	.422 (1097)	.359 (1097)	.500 (1097)	.620 (1097)	.625 (1073)	.384 (1115)	.781 (1135)								
ETS ENUMERATION TOUCHING	.282 (1075)	.358 (1097)	.293 (1097)	.196 (1097)	.271 (1097)	.383 (1097)	.382 (1073)	.308 (1115)	.721 (1135)	.390 (1135)							
ETS ENUMERATION SAME # MATCHING	.237 (1075)	.225 (1097)	.199 (1097)	.095 (1097)	.176 (1097)	.148 (1097)	.232 (1073)	.298 (1115)	.664 (1135)	.257 (1135)	.202 (135)						
BROWN-UNADJUSTED	.322 (2689)	.162 (2753)	.243 (2753)	.145 (2753)	.173 (2753)	.270 (2753)	.323 (2689)	.261 (1145)	.228 (1073)	.271 (1073)	.166 (1073)	.054 (1073)					
BROWN-ADJUSTED	.239 (2689)	.127 (2753)	.166 (2753)	.100 (2753)	.124 (2753)	.194 (2753)	.259 (2689)	.215 (1145)	.159 (1073)	.172 (1073)	.134 (1073)	.034 (1073)	.637 (2679)				
MI-TRUCK[3]	.174 (607)	.061 (625)	.048 (625)	.083 (625)	.121 (625)	.006 (625)	.164 (608)	.032 (637)	.136 (597)	.136 (597)	.047 (597)	.107 (597)	.118 (610)	.102 (610)			
EIGHT-BLOCK PLACEMENT	.304 (1119)	.222 (1148)	.271 (1148)	.145 (1148)	.207 (1148)	.304 (1148)	.305 (1090)	.303 (1096)	.322 (1032)	.413 (1032)	.200 (1032)	.180 (1032)	.212 (1113)	.183 (1113)	.005 (573)		
EIGHT-BLOCK REASON	.445 (1119)	.364 (1148)	.333 (1148)	.286 (1148)	.372 (1148)	.390 (1148)	.443 (1090)	.416 (1096)	.405 (1032)	.402 (1032)	.258 (1032)	.211 (1032)	.178 (1113)	.168 (1113)	.063 (573)	.520 (1211)	
EIGHT-BLOCK SUCCESS TOTAL	.439 (1119)	.346 (1148)	.351 (1148)	.257 (1148)	.344 (1148)	.404 (1148)	.440 (1090)	.422 (1096)	.422 (1032)	.416 (1032)	.266 (1032)	.226 (1032)	.220 (1113)	.200 (1113)	.046 (573)	.839 (1211)	.901 (1211)

1 Sample size for each correlation is included in parenthesis. Children in sample are those with adequate information not in Level I sites.

2 ETS ENUMERATION Score= sum of counting, touching and same number matching subtest scores.

3 MI scores are log transformations of slow times.

that other tests cannot. The PSI seems to be an adequate measure of a young child's achievement. The 32-item version may be more useful in future large scale evaluations than the 64-item version since, in addition to being shorter to give, it has adequate reliability coefficients without any ceiling effects in scores. Despite the excellent technical information already available on various forms of the PSI, there are some limitations which need to be resolved in future studies:

1. The ETS Study (Shipman, 1972) finding that there are large differences among SES groups on the 64-item PSI indicates that experience is necessary for the development of general knowledge and substantiates the test designers' claim that the test is not "culture-fair." The designers' refusal to create a culture-fair test was based on the assumption that there are a number of skills which every child, whatever his background, will have to possess to be successful in kindergarten. It was argued that such a test should reflect the biases of the school rather than mask them, since all children sooner or later have to succeed or fail according to school-defined notions of achievement. This assumptions seems defensible, and even laudable, if the test really does tap generally necessary skills and knowledge. But some critics have suggested that the answers to certain PSI items reflect regional or ethnic biases which do not have any influence on school success. Thus, when a child is asked where he would expect to find a lion, he might answer, "in a book" or "in the woods" as easily as "in

the zoo"; but by the PSI scoring system such an answer would be marked wrong. Likewise, in the case of the question, "Who do you go to when you feel sick?" it is wrong in the PSI for the child to say he would go to the hospital. The correct answer is "to a doctor" or "to a nurse." Continued item development should rectify such problems.

2. On some PSI items, the child is required to identify or reproduce two or more attributes simultaneously in giving his answer. A problem arises since some items are scored to allow a partially correct answer and some are not. Thus, on the test item requesting that the child "color the triangle orange", one point is given for selecting the correct geometric configuration and another for using the right color. But on the item requesting that the child "put the yellow car on the little box", the child's response is either marked entirely correct or entirely incorrect, regardless of the fact that a judgment of color, of size, and of relation must be made. Critics suggest that credit should always be given for the understanding of individual task dimensions.

3. The PSI may have stronger practice effects than other tests. Study of such effects is needed.

4. Predictive validity estimates are needed for all forms of the PSI.

References

Caldwell, B. The Preschool Inventory: Technical Report. Princeton, N.J.: Educational Testing Service, 1967.

Cooperative Tests and Services. Preschool Inventory Revised Edition - 1970: Handbook. Princeton, N.J.: Educational Testing Service, 1970.

Educational Testing Service. Disadvantaged children and their first school experiences: preliminary description of the initial sample prior to school enrollment. Princeton, N.J.: Educational Testing Service, 1970.

Emrick, J. A. Results of Fall 1971 Follow Through Supplemental Testing. Stanford Research Institute report to the Bureau of Elementary and Secondary Education, Office of Education, H.E.W. under contract # OEC-0-8-522480-4633. Menlo Park, Cal.: SRI, 1972.

Hi/Scope Foundation/Abt Associates. Home Start Evaluation Study. Draft interim report to Office of Child Development, H.E.W. under contract # HEW-OS-72-127. Ypsilanti, Michigan: Hi/Scope Foundation/Abt Associates, 1973.

Research Triangle Institute. A report on two national samples of Head Start classes: Some aspects of child development of participants in full year 1967-1968 and 1968-1969 programs. Final report to Office of Child Development, H.E.W. Research Triangle Park, N.C.: R.I.T., 1972.

Shipman, V. C., J. Barone; A. Beaton, W. Emmerich and W. Ward. Disadvantaged children and their first school experiences: Structure and development of cognitive competencies and styles prior to school entry. Princeton, N.J.: Educational Testing Service, 1971.

Shipman, V. C., ed. Disadvantaged children and their first school experiences: Technical report series. Prepared for the Office of Child Development, H.E.W. under grant # H-8526. Princeton, N.J.: Educational Testing Service, 1972.

Relevant Redundant Cue Concept Acquisition Task

Purpose

The Relevant Redundant Cue Concept Acquisition Test (RRC), also called "Zings and Poggles," was designed to measure concept acquisition, learning ability, and attention to the learning process. Tests of this type provide a means of studying inductive reasoning processes in children. In addition to showing something about learning rates, the concept acquisition task yields information about the strategies the child uses in learning a task in which two or more dimensions are redundant. The concept acquisition strategies of children seem especially important to study during an age period when these strategies are hypothesized to be changing (Weir, 1964).

Description

The task consists of 64 cards on which circles, rectangles and triangles are drawn. The first 48 cards are used as part of the "training series" while the remaining 16 are used as the "transfer series" or test. In the training series, the child is shown a set of cards one at a time. The child is asked to guess if the card is a "Zing" (green and/or rectangle) or a "Poggle" (red and/or diamond). During the training period the child is told if his guess is correct and encouraged to study the cards to determine the difference between "Zings" and "Poggles." The training period is continued until the

entire deck is exhausted or until the child has given twelve consecutive correct responses. During the "transfer" or "testing" period the child is asked to identify the "zings" and "poggles" in a set of 16 cards with no help from the tester. A score of "1" is given for each correct answer in the "transfer series." All other responses during the testing/transfer series are coded as follows: refusal, don't know, request aid, no response, black, green, red, other color, oval, circle, square, rectangle, diamond, other shape, other name.

Development of instrument

The RRC was developed by Educational Testing Service in the late 1960's for use with four-to-nine-year-old children in their longitudinal study of disadvantaged children (1968). Analyses of the RRC results will be published in a future report on year II of their study. Since the RRC is a new instrument, no other researchers have used it in studies.

Reliability

Internal consistency reliability coefficients (KR-20's) for a portion of the Spring 1972 Head Start Planned Variation sample are listed in Table 1. The KR-20 for the total sample (n = 803) was .203. The estimates for approximately 85 subsamples with a sample size greater than 20 ranged from .021 for older white males with no previous preschool experience (n = 62) to .556 for older white males with previous preschool experience (n = 26). Only 10% of the KR-20 estimates were greater than .40. Most of them were under .20 and a few were

TABLE 1

KR-20 RELIABILITIES FOR SPRING 1972 HSPV RRC SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	03	8.824	2.207	0.203
Black	46	8.662	2.240	0.223
White	318	8.865	2.204	0.207
Mexican-American	129	9.116	2.130	0.156
Male	416	8.964	2.179	0.190
Female	387	8.674	2.227	0.216
Young[3]	322	8.770	2.199	0.192
Old	477	8.853	2.209	0.211
Previous Preschool	192	8.656	2.520	0.385
No Previous Preschool	592	8.873	2.104	0.129

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 16.

[3] Young is less than 57 months; old is greater than 56 months.

negative. This random fluctuation of estimates around zero indicates that there was a great deal of guessing occurring on the test.

Head Start Planned Variation Score Characteristics

The distribution of the Relevant Redundant Cue Scores for all children in the Spring 1972 sample is presented in Table 2. The mean score and standard deviation for each three month age interval from 36-38 months to 78-80 months are included. The mean score for the total sample (n = 799) is 8.820 (S.D. = 2.205).

The distributions of scores in the spring for all planned variation children and all non-planned variation children are bimodal (see Tables 3 and 4). These distributions may be explained by the fact that children's scores are partially determined by guessing and/or by knowing only one of the two dimensions of a "zing" or a "poggle". If the children were only guessing, the scores would have been lower. If a child knew one of the two dimensions of a "zing" or a "poggle", he would get approximately one-half of the items correct all the time. If the child knew one characteristic (such as a "zing" is green) and guessed on items without that characteristic (such as a black rectangle), he would get a score slightly under or over the mean.

TABLE 2

DISTRIBUTION OF PRC (ZINGS AND POGGLES) SCORES FOR ALL CHILDREN IN THE SPRING 1972 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	1	7.000	-----
39-41	1	11.000	-----
42-44	5	8.200	1.720
45-47	17	8.235	1.733
48-50	65	8.938	2.423
51-53	113	8.823	2.215
54-56	120	8.725	2.117
57-59	125	8.704	2.102
60-62	104	9.096	1.949
63-65	77	9.026	2.313
66-68	85	8.882	2.198
69-71	53	8.528	2.270
72-74	32	8.563	2.783
75-77	1	13.000	-----
78-80	---	-----	-----
TOTAL	799	8.820	2.205

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 16.

TABLE 3

DISTRIBUTION OF RRC SCORES FOR ALL PLANNED VARIATION CHILDREN IN SPRING 1972

Score	# of children	
0	5	x , Each x represents 5 children
1	0	
2	3	x
3	1	
4	7	x
5	19	xxxx
6	32	xxxxxx
7	78	xxxxxxxxxxxxxxx
8	114	xxxxxxxxxxxxxxxxxxxxxx
9	88	xxxxxxxxxxxxxxxxx
10	102	xxxxxxxxxxxxxxxxxx
11	45	xxxxxxxxx
12	63	xxxxxxxxxxx
13	13	xx
14	0	
15	2	
16	1	

N = 573

TABLE 4

DISTRIBUTION OF RRC SCORES FOR NON-PLANNED VARIATION CHILDREN IN SPRING 1972

Score	# of children	
0	1	x
1	0	
2	0	
3	1	x
4	2	xx
5	7	xxxxxxx
6	13	xxxxxxxxxxxxx
7	39	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
8	40	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
9	32	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
10	56	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
11	12	xxxxxxxxxxxx
12	23	xxxxxxxxxxxxxxxxxxxxxxx
13	3	xxx
14	1	x
15	0	
16	0	

N = 230

Remarks

Because there is not yet available technical information on this instrument from the ETS Longitudinal Study and only a limited amount of information from the Head Start Planned Variations Study since it was only given in the spring, the Relevant Redundant Cue Test must be considered as an experimental instrument in the beginning stages of development. From the limited information available, it appears that the test is too difficult for young children; perhaps it should only be used with older children in future evaluations.

References

Educational Testing Service. Disadvantaged children and their first school experiences: theoretical considerations and measurement strategies. Princeton, N. J.: ETS, 1968.

Weir, M. Developmental changes in problem solving strategies. Psychological Review, 1964, 71, 473-90.

Stanford-Binet Intelligence Scale

Purpose

The Stanford-Binet Intelligence Scale is a measure of "general intelligence" which is widely used in the United States. Although it is called a test of intelligence, it is just as much a measure of experience and achievement. It is most often defined as a measure of general mental adaptability for populations exposed to similar experiences. It has high predictive validity in terms of future school success.

Description

The Stanford-Binet Intelligence Scale, revised edition, Form L-M, consists of different subtests graduated in difficulty according to age. It is an age scale test based on the assumption that general intelligence is a trait that develops with age. The primary criteria used in constructing such a test are that the subtests be arranged in a scale so that the mean mental age of unselected subjects is the same as their mean chronological age and that the variability of IQ scores remains approximately constant from age to age. Early subtests (ages 2-3) contain non-verbal tasks such as building blocks, the three hole board, and stringing beads. Later subtests have more verbal tasks such as vocabulary, analogies, and number problems. A complete description of the subtests is in the manual by Terman and

Merrill (1960). Basal age is that level at which all tests are passed which just precedes the level where the first failure occurs. After a child's mental age (MA) is determined, it is converted to an IQ estimate using Pinneau's revised IQ tables (Terman & Merrill, 1960). Only experienced Binet testers were used in the HSPV Study.

Development of Instrument

The Stanford-Binet was developed by Alfred Binet in the late 1800's to identify the mentally defective children in the Paris elementary schools. L. Terman published the first revision of the original scale in 1916, standardizing it for American children, ages 3-16. Terman defined Intelligence Quotient as the ratio of mental age to chronological age (MA/CA). In 1937 Terman and Merrill revised the test again, making use of the age standards of performance gathered from the previous test data. At this time two forms (L and M), differing in content but not in type of question, were developed. The test was last revised in 1960 at which time the two separate forms (L and M) were combined. Emphasis was placed on correlation between individual subtest items and total score. In selecting items for the L-M Form, factor loadings of McNemar's analysis of the 1937 revision were taken into account. Items were updated and those judged obsolete were replaced with more current items.

Technical Information

Much of the reliability and validity of the 1960 scale revision rests on the 1937 scale. The 1937 scale has been found to be more reliable for older than for younger children and for lower than for higher IQ's. At ages 2 1/2 to 5 1/2, the reliability coefficients range from .83 (IQ's 140-149) to .91 (IQ's 60-69); at ages 6 to 13, .91 (IQ's 140-149) to .97 (IQ's 60-69); and at ages 14-18, .95 (IQ's 140-149) to .98 (IQ's 60-69). Since only the most reliable items of the 1937 scale were included in the 1960 revision, the 1960 scale is at least as reliable as the 1937 scale (Terman & Merrill, 1960).

Biserial correlations were done for the tests included in the L-M form. The mean correlation for the 1960 scale is .66, compared with a mean of .61 for all tests in both forms of the 1937 scale. At the preschool level (ages 2 1/2 to 5) the mean was .61, compared with the 1937 mean of .62. Verbal tests have a higher correlation (.65) with the total than the non-verbal tests (.58) (Terman & Merrill, 1960).

Correlations between retests are high when subjects are retested at fairly frequent intervals. In general, correlations decrease as interval time is lengthened and correlations increase as the child grows older if the interval between the two tests is held constant. Data from the Fels survey show that the correlation between tests given

at age three with retests at age 4 is .83; retest correlations with each successive year away from three decrease until the correlation at age 12 is .46 (Sontag et al., 1958). Test-retest correlations with later ages and age 5 or 6 are much higher than those with ages under five. For example, Bayley (1949) found that the correlations between age 10 and ages 2, 4, 6 and 8 were .42, .73, .74, and .82.

Remarks

Recent questioning of the cultural and socio-economic biases in test items has led to a reexamination of the validity of tests such as the Stanford-Binet. Significant questioning and pressures from minority groups resulted in omitting the Stanford-Binet from the 1971-72 battery. Some of the major areas of concern in considering the use of the Stanford-Binet and other intelligence tests are listed below:

1. Standardization on white samples. The Stanford-Binet was last standardized on an American white population. No standardization figures are available for a non-white population.

2. Socio-economic status. A number of studies (Charters, 1963; Willerman et al., 1970) have shown that children from lower socio-economic backgrounds score lower than those from higher socio-economic backgrounds. Specific items on the test may be foreign to the particular cultural setting of some children. There is no evidence that the Stanford-Binet is more biased toward lower socio-economic

children than other general intelligence tests.

3. Language. Both Anastasi (1958) and Freeman (1962) have specifically criticized the heavy verbal loadings on intelligence tests, which may present particular problems to lower class children. Verbal tasks on the Stanford-Binet are more frequent throughout the older age subtests.

4. Motivation. Zigler and Butterfield (1968) found that lack of motivation in culturally deprived children led to depressed Stanford-Binet IQ scores. After a preschool experience a reduction in the dehabilitating factors of motivation occurred and IQ scores increased.

5. Test administration. Testing younger children is especially difficult. The use of non-white testers with children of different ethnic backgrounds needs to be further investigated.

While the preceding general problems need to be explored further in future studies using all intelligence tests, actual biases specific to the Stanford-Binet have not been documented. Even though standardization with non-white populations and certain revisions in vocabulary and tasks seem crucial, the Stanford-Binet appears to be one of the best tests of general individual intellectual assessment.

References

Anastasi, A. Differential psychology. New York: Macmillan, 1958.

Bayley, N. Consistency and variability in the growth of intelligence from birth to eighteen years. Journal of Genetic Psychology, 1949, 75, 165-196.

Charters, W. W. Social class and intelligence tests. In W. Charters & N. Gage (Eds.), Readings in the social psychology of education. Boston: Allyn & Bacon, 1963.

Freeman, F. S. Theory and practice of psychological testing. (3rd. ed.) New York: Holt, Rinehart and Winston, 1962.

Terman, L. M., & Merrill, M. A. Stanford-Binet Intelligence Scale; Manual for the third revision-Form L-M. Boston: Houghton-Mifflin, 1960.

Willerman, L., Broman, S. H., & Fiedler, M. Infant development, preschool IQ, and social class. Child Development, 1970, 41, 69-77.

Zigler, E., & Butterfield, E. C. Motivational aspects of changes in IQ test. Child Development, 1968, 39, 1-14.

Wide Range Achievement Test

Purpose

The Wide Range Achievement Test (WRAT) is an achievement test designed to measure skills in the areas of reading, spelling and arithmetic. Most preschool programs for disadvantaged children emphasize the acquisition of these skills.

Description

The spring '72 form of the WRAT used in the HSPV Study is longer than the fall '71 form because most children have more cognitive skills after one year in a Head Start program than before. The fall '71 form has four subtests: copy marks, recognizing and naming letters, dot counting, and reading numbers. The spring '72 form has eight subtests: copy marks, name spelling, recognizing and naming letters, spelling, dot counting, reading numbers and arithmetic, written arithmetic, and word reading. Descriptions of the subtests in the three skill areas follow:

A. Spelling Skills:

1. Copying Marks. In a one minute time interval the child is to copy as many marks as possible.
2. Name Spelling. Part I asks the child to print his name in a one minute time interval on a line provided. Part II asks the child to name all the recognizable letters he has printed.

B. Reading Skills:

1. Recognizing and Naming Letters. Part I asks the child to recognize and match letters. The tester points to a series of letters in the row; the child picks out the matching letters from a different series. Part II asks the child to read aloud the letters in the second row.

2. Word Reading. The child is asked to read aloud a list of 14 words: cat, see, red, to, big, work, book, eat, was, him, how, then, open, letter.

C. Arithmetic Skills:

1. Dot Counting. The child is asked to count dots arranged in a row.

2. Reading Numbers. In a one minute time interval the child is asked to read aloud the numbers "3, 5, 6, 17, 41".

3. Arithmetic. The child is asked to respond to three arithmetic problems, such as "How many are three apples and four apples?".

4. Arithmetic (written computation). In a 30 second time interval the child is asked to read an arithmetic problem and write the answer in the box provided.

Development of Instrument

The WRAT was developed in 1940; revised in 1946 by J. Jastak and S. Bijou (Buros, 1965); and revised in 1965 by J. Jastak, S. Jastak and S. Bijou (Buros, 1972). The 1965 revised edition was prepared in two forms: Level I for ages 5 to 12 and Level II for age 12 and over. The WRAT used in the HSPV Study is a revised version of the Level I 1965 edition. A similar version was used during two years of the Follow Through evaluation (1970-72).

Norms

Norms for the five subtests given in Fall 1971 are available in Tables 1 - 40. These tables give the number of children, the mean score and the standard deviation for each of 15 three-month age intervals (from 36-38 months to 78-80 months) for the following groups in the HSPV sample: total (Table 1 - Copying Marks, Table 9 - Recognizing Letters, Table 17 - Naming Letters, Table 25 - Reading Numbers, and Table 33 - Dot Counting), males (Table 2 - Copying Marks, Table 10 - Recognizing Letters, Table 18 - Naming Letters, Table 26 - Reading Numbers, Table 34 - Dot Counting), females (Table 3 - Copying Marks, Table 11 - Recognizing Letters, Table 19 - Naming Letters, Table 27 - Reading Numbers, Table 35 - Dot Counting), children with previous preschool experience (Table 4 - Copying Marks, Table 12 - Recognizing Letters, Table 19 - Naming Letters,

Table 28, Reading Numbers, Table 36 - Dot Counting), children with no previous preschool (Table 5 - Copying Marks, Table 13 - Recognizing Letters, Table 21 - Naming Letters, Table 29 - Reading Numbers, Table 37 - Dot Counting), white children (Table 6 - Copying Marks, Table 14 - Recognizing Letters, Table 22 - Naming Letters, Table 30 - Reading Numbers, Table 38 - Dot Counting), black children (Table 7 - Copying Marks, Table 15 - Recognizing Letters, Table 23 - Naming Letters, Table 31 - Reading Numbers, Table 39 - Dot Counting) and Mexican-American children (Table 8 - Copying Marks, Table 16 - Recognizing Letters, Table 24 - Naming Letters, Table 32 - Reading Numbers, Table 40 - Dot Counting).

Mean scores for the total Fall 1971 sample on the fall subtests were 1.921 (S.D.=2.666) for Copying Marks, 6.554 (S.D. = 3.205) for Recognizing Letters, 1.195 (S.D. = 2.632) for Naming Letters, .613 (S.D. = 1.103) for Reading Numbers, and 6.708 (S.D. = 5.294) for Dot Counting. Scores for the Copying Marks, Recognizing Letters, and Dot Counting subtests definitely increased with age while scores of the Naming Letters and Reading Numbers subtests improved minimumly with age.

TABLE 1

DISTRIBUTION OF WRAT COPYING MARKS SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	6	0.333	0.745
39-41	4	0.500	0.500
42-44	22	0.273	0.617
45-47	77	0.636	1.643
48-50	262	0.492	0.984
51-53	478	0.705	1.204
54-56	481	1.158	1.793
57-59	464	1.672	2.192
60-62	379	2.517	2.784
63-65	252	2.968	2.766
66-68	255	3.443	3.333
69-71	205	5.317	3.631
72-74	87	4.345	3.835
75-77	5	4.200	3.816
78-80	3	1.333	0.943
TOTAL	2980	1.921	2.666

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 18.

TABLE 2

DISTRIBUTION OF WPAT COPYING MAPES SCORES FOR MALES[1] IN THE FALL 1971 HSPV SAMPLE

Age (Months)	N	Mean Score[2]	S.D.
36-38	1	-----	-----
39-41	1	-----	-----
42-44	8	0.125	0.331
45-47	46	0.413	0.946
48-50	132	0.492	1.048
51-53	255	0.612	1.149
54-56	243	0.881	1.539
57-59	239	1.247	1.772
60-62	206	2.335	2.732
63-65	115	2.374	2.386
66-68	137	2.591	2.814
69-71	94	3.649	3.426
72-74	46	3.891	3.789
75-77	3	7.000	2.160
78-80	2	2.000	-----
TOTAL	1528	1.577	2.372

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 18.

TABLE 3

DISTRIBUTION OF WRAT COPYING MARKS SCORES FOR FEMALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	5	0.400	0.800
39-41	3	0.667	0.471
42-44	14	0.357	0.718
45-47	31	0.968	2.279
48-50	130	0.492	0.914
51-53	223	0.812	1.257
54-56	238	1.441	1.980
57-59	225	2.124	2.486
60-62	173	2.734	2.828
63-65	137	3.467	2.957
66-68	118	4.432	3.604
69-71	111	4.883	3.702
72-74	41	4.854	3.823
75-77	2	-----	-----
78-80	1	-----	-----
TOTAL	1452	2.284	2.899

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 18.

TABLE 4

DISTRIBUTION OF WRAT COPYING MARKS SCORES FOR ALL CHILDREN WITH PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	---	-----	-----
42-44	2	1.000	1.000
45-47	8	1.625	1.867
48-50	34	0.971	1.543
51-53	64	0.703	1.056
54-56	62	0.871	1.301
57-59	79	2.392	3.128
60-62	119	2.840	2.750
63-65	98	2.949	3.167
66-68	92	3.554	3.595
69-71	95	4.526	3.963
72-74	37	5.189	4.190
75-77	3	7.000	2.160
78-80	2	1.000	1.000
TOTAL	695	2.784	3.297

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 18.

TABLE 5

DISTRIBUTION OF WRAT COPYING MARKS SCORES FOR ALL CHILDREN WITH NO PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	6	0.333	0.745
39-41	4	0.500	0.500
42-44	19	0.211	0.521
45-47	68	0.529	1.586
48-50	221	0.403	0.827
51-53	397	0.708	1.240
54-56	405	1.190	1.847
57-59	373	1.507	1.920
60-62	249	2.369	2.816
63-65	146	2.966	2.478
66-68	157	3.376	3.205
69-71	107	4.047	3.175
72-74	50	3.720	3.418
75-77	2	-----	-----
78-80	1	2.000	-----
TOTAL	2205	1.647	2.377

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 18.

TABLE 6

DISTRIBUTION OF WRAT COPYING MARKS SCORES FOR WHITE CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	3	0.667	0.471
42-44	8	0.125	0.331
45-47	32	0.500	0.968
48-50	93	0.548	1.122
51-53	188	0.798	1.419
54-56	212	1.311	1.842
57-59	187	1.706	2.309
60-62	163	2.650	2.851
63-65	94	2.989	2.988
66-68	99	3.162	3.335
69-71	79	5.025	3.486
72-74	51	4.824	4.264
75-77	3	7.000	2.160
78-80	2	1.000	1.000
TOTAL	1214	2.067	2.812

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 18.

TABLE 7

DISTRIBUTION OF WRAT COPYING MARKS SCORES FOR BLACK CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	6	0.333	0.745
39-41	1	-----	-----
42-44	14	0.357	0.718
45-47	45	0.733	1.982
48-50	140	0.421	0.854
51-53	210	0.571	0.950
54-56	205	1.010	1.778
57-59	193	1.446	1.966
60-62	137	1.825	2.018
63-65	106	2.396	2.398
66-68	102	3.167	3.116
69-71	103	3.485	3.444
72-74	34	3.559	2.932
75-77	2	-----	-----
78-80	1	2.000	-----
TOTAL	1299	1.550	2.309

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 18.

TABLE 8

DISTRIBUTION OF WRAT COPYING SCORES FOR MEXICAN-AMERICAN CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	---	-----	-----
42-44	---	-----	-----
45-47	---	-----	-----
48-50	23	0.783	1.140
51-53	73	0.836	1.250
54-56	53	1.094	1.640
57-59	70	2.214	2.461
60-62	69	3.174	3.189
63-65	50	4.100	2.744
66-68	51	4.451	3.339
69-71	19	5.421	3.345
72-74	---	-----	-----
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	408	2.564	2.914

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 18.

TABLE 9

DISTRIBUTION OF WRAT RECOGNIZING LETTERS SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	6	3.667	3.682
39-41	4	6.750	1.299
42-44	22	4.455	3.115
45-47	77	4.286	3.129
48-50	262	5.172	3.229
51-53	478	5.651	3.217
54-56	481	6.403	3.106
57-59	464	6.547	3.041
60-62	379	7.179	3.152
63-65	252	7.095	3.152
66-68	255	7.745	2.820
69-71	205	7.780	3.015
72-74	87	8.655	2.100
75-77	5	6.000	4.517
78-80	3	6.000	1.414
TOTAL	2980	6.554	3.205

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 10.

TABLE 10

DISTRIBUTION OF WRAT RECOGNIZING LETTERS SCORES FOR MALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	1	-----	-----
39-41	1	5.000	-----
42-44	8	5.125	2.522
45-47	46	3.957	2.881
48-50	132	5.364	3.222
51-53	255	5.443	3.096
54-56	243	6.144	3.145
57-59	239	6.301	3.113
60-62	206	7.175	2.903
63-65	115	6.800	3.149
66-68	137	7.416	2.843
69-71	94	7.351	3.178
72-74	46	8.674	2.001
75-77	3	9.6671	0.471
78-80	2	7.000	-----
TOTAL	1528	6.369	3.180

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 10.

TABLE 11

DISTRIBUTION OF WRAT RECOGNIZING LETTERS SCORES FOR FEMALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	5	4.400	3.611
39-41	3	7.333	0.943
42-44	14	4.071	3.348
45-47	31	4.774	3.405
48-50	130	4.977	3.224
51-53	223	5.888	3.334
54-56	238	6.668	3.042
57-59	225	6.809	2.940
60-62	173	7.185	3.041
63-65	137	7.343	3.133
66-68	118	8.127	2.745
69-71	111	8.144	2.818
72-74	41	8.634	2.206
75-77	2	0.500	0.500
78-80	1	4.000	-----
TOTAL	1452	6.749	3.220

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 10.

TABLE 12

DISTRIBUTION OF WRAT RECOGNIZING LETTERS SCORES FOR ALL CHILDREN WITH PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	---	-----	-----
42-44	2	8.500	0.500
45-47	8	6.125	3.219
48-50	34	5.706	3.650
51-53	64	6.375	2.809
54-56	62	6.581	3.124
57-59	79	7.228	2.846
60-62	119	7.445	2.866
63-65	98	7.449	2.935
66-68	92	8.185	2.231
69-71	95	8.053	2.766
72-74	37	8.784	2.120
75-77	3	9.667	0.471
78-80	2	5.500	1.500
TOTAL	695	7.404	2.909

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 10.

TABLE 13

DISTRIBUTION OF WRAT RECOGNIZING LETTERS SCORES FOR ALL CHILDREN WITH NO PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE

Age (Months)	N	Mean Score [2]	S.D.
36-38	6	3.667	3.682
39-41	4	6.750	1.299
42-44	19	3.895	2.972
45-47	68	3.985	2.983
48-50	221	5.100	3.178
51-53	397	5.547	3.247
54-56	405	6.410	3.078
57-59	373	6.378	3.067
60-62	249	7.052	2.993
63-65	146	7.137	3.049
66-68	157	7.592	2.956
69-71	107	7.645	3.100
72-74	50	8.560	2.080
75-77	2	0.500	0.500
78-80	1	7.000	-----
TOTAL	2205	6.323	3.224

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score=10.

TABLE 14

DISTRIBUTION OF WPAT RECOGNIZING LETTERS SCORES FOR WHITE CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	3	7.000	1.414
42-44	8	4.750	2.586
45-47	32	4.656	3.058
48-50	93	5.323	3.111
51-53	188	5.622	3.135
54-56	212	6.476	3.142
57-59	187	6.909	2.734
60-62	163	7.485	2.601
63-65	94	7.628	2.621
66-68	99	7.939	2.440
69-71	79	8.532	2.055
72-74	51	8.804	1.645
75-77	3	9.667	0.471
78-80	2	5.500	1.500
TOTAL	1214	6.846	2.985

[1] Includes all children with adequate age information not in Level 1 sites.

[2] maximum score=10.

TABLE 15

DISTRIBUTION OF WRAT RECOGNIZING LETTERS SCORES FOR BLACK CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	6	3.667	3.682
39-41	1	6.000	------
42-44	14	4.286	3.369
45-47	45	4.022	3.152
48-50	140	5.100	3.332
51-53	210	5.505	3.256
54-56	205	6.371	3.105
57-59	193	6.104	3.196
60-62	137	6.869	3.073
63-65	106	7.264	2.779
66-68	102	7.637	2.920
69-71	103	7.757	2.951
72-74	34	8.588	2.415
75-77	2	0.500	0.500
78-80	1	7.000	-----
TOTAL	1299	6.322	3.278

[1]Includes all children with adequate age information not in Level I sites.

[2]maximum score=10.

TABLE 16

DISTRIBUTION OF WRAT RECOGNIZING LETTERS SCORES FOR MEXICAN-AMERICAN CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	---	-----	-----
42-44	---	-----	-----
45-47	---	-----	-----
48-50	23	5.826	2.792
51-53	73	6.068	3.215
54-56	53	6.321	2.800
57-59	70	6.786	3.295
60-62	69	7.014	3.317
63-65	50	5.620	4.204
66-68	51	7.451	3.268
69-71	19	4.526	4.453
72-74	---	-----	-----
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	408	6.417	3.458

[1] Includes all children with adequate age information not in Level I sites.

[2] maximum score=10.

TABLE 17

DISTRIBUTION OF WRAT NAMING LETTERS SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	6	-----	-----
39-41	4	-----	-----
42-44	22	0.318	0.555
45-47	77	1.013	2.535
48-50	262	0.649	1.736
51-53	478	0.722	2.017
54-56	481	1.073	2.540
57-59	464	1.136	2.550
60-62	379	1.417	2.850
63-65	252	1.115	2.225
66-68	255	1.675	3.001
69-71	205	2.137	3.620
72-74	87	2.655	3.856
75-77	5	0.800	0.748
78-80	3	-----	-----
TOTAL	2980	1.195	2.632

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score=13.

TABLE 18

DISTRIBUTION OF WPAT NAMING LETTERS SCORES FOR MALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	1	-----	-----
39-41	1	-----	-----
42-44	8	0.125	0.331
45-47	46	0.783	2.074
48-50	132	0.644	1.508
51-53	255	0.592	1.887
54-56	243	1.037	2.483
57-59	239	0.858	2.242
60-62	206	1.422	3.074
63-65	115	1.226	2.296
66-68	137	1.343	2.822
69-71	94	2.064	3.784
72-74	46	2.283	3.405
75-77	3	1.000	0.816
78-80	2	-----	-----
TOTAL	1528	1.080	2.550

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score=13.

TABLE 19

DISTRIBUTION OF WRAT NAMING LETTERS SCORES FOR FEMALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	5	-----	-----
39-41	3	-----	-----
42-44	14	0.429	0.623
45-47	31	1.355	3.064
48-50	130	0.654	1.940
51-53	223	0.870	2.146
54-56	238	1.109	2.595
57-59	225	1.431	2.810
60-62	173	1.410	2.558
63-65	137	1.022	2.160
66-68	118	2.059	3.152
69-71	111	2.198	3.474
72-74	41	3.073	4.268
75-77	2	0.500	0.500
78-80	1	-----	-----
TOTAL	1452	1.316	2.709

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score=13.

TABLE 20

DISTRIBUTION OF WRAT NAMING LETTERS SCORES FOR ALL CHILDREN WITH PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	---	-----	-----
42-44	2	0.500	0.500
45-47	8	2.250	3.455
48-50	34	0.794	2.011
51-53	64	1.266	2.763
54-56	62	1.387	2.672
57-59	79	1.468	2.920
60-62	119	2.008	3.506
63-65	98	0.980	1.985
66-68	92	1.891	2.928
69-71	95	2.379	3.787
72-74	37	3.622	4.277
75-77	3	1.000	0.816
78-80	2	-----	-----
TOTAL	695	1.728	3.134

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score=13.

TABLE 21

DISTRIBUTION OF WRAT NAMING LETTERS SCORES FOR ALL CHILDREN WITH NO PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	6	-----	-----
39-41	4	-----	-----
42-44	19	0.211	0.408
45-47	68	0.882	2.380
48-50	221	0.624	1.699
51-53	397	0.597	1.770
54-56	405	1.049	2.554
57-59	373	1.013	2.335
60-62	249	1.129	2.373
63-65	146	1.233	2.407
66-68	157	1.535	2.964
69-71	107	1.972	3.492
72-74	50	1.940	3.337
75-77	2	0.500	0.500
78-80	1	-----	-----
TOTAL	2205	1.022	2.405

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 13.

TABLE 22

DISTRIBUTION OF WRAT NAMING LETTERS SCORES FOR WHITE CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	3	-----	-----
42-44	8	0.500	0.707
45-47	32	1.688	3.025
48-50	93	0.634	1.664
51-53	188	1.043	2.667
54-56	212	1.476	3.136
57-59	187	1.390	2.853
60-62	163	1.503	2.872
63-65	94	1.319	2.586
66-68	99	2.131	3.541
69-71	79	3.367	4.401
72-74	51	2.922	3.814
75-77	3	1.000	0.816
78-80	2	-----	-----
TOTAL	1214	1.552	3.083

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 13.

TABLE 23

DISTRIBUTION OF WRAT NAMING LETTERS SCORES FOR BLACKS CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	6	-----	-----
39-41	1	-----	-----
42-44	14	0.214	0.410
45-47	45	0.533	1.984
48-50	140	0.714	1.910
51-53	210	0.505	1.378
54-56	205	0.712	1.856
57-59	193	1.109	2.418
60-62	137	1.438	2.909
63-65	106	0.858	1.501
66-68	102	1.382	2.594
69-71	103	1.534	2.949
72-74	34	2.265	3.950
75-77	2	0.500	0.500
78-80	1	-----	-----
TOTAL	1299	0.968	2.26

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 13.

TABLE 24

DISTRIBUTION OF WRAT NAMING LETTERS SCORES FOR MEXICAN-AMERICAN CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score	S.D.
36-38	---	-----	-----
39-41	---	-----	-----
42-44	---	-----	-----
45-47	---	-----	-----
48-50	23	0.478	0.878
51-53	73	0.534	1.536
54-56	53	0.566	0.981
57-59	70	0.629	2.126
60-62	69	0.826	1.918
63-65	50	1.160	2.533
66-68	51	1.431	2.553
69-71	19	0.632	1.563
72-74	---	-----	-----
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	408	0.794	1.950

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 13.

TABLE 25

DISTRIBUTION OF WRAT READING NUMBERS SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	6	------	------
39-41	4	0.250	0.433
42-44	22	0.136	0.457
45-47	77	0.325	0.829
48-50	262	0.271	0.670
51-53	478	0.299	0.724
54-56	481	0.493	0.993
57-59	464	0.433	0.967
60-62	379	0.734	1.187
63-65	252	0.742	1.182
66-68	255	0.973	1.335
69-71	205	1.312	1.335
72-74	87	1.563	1. 51
75-77	5	0.800	1.166
78-80	3	------	------
TOTAL	2980	0.613	1.103

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 5.

TABLE 26

DISTRIBUTION OF WRAT READING NUMBERS SCORES FOR MALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	1	-----	-----
39-41	1	-----	-----
42-44	8	0.125	0.331
45-47	46	0.348	0.914
48-50	132	0.265	0.638
51-53	255	0.286	0.681
54-56	243	0.539	1.047
57-59	239	0.377	0.834
60-62	206	0.748	1.224
63-65	115	0.722	1.184
66-68	137	0.774	1.256
69-71	94	1.309	1.414
72-74	46	1.543	1.611
75-77	3	1.333	1.247
78-80	2	-----	-----
TOTAL	1528	0.580	1.085

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 5.

TABLE 27

DISTRIBUTION OF WRAT READING NUMBERS SCORES FOR FEMALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	5	-----	-----
39-41	3	0.333	0.471
42-44	14	0.143	0.515
45-47	31	0.290	0.681
48-50	130	0.277	0.702
51-53	223	0.314	0.769
54-56	238	0.445	0.932
57-59	225	0.596	1.079
60-62	173	0.717	1.141
63-65	137	0.759	1.181
66-68	118	1.203	1.387
69-71	111	1.315	1.401
72-74	41	1.585	1.481
75-77	2	-----	-----
78-80	1	-----	-----
TOTAL	1452	0.647	1.121

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 5.

TABLE 28

DISTRIBUTION OF WRAT READING NUMBERS SCORES FOR ALL CHILDREN WITH PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	---	-----	-----
42-44	2	1.500	0.500
45-47	8	1.125	1.053
48-50	34	0.324	0.629
51-53	64	0.438	0.916
54-56	62	0.661	1.062
57-59	79	0.557	1.076
60-62	119	0.924	1.291
63-65	98	0.776	1.093
66-68	92	1.109	1.363
69-71	95	1.295	1.368
72-74	37	2.081	1.583
75-77	3	1.333	1.247
78-80	2	-----	-----
TOTAL	695	0.904	1.265

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 5.

TABLE 29

DISTRIBUTION OF WRAT READING NUMBERS SCORES FOR ALL CHILDREN WITH NO PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score [2]	S.D.
36-38	6	------	------
39-41	4	0.250	0.433
42-44	19	-----	-----
45-47	68	0.235	0.750
48-50	221	0.267	0.684
51-53	397	0.290	0.698
54-56	405	0.472	0.987
57-59	373	0.456	0.927
60-62	249	0.659	1.134
63-65	146	0.719	1.232
66-68	157	0.904	1.305
69-71	107	1.355	1.449
72-74	50	1.180	1.410
75-77	2	-----	-----
78-80	1	-----	-----
TOTAL	2205	0.529	1.036

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 5.

TABLE 30

DISTRIBUTION OF WPAT READING NUMBERS SCORES FOR WHITE CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	3	0.333	0.471
42-44	8	-----	-----
45-47	32	0.375	0.927
48-50	93	0.312	0.816
51-53	188	0.330	0.770
54-56	212	0.571	1.103
57-59	187	0.572	1.028
60-62	163	0.920	1.306
63-65	94	0.798	1.190
66-68	99	1.202	1.400
69-71	79	1.734	1.482
72-74	51	1.686	1.627
75-77	3	1.333	1.247
78-80	2	-----	-----
TOTAL	1214	0.744	1.217

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 5.

TABLE 31

DISTRIBUTION OF WRAT READING NUMBERS SCORES FOR BLACK CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	6	-----	-----
39-41	1	-----	-----
42-44	14	0.214	0.558
45-47	45	0.289	0.749
48-50	140	0.271	0.596
51-53	210	0.276	0.669
54-56	205	0.415	0.860
57-59	193	0.466	0.982
60-62	137	0.577	1.065
63-65	106	0.509	0.934
66-68	102	0.725	1.181
69-71	103	1.039	1.277
72-74	34	1.382	1.415
75-77	2	-----	-----
78-80	1	-----	-----
TOTAL	1299	0.499	0.971

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 5.

TABLE 32

DISTRIBUTION OF WRAT READING NUMBERS SCORES FOR MEXICAN-AMERICAN CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	---	-----	-----
42-44	---	-----	-----
45-47	---	-----	-----
48-50	23	0.174	0.480
51-53	73	0.315	0.774
54-56	53	0.453	0.943
57-59	70	0.300	0.744
60-62	69	0.580	0.999
63-65	50	1.040	1.442
66-68	51	1.000	1.372
69-71	19	1.105	1.372
72-74	---	-----	-----
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	408	0.578	1.084

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 5.

TABLE 33

DISTRIBUTION OF WRAT DOT COUNTING SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
39-41	4	2.000	.707
42-44	22	2.455	3.299
45-47	77	3.714	4.360
48-50	262	4.050	4.336
51-53	478	4.360	4.403
54-56	481	5.699	5.048
57-59	464	6.349	5.019
60-62	379	8.024	5.112
63-65	252	8.794	5.009
66-68	255	9.596	4.883
69-71	205	10.444	4.832
72-74	87	10.138	5.052
75-77	5	5.600	4.499
78-80	3	5.333	6.182
TOTAL	2974	6.708	5.294

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 15.

TABLE 34

DISTRIBUTION OF WRAT DOT COUNTING SCORES FOR MALE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
39-41	1	1.000	-
42-44	8	.875	1.053
45-47	46	3.913	4.496
48-50	132	4.182	4.627
51-53	255	3.847	4.142
54-56	243	5.078	4.878
57-59	239	5.527	4.669
60-62	206	7.850	5.169
63-65	115	8.148	5.166
66-68	137	8.642	4.927
69-71	94	9.638	4.935
72-74	46	9.109	5.301
75-77	3	9.000	2.160
78-80	2	8.000	6.000
TOTAL	1527	6.144	5.175

[1]Includes all children with adequate age information not in Level I sites.

[2]Maximum score = 15.

TABLE 35

DISTRIBUTION OF WRAT DOT COUNTING SCORES

FOR FEMALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score [2]	S.D.
39-41	3	2.333	.471
42-44	14	3.357	3.772
45-47	31	3.419	4.133
48-50	130	3.915	4.015
51-53	223	4.946	4.614
54-56	238	6.332	5.139
57-59	225	7.222	5.227
60-62	173	8.231	5.034
63-65	137	9.336	4.808
66-68	118	10.703	4.589
69-71	111	11.126	4.635
72-74	41	11.293	4.484
75-77	2	.500	.500
78-80	1	-	-
TOTAL	1497	7.304	5.353

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 15.

TABLE 36

DISTRIBUTION OF WRAT DOT COUNTING SCORES FOR ALL CHILDREN WITH PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
39-41	---	-----	-----
42-44	2	5.500	4.500
45-47	8	5.875	5.555
48-50	34	5.324	5.132
51-53	64	5.375	4.827
54-56	62	5.758	5.148
57-59	79	8.038	5.232
60-62	119	9.126	4.965
63-65	98	9.204	4.863
66-68	92	10.717	4.507
69-71	95	10.968	4.644
72-74	37	11.865	4.134
75-77	3	9.000	2.160
78-80	2	1.000	1.000
TOTAL	695	8.718	5.290

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 15.

TABLE 37

DISTRIBUTION OF WRAT DOT COUNTING SCORES FOR ALL CHILDREN WITH NO PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
39-41	4	2.000	.707
42-44	19	2.158	3.065
45-47	68	3.368	4.080
48-50	221	3.891	4.201
51-53	397	4.191	4.297
54-56	405	5.716	5.077
57-59	373	5.954	4.896
60-62	249	7.462	5.120
63-65	146	8.342	5.108
66-68	157	9.064	4.919
69-71	107	10.093	4.860
72-74	50	8.860	5.284
75-77	2	.500	.500
78-80	1	14.000	-
TOTAL	2199	6.082	5.146

[1] Includes all children with adequate age information not in Level I sites.
[2] Maximum score = 15.

TABLE 38

DISTRIBUTION OF WRAT DOT COUNTING SCORES FOR WHITE CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
39-41	3	2.000	.816
42-44	8	2.250	3.455
45-47	32	4.156	4.658
48-50	93	4.118	4.093
51-53	188	4.340	4.508
54-56	212	5.297	5.001
57-59	187	5.861	4.832
60-62	163	7.908	5.178
63-65	94	8.074	5.068
66-68	99	9.293	5.109
69-71	79	10.722	4.698
72-74	51	10.118	4.910
75-77	3	9.000	2.160
78-80	2	1.000	1.000
TOTAL	1214	6.536	5.280

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 15.

TABLE 39

DISTRIBUTION OF WRAT COPYING MARKS SCORES FOR BLACK CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
39-41	1	2.000	-
42-44	14	2.571	3.201
45-47	45	3.400	4.106
48-50	140	4.186	4.607
51-53	210	4.414	4.416
54-56	205	6.098	5.194
57-59	193	6.839	5.222
60-62	137	8.182	4.863
63-65	106	8.991	4.759
66-68	102	10.088	4.655
69-71	103	10.243	4.900
72-74	34	10.294	5.096
75-77	2	.500	.500
78-80	1	14.000	-
TOTAL	1293	6.804	5.324

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 15.

TABLE 40

DISTRIBUTION OF WRAT DOT COUNTING SCORES FOR MEXICAN-AMERICAN CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
39-41	-	----	----
42-44	-	----	----
45-47	-	----	----
48-50	23	3.435	3.411
51-53	73	4.507	4.188
54-56	53	5.377	4.594
57-59	70	6.429	4.795
60-62	69	7.899	5.344
63-65	50	9.560	5.258
66-68	51	9.039	4.867
69-71	19	10.316	5.242
72-74	-	----	----
75-77	-	----	----
78-80	-	----	----
TOTAL	408	6.919	5.208

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 15.

Score Characteristics

Frequency distributions of the total scores for the five Fall 1971 WRAT subtests are presented in Tables 41 - 45 for the total HSPV sample. The distribution of scores of the Copying Marks subtest has a definite floor effect (See Table 41). Forty-four percent of the children (n=3033) copied no marks correctly and 16% copied only one correctly. The distribution of scores for Recognizing Letters, a matching test, is rather flat across all scores from 0 to 10 (see Table 42). The largest number of children (22%) scored the highest score. In the spring, this subtest had a definite ceiling effect. The distribution of scores for Naming Letters has a definite floor effect in the fall (see Table 43). Sixty-four percent of the children (n=3033) got no letters correct, while 15.6% named one correctly. The Reading Numbers subtest distribution of scores is also very positively skewed (see Table 44). Seventy-one percent of the children (n=3033) received scores of zero, while 11.4% received scores of one. It should be pointed out that the Reading Numbers subtest scores do not necessarily form a uniform scale since the first three numbers are single digit numbers and considerably easier to read than the last two numbers which are two-digit numbers. The Dot Counting scores from a bimodal distribution with each of the end scores being the most frequently obtained (12%) (See Table 45). The scores of this subtest do not represent a true continuous scale since the subtest consists of only one item scored from 0 - 15: the total score is not the result of scores on 15 separate items.

TABLE 41

DISTRIBUTION OF WRAT COPYING MARKS

SCORES FOR FALL 1971 HSPV TOTAL SAMPLE[1]

Score	# of Children	x = nearest 100 children
0	1338	xxxxxxxxxxxx
1	482	xxxxx
2	313	xxx
3	269	xxx
4	233	xx
5	125	x
6	76	x
7	56	x
8	37	
9	27	
10	25	
11	14	
12	10	
13	10	
14	8	
15	4	
16	1	
17	3	
18	2	

Total N =3033

[1] Includes PV & non-PV children

TABLE 42

DISTRIBUTION OF WRAT RECOGNIZING LETTERS SCORES FOR FALL 1971 HSPV TOTAL SAMPLE [1]

x = nearest 10 children

Score	# of Children	
0	240	xxxxxxxxxxxxxxxxxxxxxxxx
1	134	xxxxxxxxxxxxx
2	135	xxxxxxxxxxxxxx
3	113	xxxxxxxxxxx
4	161	xxxxxxxxxxxxxxxx
5	222	xxxxxxxxxxxxxxxxxxxxxx
6	244	xxxxxxxxxxxxxxxxxxxxxxxx
7	294	xxxxxxxxxxxxxxxxxxxxxxxxxxxxx
8	375	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
9	341	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
10	674	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Total N =3033

[1] Includes PV and non-PV children.

TABLE 43

DISTRIBUTION OF WRAT NAMING LETTERS SCORES

FOR FALL 1971 HSPV TOTAL SAMPLE[1]

Score	# of Children	x = nearest 100 children
0	1955	xxxxxxxxxxxxxxxxxxx
1	475	xxxxx
2	213	xx
3	106	x
4	47	
5	29	
6	27	
7	29	
8	23	
9	16	
10	17	
11	25	
12	26	
13	47	

Total N =3033

[1] Includes PV and non-PV children.

TABLE 44

DISTRIBUTION OF WRAT READING NUMBERS

SCORES FOR FALL 1971 HSPV TOTAL SAMPLE [1]

Score	# of Children	x = nearest 100 children
0	2154	xxxxxxxxxxxxxxxxxxxxx
1	345	xxx
2	181	xx
3	284	xxx
4	62	x
5	7	

Total N =3033

[1] Includes PV and non-PV children.

TABLE 45

DISTRIBUTION OF WRAT DOT COUNTING SCORES

FOR FALL 1971 HSPV TOTAL SAMPLE[1]

Score	# of Children	x = nearest 100 children
0	423	xxxx
1	375	xxxx
2	244	xx
3	215	xx
4	199	xx
5	159	xx
6	171	xx
7	115	x
8	78	x
9	103	x
10	168	xx
11	193	xx
12	112	x
13	107	x
14	139	x
15	403	xxxx

Total N = 3204

[1] Includes PV & non-PV children.

Reliability

KR-20 reliability coefficients for four of the five Fall subtests are reported in Tables 46 - 49 for the total sample (n=3205) and nine subsamples. They were not computed for the Dot Counting subtest since this is essentially a one item test. The KR-20 for the Copying Marks subtest was .794 for the total sample (see Table 46). The estimates for 91 subsamples with a size greater than 20 ranged from .462 for young black males with previous preschool experience (n = 47) to .848 for white females with previous preschool experience (n = 135) and for old white males with previous preschool experience (n = 94). Twenty-two percent of the KR-20's were greater than .80; one-half were in the .70's. The KR-20 for the Recognizing Letters subtest was .794 for the total sample (see Table 47). The estimates for 91 subsamples with a size greater than 20 ranged from .696 for young female Mexican-American children with no previous preschool experience (n = 60) to .854 for old Mexican-American females (n = 130). Almost all of the estimates were from .70 to .85. The KR-20 for the Naming Letters subtest was .848 for the total sample (see Table 48). The estimates for 91 subsamples with a size greater than 20 ranged from .376 for young female Mexican-American children with no previous preschool experience (n = 60) to .902 for young white males with previous preschool experience (n = 34). Eighty-eight percent of the KR-20's were greater than .80. The KR-20 for the Reading Numbers subtest was .593 for the total sample (see Table 49). The estimates for 91 subsamples with a size greater than 20 ranged from .456 for young black females with no previous preschool experience (n = 296)

to .699 for Mexican-American males with previous preschool experience (n = 33). About one-third of the KR-20's were greater than .60.

KR-20's for four of the subtests given only in the spring battery (Spelling, Oral Arithmetic, Written Arithmetic, Word Reading) are presented in Table 50 for the total sample, males and females. Since these KR-20's are computed on the spring sample after the HSPV treatment was introduced, they can not be compared to the KR-20's reported for the fall subtests.

TABLE 46

KR-20 RELIABILITIES FOR FALL 1971 COPYING MARKS

SUBTEST-WRAT SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	3205	1.899	2.647	.794
Black	1392	1.527	2.298	.774
White	1301	2.045	2.793	.802
Mexican-American	446	2.534	2.883	.790
Male	1596	1.571	2.374	.780
Female	1525	2.257	2.896	.804
Young[3]	1416	0.817	1.469	.678
Old	1684	2.814	3.059	.797
Previous Preschool	751	2.746	3.272	.821
No Previous Preschool	2371	1.624	2.357	.773

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 18.

[3] Young is less than 57 months; old is greater than 56 months.

TABLE 47

KR-20 RELIABILITIES FOR FALL 1971 RECOGNIZING LETTERS SUBTEST-WRAT SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	3205	6.467	3.244	.794
Black	1392	6.230	3.306	.796
White	1301	6.729	3.076	.781
Mexican-American	446	6.453	3.403	.811
Male	1596	6.330	3.196	.785
Female	1525	6.655	3.268	.801
Young[3]	1416	5.638	3.268	.784
Old	1684	7.200	3.034	.791
Previous Preschool	751	7.388	2.892	.780
No Previous Preschool	2371	6.205	3.276	.792

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 10.

[3] Young is less than 57 months; old is greater than 56 months.

TABLE 48

KR-20 RELIABILITIES FOR FALL 1971 NAMING LETTERS

SUBTEST-WRAT SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	3205	1.204	2.642	.848
Black	1392	.978	2.290	.832
White	1301	1.548	3.063	.860
Mexican-American	446	0.872	2.116	.825
Male	1596	1.104	2.576	.852
Female	1525	1.323	2.723	.845
Young[3]	1416	0.854	2.217	.841
Old	1684	1.506	2.929	.850
Previous Preschool	751	1.758	3.158	.853
No Previous Preschool	2371	1.024	2.411	.842

[1] Includes all children with adequate age information not in Level I sites.

[2] Maximum score = 13.

[3] Young is less than 57 months; old is greater than 56 months.

TABLE 49

KR-20 RELIABILITIES FOR FALL 1971 READING NUMBERS

SUBTEST-WRAT SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	3205	0.604	1.098	.593
Black	1392	0.488	0.965	.558
White	1301	0.736	1.213	.612
Mexican-American	446	0.574	1.081	.596
Male	1596	0.578	1.085	.596
Female	1525	0.640	1.118	.591
Young[3]	1416	0.360	0.828	.537
Old	1684	0.818	1.249	.607
Previous Preschool	751	0.908	1.271	.595
No Previous Preschool	2371	0.515	1.025	.588

[1] Includes all children with adequate information not in Level I sites.

[2] Maximum score = 5.

[3] Young is less than 57 months; old is greater than 56 months.

TABLE 50

KR-20 RELIABILITIES FOR SPRING 1972 WRAT SUBTESTS FOR TOTAL HSPV SAMPLE, MALES AND FEMALES[1]

sample	n	mean score	S.D.	KR-20
	A. Spelling (Maximum Score = 8)			
Total	2792	.116	.626	.712
Male	1411	.099	.585	.719
Female	1381	.133	.664	.707
	B. Oral Arithmetic (Maximum Score = 7)			
Total	2792	2.320	1.660	.550
Male	1411	2.327	1.684	.557
Female	1381	2.312	1.635	.544
	C. Written Arithmetic (Maximum Score = 4)			
Total	2792	.078	.388	.499
Male	1411	.060	.355	.534
Female	1381	.096	.419	.474
	D. Word Reading (Maximum Score = 15)			
Total	2792	.078	.388	.499
Male	1411	.060	.355	.534
Female	1381	.096	.419	.474

[1] These subtests were only given in Spring 1972.

Validity

Almost all of the studies using the WRAT deal with populations which are not comparable with the HSPV sample. One exception is a study by Washington & Teska (1970) in which they individually administered the WRAT, ITPA, California Achievement Tests (CAT) Primary Forms, and the Stanford-Binet to 96 disadvantaged children (ages 5-7 to 7-5, mean age = 6-9). Pearsonian correlations of the WRAT with the CAT are listed below:

	WRAT		
CAT	Reading	Spelling	Arithmetic
Reading	.86	.82	.72
Arithmetic	.87	.82	.84
Language	.80	.84	.69
Total	.89	.87	.79

These high correlations are evidence of good concurrent validity for the WRAT. Correlations with the Stanford-Binet and ITPA Verbal Expression Subtest are listed below:

	Stanford-Binet		ITPA	
WRAT	MA	IQ	Ver. Exp.	Total
Reading	.74	.46	.38	.72
Spelling	.77	.57	.37	.71
Arithmetic	.70	.41	.31	.68

The intercorrelations of the five WRAT subtests given in Fall 1971 and the correlations of these subtests with other tests in the Fall 1971 battery are given in Table 51. The intercorrelations between these subtests ranged from .302 for Naming Letters and Recognizing letters to .600 for Reading Numbers and Naming Letters. Some of the highest correlations between a WRAT subtest and other tests were .620 (Dot Counting and ETS Enumeration Counting), .589 (Dot Counting and the 32-item PSI), .542 (Dot Counting and ETS Enumeration Total), .537 (Recognizing Letters and the PPVT), .551 (Copying Marks and the 32-item PSI), .508 (Copying Marks and the ETS Enumeration Total), .508 (Reading Numbers and the 32-item PSI), .504 (Copying Marks and the ETS Enumeration Counting Subtest), and .500 (Reading Numbers and the ETS Enumeration Counting Subtest). Correlations between the various WRAT subtests and the ITPA Verbal Expression Subtest were around .30, and thus similar to the Washington and Teska (1970) findings.

All of the subtests correlated in the .40 - .50 range with the other achievement measure in the battery (32-item PSI): .55-Copying Marks, .48-Recognizing Letters, 41-Naming Letters, and .51-Reading Numbers. If the correlations are corrected for unreliability[1], the estimated correlation between the true score components of the 32-item PSI and these WRAT subtests are higher: .7-Copying Marks, .6-Recognizing Letters, .5-Naming Letters, and .7-Reading Numbers.

---

[1] Using $\frac{r_{1.2}}{\sqrt{t_1 \cdot t_2}}$ where $r_{1.2}$ is the correlation between tests and t_1, t_2 are reliability estimates for the tests.

TABLE 51 INTERCORRELATIONS OF FALL 1971 SCORES FROM THE PPVT, WRAT SUBTESTS, 32-ITEM PSI, ITPA VERBAL EXPRESSION SUBTEST, ETS ENUMERATION SUBTESTS, BROWN, MI-TRUCK SUBTEST, AND EIGHT-BLOCK SORT SUCCESS SCORES[1]

	PPVT	WRAT-COPY MARKS	WRAT-RECOG. LETTERS	WRAT-NAME LETTERS	WRAT-READ #	WRAT-DOT COUNT	PSI 32-ITEM	ITPA-VERBAL EXPRESS.	ETS. ENUM. TOTAL	ETS. ENUM. COUNT.	ETS ENUM. TOUCH.	ETS ENUM. SAME # MATCH.	BROWN UNADJ.	BROWN ADJ.	MI-TRUCK	EIGHT-BLOCK PLACE.	EIGHT-BLOCK REASON
WRAT-COPY MARKS	.413 (2881)																
WRAT-RECOG. LETTERS	[illegible] (2881)	.375 ([illegible])															
WRAT-NAME LETTERS	[illegible] (2881)	[illegible]	[illegible]														
WRAT-READ NUMBERS	[illegible] (2881)	[illegible]	.325 ([illegible])	.600 ([illegible])													
WRAT-DOT COUNTING	.453 (2881)	[illegible]	[illegible]	.324 ([illegible])	.451 ([illegible])												
PSI (32-item)	.665 ([illegible])	.551 ([illegible])	.481 ([illegible])	.414 ([illegible])	.508 ([illegible])	[illegible] ([illegible])											
ITPA-VERBAL EXPRESSION	.467 ([illegible])	.353 ([illegible])	[illegible]	.270 ([illegible])	.341 ([illegible])	.368 ([illegible])	.506 (1138)										
ETS ENUMERATION TOTAL[2]	.475 ([illegible])	[illegible]	.327 ([illegible])	[illegible]	[illegible]	.342 (1097)	.584 (1073)	.459 (1115)									
ETS ENUMERATION COUNTING	.432 ([illegible])	[illegible]	.422 ([illegible])	.359 ([illegible])	.500 ([illegible])	[illegible] ([illegible])	.625 (1073)	.354 (1115)	.781 (1135)								
ETS ENUMERATION TOUCHING	[illegible]	.359 ([illegible])	.207 ([illegible])	[illegible]	[illegible]	[illegible] ([illegible])	.382 (1073)	.308 (1115)	.721 (1135)	.390 (1135)							
ETS ENUMERATION SAME # MATCHING	.237 ([illegible])	.225 ([illegible])	[illegible]	[illegible]	[illegible]	[illegible]	.232 (1073)	.298 (1115)	.664 (1135)	.257 (1135)	.202 (135)						
BROWN-UNADJUSTED	.322 ([illegible])	.362 (2753)	[illegible] (2753)	[illegible] (2753)	[illegible] (2753)	[illegible]	.323 (2669)	.261 (1145)	.226 (1073)	.271 (1073)	.160 (1073)	.054 (1073)					
BROWN-ADJUSTED	.239 (2669)	.127 (2753)	.156 (2753)	.100 (2753)	.124 (2753)	.124 (2753)	.239 (2669)	.215 (1145)	.159 (1073)	.172 (1073)	.134 (1073)	.034 (1073)	.637 (2879)				
MI-TRUCK[3]	.174 (607)	.061 (625)	.048 (625)	.085 (625)	.121 (625)	.006 (625)	.164 (618)	.032 (637)	.136 (597)	.175 (597)	.047 (597)	.107 (597)	.118 (610)	.109 (610)			
EIGHT-BLOCK PLACEMENT	.301 (1119)	.222 ([illegible])	.271 ([illegible])	.145 ([illegible])	[illegible] ([illegible])	.304 (1148)	.305 (1090)	.303 (1096)	.322 (1032)	.413 (1032)	.200 (1032)	.180 (1032)	.212 (1113)	.183 (1113)	.005 (573)		
EIGHT-BLOCK REASON	.445 (1119)	.364 ([illegible])	.353 (1148)	.266 (1148)	.372 ([illegible])	.390 (1148)	.423 (1090)	.416 ([illegible])	.405 (1032)	.402 (1032)	.258 (1032)	.211 (1032)	.178 (1113)	.168 (1113)	.063 (573)	.520 (1211)	
EIGHT-BLOCK SUCCESS TOTAL	.433 (1119)	.346 (1148)	.351 (1148)	.257 (1148)	.344 (1148)	.404 (1148)	.440 (1090)	.422 (1096)	.422 (1032)	.416 (1032)	.266 (1032)	.226 (1032)	.220 (1113)	.200 (1113)	.046 (573)	.839 (1211)	.901 (1211)

[1] Sample size for each correlation is included in parenthesis. Children in sample are those with adequate information not in Level I sites.

[2] ETS ENUMERATION Score= sum of counting, touching and same number matching subtest scores.

[3] MI scores are log transformations of slow times.

Remarks

Many of the WRAT subtests appear to be too difficult for young children. Even though some of the internal consistency reliabilities are adequate for subtests of this length, the skewed distribution of most subtest scores limits the usefulness of this data. Technical information is still needed on the subtests used only in the Spring 1972; this was not calculated for the HSPV data since the results would be confounded by treatment effects and could not be compared to similar estimates for the Fall subtests. Because of these problems, it is recommended that the WRAT subtests be used only as a set of criterion-reference measures.

Several other questions need to be explored also in future analyses using the WRAT:

1. Why are the subtests timed? Is speed really important, especially for younger children? It is clear that the time constraints would be a disadvantage to the youngest children in the HSPV sample.

2. Are the instructions truly standardized? Since there is such a wide variety of subtests, it is not certain whether all testers give every item in a standard way. In addition, there are the recurrent problems in test administration with younger children of prompting, verbal reinforcement, and gestural cueing.

3. Is the Copying Marks subtest more a measure of

motor coordination than achievement especially when used with younger children?

4. Should more attention be placed on individual children's response styles in addition to correct responses?

5. Is there a meaningful way to aggregate scores across subtests to produce one composite achievement score?

References

Buros, O. K. (Ed.) The sixth mental measurements yearbook. Highland Park, N.J.: Gryphon, 1965.

Buros, O. K. (Ed.) The seventh mental measurements yearbook. Highland Park, N.J.: Gryphon, 1972. 2 vols.

Washington, E. D., & Teska, J. A. Correlations between the Wide Range Achievement Test, the California Achievement Tests, the Stanford-Binet, and the Illinois Test of Psycholinguistic Abilities. Psychological Reports, 1970, 26, 291-4.

PART III: APPENDICES RELATING TO THE PROCEDURES USED IN EXAMINING THE QUALITY OF THE DATA

APPENDIX A

Test-Retest / Inter-Tester Reliability Study

Introduction

There are two important issues which this reliability study has taken under consideration: 1) the estimation of test stability (i.e., test-retest reliability), and 2) the assessment of inter-tester effects for those elements of the HSPV battery where there is reason to suspect that the tester may have an important effect on child performance. The first piece of information is useful for two reasons: 1) it is evidence on which to decide whether a particular test should be used in subsequent "program-effects" analysis, and 2) if it is used for such a purpose, it provides reliability estimates which are necessary for estimating true scores. As for the second piece of information, in any test which requires a significant amount of interaction between the tester and the test subject, the objectivity and expertise of the tester become very important. This is particularly important when testers are nested within sites, and when the training of these testers is also nested within sites. This is the case in the HSPV evaluation. Any tester biases (i. e., "level effects") become completely con-

founded with site effects, and could render the interpretation of such site and model effects impossible.

Design

The experiment was conducted at two sites (209, Salt Lake; and 2001, Kansas City), with three testers, two paraprofessionals and one expert, at each site. The test battery was to be administered twice to each child according to the following design on tester assignment:

FIGURE 1

TIME 1

		PP1	PP2	E	
	PP1	11	11	0	22
TIME 2	PP2	11	11	11	33
	E	11	0	15	26
		33	22	26	81

This design was to be implemented by SRI at both the Salt Lake and Kansas City sites.

The classes were to be chosen at random from those available to make up the 81 children/site. These children were to be assigned at random to the 7 cells of the design.

Sample

329 individual test batteries were forwarded to us by SRI. Of these, 33 were totally unusable (i.e., no data, missing identification fields, etc.). Of the remaining 296 there were 136 usable test pairs (the

remaining 24 were missing one of the two observations). Of these, 7 had one or more incomplete test codes. The remaining 129 units were used in all analyses, with the exception of the test-retest correlations which may be based on a slightly higher number (136 = max.). The breakdown of these 129 units is as follows:

FIGURE 3

Salt Lake City

	PP1	PP2	E1	
PP1	8	11	0	19
PP2	10	8	10	28
E1	11	0	13	24
	29	19	23	71

Kansas City

	PP3	PP4	E2	
PP3	11	7	0	18
PP4	6	10	9	25
E2	9	0	6	15
	26	17	15	58

Analysis:

The data were analyzed as a repeated measures model with Groups, i.e., tester pairing (a fixed effect) nested within sites. An Unweighted Means Analysis using the Datatext, Release 3 program and an exact least-squares analysis using Multivariance, Version 4 were performed. The results were thoroughly consistent. Table 1 contains the ANOVA tables for the Unweighted Means Analysis. The means and standard deviations on each of the tests

TABLE 1

UNWEIGHTED MEANS ANALYSIS OF VARIANCE TABLE
Groups within Sites Repeated Measures Model

CLASSIFYING FACTORS

SITE	SITE
GROUP	GROUP
TIME	TEST TIME
UNIT	SUBJECTS OR UNITS OF ANALYSIS

PSI

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F-TEST
SITE	837.288	1	837.288	13.503***
GROUP	336.408	12	29.701	0.479
UNIT	7131.047	115	62.009	NOT TESTED
TIME	37.318	1	37.318	8.940***
SITE X TIME	0.928	1	0.928	0.222
GROUP X TIME	30.369	12	2.531	0.606
TIME X UNIT	480.067	115	4.174	NOT TESTED
TOTAL	8873.414	257	34.527	

ITPA

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F-TEST
SITE	302.278	1	302.278	5.410*
GROUP	433.957	12	36.163	0.647
UNIT	6425.625	115	55.875	NOT TESTED
TIME	0.290	1	0.290	0.043
SITE X TIME	19.466	1	19.466	2.913
GROUP X TIME	213.800	12	17.817	2.666**
TIME X UNIT	768.474	115	6.682	NOT TESTED
TOTAL	8873.414	257	31.766	

Log TRANSFORM OF MOTOR INHIBITION

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F-TEST
SITE	0.004	1	0.004	0.022
GROUP	4.959	12	0.413	2.220*
UNIT	21.411	115	0.186	NOT TESTED
TIME	0.015	1	0.015	0.312
SITE X TIME	0.075	1	0.075	1.574
GROUP X TIME	2.210	12	0.184	3.860***
TIME X UNIT	5.38	115	0.048	NOT TESTED
TOTAL	34.113	257	0.133	

TABLE 1 (Con't.)

UNWEIGHTED MEANS ANALYSIS OF VARIANCE TABLE
Groups within Sites Repeated Measures Model

CLASSIFYIN FACTORS

SITE	SITE
GROUP	GROUP
TIME	TEST TIME
UNIT	SUBJECTS OR UNITS OF ANALYSIS

MOTOR INHIBITION

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F-TEST
SITE	43202.371	1	43202.371	0.792
GROUP	1151927.000	12	95993.875	1.759
UNIT	6274506.000	115	54560.918	NOT TESTED
TIME	17345.016	1	17345.016	1.303
SITE X TIME	30126.953	1	30126.953	2.264
GROUP X TIME	471969.563	12	39330.797	2.955**
TIME X UNIT	1530385.000	115	13307.695	NOT TESTED
TOTAL	9519460.000	257	37040.699	

ENUM: COUNTING

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F-TEST
SITE	30.535	1	30.535	3.282
GROUP	52.722	12	4.393	0.472
UNIT	1070.097	115	9.305	NOT TESTED
TIME	0.946	1	0.946	0.768
SITE X TIME	3.963	1	3.963	3.218
GROUP X TIME	22.023	12	1.835	1.490
TIME X UNIT	141.634	115	1.232	NOT TESTED
TOTAL	1321.920	257	5.144	

ENUM: POINTING AND TOUCHIN

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F-TEST
SITE	28.734	1	28.734	8.334**
GROUP	28.732	12	2.394	0.694
UNIT	396.493	115	3.448	NOT TESTED
TIME	0.440	1	0.440	0.498
SITE X TIME	0.005	1	0.005	0.006
GROUP X TIME	16.321	12	1.360	1.541
TIME X UNIT	101.554	115	0.883	NOT TESTED
TOTAL	572.260	257	2.227	

TABLE 1 (Con't.)

UNWEIGHTED MEANS ANALYSIS OF VARIANCE TABLE
Groups within Sites Repeated Measures Model

CLASSIFYING FACTORS

SITE	SITE
GROUP	GROUP
TIME	TEST TIME
UNIT	SUBJECTS OR UNITS OF ANALYSIS

ENUM: SAME ORDER

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F-TEST
SITE	0.186	1	0.186	0.038
GROUP	24.771	12	2.064	0.417
UNIT	569.713	115	4.954	NOT TESTED
TIME	2.603	1	2.603	2.381
SITE X TIME	2.663	1	2.663	2.435
GROUP X TIME	23.645	12	1.970	1.802
TIME X UNIT	125.764	115	1.094	NOT TESTED
TOTAL	749.345	257	2.916	

ENUM: SAME NUMBER

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F-TEST
SITE	0.290	1	0.290	0.176
GROUP	16.146	12	1.345	0.817
UNIT	189.486	115	1.648	NOT TESTED
TIME	2.834	1	2.834	3.666
SITE X TIME	0.114	1	0.114	0.148
GROUP X TIME	7.998	12	0.666	0.862
TIME X UNIT	88.903	115	0.773	NOT TESTED
TOTAL	305.771	257	1.190	

ENUM: SAME TOTAL

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F-TEST
SITE	0.942	1	0.942	0.102
GROUP	46.189	12	3.849	0.415
UNIT	1066.222	115	9.271	NOT TESTED
TIME	10.870	1	10.870	5.052*
SITE X TIME	3.881	1	3.881	1.804
GROUP X TIME	31.348	12	2.612	1.214
TIME X UNIT	247.462	115	2.152	NOT TESTED
TOTAL	1406.914	257	5.474	

broken down by tester pairing and test-retest are presented in Table 2. For a description of the tests and the scoring procedure, see Part II and Chart 1.

Of the elements of the test battery, the PSI, the ITPA, and the Pointing and Touching Subtest (Enumeration) demonstrated statistically significant site effects. Significant time effects (the test-retest period being separated by a period of 10 days to two weeks) were found for the PSI and the "same-total" subtest (Enumeration).

In terms of assessing tester bias, the important ANOVA term is the GROUP X TIME interaction term. A significant result here indicates that within a group (i.e., a tester pairing) at a given time point a result deviant from what might have been expected has occurred. The most reasonable interpretation of such an "interaction effect" is a tester level effect ("bias"). Significant results on the GROUP X TIME interaction were found for the ITPA, Log of the Motor Inhibition, and the Motor Inhibition. Examination of the 1 d.f. contrasts of the Group x Time Interaction from the exact least-squares analysis indicates the source of these significant results (see Tables 3A, 3B, 3C).

A word of explanation about these contrasts is perhaps in order. These are simple contrasts in which each tester pairing group is compared to Expert-Expert groups at that site.

TABLE 2 (1 of 5)

Mean and S.D.'s for Test and Retest Scores of the PSI (32-Item) ITPA Verbal Expression Subtest, MI, and ETS Enumeration.

Variable Description Tester ID Pre Post	NAME	TEST MEAN	TEST SD	TEST N	RETEST MEAN	RETEST SD	RETEST N
	PSI	16.667	5.826	129	17.434	5.935	129
PP1-PP1	1	17.500	4.770	8	17.750	5.238	8
PP1-PP2	2	17.100	3.885	10	18.800	4.771	10
PP1-E1	3	21.273	5.429	11	21.091	5.053	11
PP2-PP1	4	18.818	7.133	11	18.818	7.107	11
PP2-PP2	5	17.000	5.590	8	19.000	6.000	8
E1-E1	6	17.923	4.843	13	19.077	5.936	13
E1-PP2	7	18.000	4.960	10	19.400	5.919	10
PP3-PP3	8	14.091	6.748	11	15.273	6.826	11
PP3-PP4	9	16.667	3.496	6	17.167	2.967	6
PP3-E2	10	17.111	6.556	9	16.778	7.003	9
PP4-PP3	11	14.286	4.463	7	14.714	3.918	7
PP4-PP4	12	14.400	4.449	10	15.000	4.690	10
E2-E2	13	12.833	4.705	6	14.500	4.113	6
E2-PP4	14	13.222	5.094	9	13.778	4.391	9
	ITPA	12.729	5.869	129	12.946	5.455	129
PP1-PP1	1	14.125	4.428	8	14.625	4.211	8
PP1-PP2	2	9.500	3.628	10	12.300	4.776	10
PP1-E1	3	12.273	4.750	11	16.455	5.034	11
PP2-PP1	4	14.727	6.538	11	15.636	6.767	11
PP2-PP2	5	12.875	6.547	8	12.125	5.395	8
E1-E1	6	14.462	7.250	13	13.846	7.794	13
E1-PP2	7	17.100	6.024	10	14.500	5.296	10
PP3-PP3	8	9.818	4.658	11	10.091	4.981	11
PP3-PP4	9	11.167	6.563	6	11.167	3.578	6
PP3-E2	10	11.667	5.249	9	11.333	4.028	9
PP4-PP3	11	10.857	1.641	7	13.000	2.330	7
PP4-PP4	12	11.400	3.292	10	10.800	2.522	10
E2-E2	13	15.833	3.543	6	12.167	5.698	6
E2-PP4	14	11.222	7.146	9	11.444	5.230	9

TABLE 2 (2 of 5)

Variable Description / Tester ID Pre Post	NAME	TEST MEAN	TEST SD	TEST N	RETEST MEAN	RETEST SD	RETEST N
	LOGMI1	6.144	0.355	129	6.159	0.386	129
PP1-PP1	1	6.206	0.391	8	6.155	0.334	8
PP1-PP2	2	5.598	0.412	10	6.183	0.300	10
PP1-E1	3	6.193	0.444	11	5.685	0.443	11
PP2-PP1	4	6.321	0.367	11	6.366	0.398	11
PP2-PP2	5	6.431	0.171	8	6.295	0.204	8
E1-E1	6	5.815	0.315	13	5.892	0.256	13
E1-PP2	7	6.065	0.525	10	6.504	0.346	10
PP3-PP3	8	6.177	0.235	11	6.204	0.320	11
PP3-PP4	9	6.303	0.113	6	6.311	0.402	6
PP3-E2	10	6.141	0.277	9	6.206	0.296	9
PP4-PP3	11	5.995	0.208	7	6.084	0.215	7
PP4-PP4	12	6.046	0.165	10	6.218	0.291	10
E2-E2	13	6.078	0.238	6	6.037	0.350	6
E2-PP4	14	6.173	0.125	9	6.207	0.302	9
	MI1	492.534	188.337	129	509.046	200.695	129
PP1-PP1	1	535.375	216.247	8	501.625	204.152	8
PP1-PP2	2	488.100	271.830	10	508.600	170.486	10
PP1-E1	3	543.636	266.518	11	326.364	158.664	11
PP2-PP1	4	595.090	223.547	11	630.000	251.558	11
PP2-PP2	5	630.500	110.728	8	553.000	110.150	8
E1-E1	6	352.769	114.121	13	374.153	95.122	13
E1-PP2	7	493.300	255.579	10	707.600	233.662	10
PP3-PP3	8	495.090	116.542	11	520.727	168.690	11
PP3-PP4	9	549.667	62.358	6	597.833	248.305	6
PP3-E2	10	482.444	130.373	9	515.333	129.113	9
PP4-PP3	11	410.000	82.925	7	449.428	101.276	7
PP4-PP4	12	428.399	71.059	10	523.600	155.694	10
E2-E2	13	448.667	105.979	6	441.667	125.387	6
E2-PP4	14	483.555	59.768	9	522.111	187.678	9

TABLE 2 (3 of 5)

Variable Description / Tester ID	NAME	TEST MEAN	TEST SD	TEST N	RETEST MEAN	RETEST SD	RETEST N
Pre Post	EN13 (Counting)	2.698	2.247	129	2.860	2.271	129
PP1-PP1	1	3.500	2.784	8	3.000	2.646	8
PP1-PP2	2	2.100	1.921	10	3.400	2.154	10
PP1-E1	3	3.182	2.249	11	3.636	1.872	11
PP2-PP1	4	2.727	1.656	11	3.455	2.271	11
PP2-PP2	5	3.375	2.288	8	4.250	2.278	8
E1-E1	6	2.923	2.433	13	2.615	2.403	13
E1-PP2	7	2.800	2.135	10	2.900	2.022	10
PP3-PP3	8	1.909	2.234	11	1.818	1.992	11
PP3-PP4	9	3.667	1.795	6	2.333	2.055	6
PP3-E2	10	3.111	2.283	9	2.667	1.944	9
PP4-PP3	11	1.571	2.060	7	1.286	1.278	7
PP4-PP4	12	2.300	2.238	10	2.600	2.458	10
E2-E2	13	2.667	1.972	6	3.167	2.339	6
E2-PP4	14	2.222	1.988	9	2.667	2.108	9
	EN14 (Touching)	4.574	1.440	129	4.488	1.536	129
PP1-PP1	1	5.250	0.829	8	5.374	0.696	8
PP1-PP2	2	4.700	1.005	10	4.600	1.428	10
PP1-E1	3	5.364	0.771	11	4.909	0.996	11
PP2-PP1	4	5.000	1.128	11	4.636	1.150	11
PP2-PP2	5	5.125	1.965	8	4.500	1.871	8
E1-E1	6	4.385	1.273	13	4.846	1.231	13
E1-PP2	7	4.500	0.806	10	4.800	1.327	10
PP3-PP3	8	4.273	1.420	11	4.455	1.827	11
PP3-PP4	9	3.667	1.700	6	4.833	1.213	6
PP3-E2	10	3.444	1.892	9	2.889	1.853	9
PP4-PP3	11	4.429	1.294	7	4.429	1.400	7
PP4-PP4	12	4.600	1.960	10	4.000	1.844	10
E2-E2	13	4.833	1.067	6	4.000	1.528	6
E2-PP4	14	4.222	1.030	9	4.333	1.155	9

TABLE 2 (4 of 5)

Variable Description Tester ID Pre Post	NAME	TEST MEAN	TEST SD	TEST N	RETEST MEAN	RETEST SD	RETEST N
	EN15 (Same Order)	3.264	1.772	129	3.457	1.628	129
PP1-PP1	1	3.125	1.166	8	3.250	1.561	8
PP1-PP2	2	2.700	2.002	10	3.500	1.962	10
PP1-E1	3	3.000	1.859	11	3.091	1.881	11
PP2-PP1	4	3.091	1.621	11	4.182	0.936	11
PP2-PP2	5	3.375	2.342	8	3.500	2.236	8
E1-E1	6	3.077	2.759	13	3.846	2.032	13
E1-PP2	7	3.900	1.814	10	3.800	1.249	10
PP3-PP3	8	3.273	1.483	11	2.818	1.641	11
PP3-PP4	9	2.333	1.374	6	3.333	0.943	6
PP3-E2	10	3.556	1.165	9	2.667	1.414	9
PP4-PP3	11	3.857	0.639	7	3.429	0.495	7
PP4-PP4	12	4.100	0.943	10	3.300	1.269	10
E2-E2	13	3.333	1.491	6	4.333	1.374	6
E2-PP4	14	2.889	1.197	9	3.444	1.423	9
	EN16 (Same Number)	2.178	1.089	129	2.403	1.096	129
PP1-PP1	1	2.000	1.000	8	2.875	1.364	8
PP1-PP2	2	2.400	1.281	10	2.300	1.187	10
PP1-E1	3	2.091	0.900	11	2.182	0.936	11
PP2-PP1	4	3.182	1.029	11	3.000	1.206	11
PP2-PP2	5	2.125	0.927	8	2.125	1.269	8
E1-E1	6	1.515	1.273	13	2.538	1.393	13
E1-PP2	7	2.000	1.000	10	2.200	1.166	10
PP3-PP3	8	1.727	1.213	11	2.273	0.962	11
PP3-PP4	9	2.167	0.687	6	2.500	0.764	6
PP3-E2	10	2.222	0.786	9	2.667	0.471	9
PP4-PP3	11	2.429	0.904	7	2.286	1.030	7
PP4-PP4	12	2.300	0.640	10	2.100	0.831	10
E2-E2	13	2.167	1.067	6	2.500	0.957	6
E2-PP4	14	2.222	1.030	9	2.111	0.567	9

TABLE 2 (5 of 5)

Variable Description Tester ID Pre	Post	NAME		TEST MEAN	TEST SD	TEST N	RETEST MEAN	RETEST SD	RETEST N
			(Same Total)	5.442	2.332	129	5.860	2.332	129
PP1	PP1	1		5.125	1.763	8	6.125	2.472	8
PP1	PP2	2		5.100	2.468	10	5.800	2.561	10
PP1	E1	3		5.091	2.065	11	5.273	2.666	11
PP2	PP1	4		6.273	2.136	11	7.182	1.641	11
PP2	PP2	5		5.500	3.162	8	5.625	3.389	8
E1	E1	6		4.692	3.750	13	6.385	3.027	13
E1	PP2	7		5.900	2.385	10	6.000	2.049	10
PP3	PP3	8		5.000	2.412	11	5.091	2.193	11
PP3	PP4	9		4.500	1.607	6	5.833	1.213	6
PP3	E2	10		5.778	1.685	9	5.333	1.563	9
PP4	PP3	11		6.286	0.881	7	5.714	1.485	7
PP4	PP4	12		6.400	1.114	10	5.400	1.497	10
E2	E2	13		5.500	1.893	6	6.833	2.267	6
E2	PP4	14		5.111	1.100	9	5.556	1.771	9

TABLE 3

TABLE OF 1 DEGREE OF FREEDOM CONTRASTS FOR GROUP (TESTER PAIRING) x TIME INTERACTION

A: ITPA

CONTRAST	DIFFERENCE	S.E.	T-RATIO
(PP1-PP1)-(E1-E1)	1.115	1.6428	0.6787
(PP1-PP2)-(E1-E1)	3.415	1.5377	2.2333
(PP1-E1)-(E1-E1)	4.797	1.4977	3.202
(PP2-PP1)-(E1-E1)	1.524	1.4977	1.018
(PP2-PP2)-(E1-E1)	-0.135	1.6428	-0.082
(E1-PP2)-(E1-E1)	-1.985	1.5377	-1.291
(PP3-PP3)-(E2-E2)	3.439	1.8554	1.853
(PP3-PP4)-(E2-E2)	1.167	2.1107	0.553
(PP3-E2)-(E2-E2)	2.833	1.9268	1.470
(PP4-PP3)-(E2-E2)	5.310	2.0339	2.611
(PP4-PP4)-(E2-E2)	2.567	1.8878	1.360
(E3-PP4)-(E2-E2)	3.389	1.9268	1.759

B: LOG TRANSFORM OF M I

CONTRAST	DIFFERENCE	S.E.	T-RATIO
(PP1-PP1)-(E1-E1)	-0.128	0.2740	-0.467
(PP1-PP2)-(E1-E1)	0.517	0.2564	2.019
(PP1-E1)-(E1-E1)	-0.584	0.2498	-2.338
(PP2-PP1)-(E1-E1)	-0.031	0.2498	-0.124
(PP2-PP2)-(E1-E1)	-0.213	0.2740	-0.777
(E1-PP2)-(E1-E1)	0.362	0.2564	1.414
(PP3-PP3)-(E2-E2)	0.068	0.3094	0.220
(PP3-PP4)-(E2-E2)	0.048	0.3520	0.136
(PP3-E2)-(E2-E2)	0.105	0.3213	0.327
(PP4-PP3)-(E2-E2)	0.130	0.3392	0.383
(PP4-PP4)-(E2-E2)	0.213	0.3148	0.676
(E3-PP4)-(E2-E2)	0.074	0.3213	0.230

TABLE 3 (cont.)

TABLE OF 1 DEGREE OF FREEDOM CONTRASTS FOR GROUP (TESTER PAIRING) x TIME INTERACTION

C: M I

CONTRAST	DIFFERENCE	S.E.	T-RATIO
(PP1-PP1)-(E1-E1)	-55.135	73.3115	-0.572
(PP1-PP2)-(E1-E1)	-0.885	68.6233	-0.013
(PP1-E1)-(E1-E1)	-238.657	66.8370	-3.571
(PP2-PP1)-(E1-E1)	13.525	66.8370	0.202
(PP2-PP2)-(E1-E1)	-98.885	73.3116	-1.096
(E1-PP2)-(E1-E1)	192.915	68.6233	2.811
(PP3-PP3)-(E2-E2)	32.636	82.8003	0.390
(PP3-PP4)-(E2-E2)	55.167	94.1831	0.586
(PP3-E2)-(E2-E2)	39.889	85.9861	0.464
(PP4-PP3)-(E2-E2)	46.429	90.7667	0.512
(PP4-PP4)-(E2-E2)	102.200	84.2488	1.213
(E3-PP4)-(E2-E2)	45.556	85.9861	0.530

Note 1: There are 115 d.f. for error.
Significance level = 0.05 for student's t = ±1.98 (two tail test)
Significance level = 0.05 for Dunnett's t statistic = ±2.60 (two tail test)

Note 2: Dunnett's t statistic is a test for multiple comparisons of treatment means or contrasts among them with a control. It is based on the probability of falsely rejecting at least 1 comparison (C.F. Winer)

For example, the Group 5 contrast is as follows:

		TIME 1	TIME 2	
PP1-PP1	Group 1	0	0	
.	.	.	.	
.	.	.	.	
.	.	0	0	Nested within site
PP2-PP2	5	1	-1	
E1-E1	6	-1	1	
E1-PP2	7	0	0	

Under the null hypothesis of "no tester effects," the expectation of this contrast is zero. Further, by definition, the experts exhibit zero "tester bias." Thus, this contrast becomes an estimate of paraprofessional tester bias.

For all three measures, the source of the significant interaction is concentrated in the paraprofessional testers at the Salt Lake site. For the ITPA, however, this is also some indication of significant interactions at the Kansas City site. One must be somewhat cautious, however, in interpreting these results for the Kansas City site in that there were only 6 children with valid test batteries for the E2-E2 tester pairing.

In general, the results appear to indicate a strong tester bias on the part of paraprofessional 1 for all three measures (i.e., significant contrasts for (PP1-E1) - (E1-E1)).

There also appears to be a tester bias on the part of paraprofessional 2 for at least the Motor Inhibition. The presence of such significant results for such a small sample indicates a considerable likelihood for extensive tester level effects ("bias") in the general HSPV study. As such, the reliability of these instruments for our purposes is indeed questionable.

Test-Retest Correlations

The test-retest correlations for Salt Lake and Kansas City are presented in Tables 4 and 5 respectively. In Table 6, the data for the parallel groups (1 and 8, 2 and 9, etc.) is pooled across sites, and the correlations are recomputed. These correlations can be interpreted as coefficients of stability. They depend on conditions of retesting (e.g., tester), and the length of time between testings. As such, they are generally less than coefficients of precision.

The results for the PSI and the ITPA both across sites and tester conditions are in general quite good. The estimates for the Motor Inhibition are less impressive. (There is in fact one negative estimate [Group 7 at Kansas City]). This combined with the information on the susceptibility of this instrument to tester bias, makes it an unsatisfactory measurement device. As for the subtests of Enumeration, with the exception of the

TABLE 4

TEST RETEST RELIABILITY ESTIMATES BY TESTER PAIRING (Group #)

SITE 209: - SALT LAKE

TEST \ GROUP #	PP1-PP1 1	PP1-PP2 2	PP1-E1 3	PP2-PP1 4	PP1-PP2 5	E1-E1 6	E1-PP2 7
PSI	.936 (8)	.778 (10)	.920 (11)	.943 (11)	.894 (10)	.873 (13)	.962 (11)
ITPA	.660 (8)	.798 (9)	.774 (11)	.898 (11)	.949 (11)	.889 (13)	.850 (11)
LOGMI	.735 (8)	.743 (9)	.705 (11)	.675 (11)	.603 (11)	.673 (13)	.541 (11)
MI	.869 (8)	.883 (9)	.770 (11)	.682 (11)	.583 (11)	.615 (13)	.616 (11)
Enumeration Pointing & Touching	.054 (8)	.334 (10)	.516 (11)	.421 (11)	.893 (10)	.480 (13)	.548 (11)
Counting	.950 (8)	.595 (10)	.664 (11)	.831 (11)	.798 (8)	.640 (13)	.922 (10)
Same No.	.326 (8)	.725 (10)	.676 (11)	.229 (11)	.875 (10)	.867 (13)	.431 (11)
Same Order	.917 (8)	.579 (10)	.088 (11)	.147 (11)	.800 (10)	.507 (13)	.375 (11)
Same Total	.656 (8)	.763 (10)	.623 (11)	.219 (11)	.893 (10)	.878 (13)	.412 (11)

These are Pearson Product Moment Correlations.

TABLE 5

TEST RETEST RELIABILITY ESTIMATES BY TESTER PAIRING
(Group #)

SITE 2001: KANSAS CITY

TEST \ GROUP #	PP3-PP3 1	PP3-PP4 2	PP3-E2 3	PP4-PP3 4	PP4-PP4 5	E2-E2 6	E2-PP4 7
PSI	.891 (11)	.343 (6)	.833 (10)	.696 (8)	.687 (10)	.866 (6)	.971 (9)
ITPA	.726 (10)	.536 (7)	.841 (10)	.793 (8)	.455 (10)	.823 (6)	.797 (9)
LOGMI	.536 (10)	.348 (7)	.748 (10)	.340 (8)	.511 (10)	.820 (6)	-0.0307 (9)
MI	.526 (10)	.226 (7)	.804 (10)	.370 (8)	.529 (10)	.831 (6)	-0.0312 (9)
Pointing and Touching	.198 (11)	.458 (7)	.775 (9)	.845 (7)	.913 (10)	.818 (6)	-0.249 (9)
Counting	.936 (11)	.753 (6)	.760 (9)	.861 (7)	.931 (10)	.915 (6)	.309 (9)
Same No.	.875 (11)	.706 (7)	.247 (9)	-0.258 (7)	.894 (10)	.759 (6)	-0.0363 (9)
Same Order	.507 (11)	.215 (7)	.200 (9)	.175 (7)	.132 (10)	-0.408 (6)	-0.233 (9)
Same Total	.753 (11)	.649 (7)	.365 (9)	-0.265 (7)	.504 (10)	0.330 (6)	-0.203 (9)

These are Pearson Product Moment Correlations.

TABLE 6

TEST-RETEST RELIABILITY: POOLED ACROSS SITES

(Group #)

TEST \ GROUP #	1	2	3	4	5	6	7
PSI	.952 (19)	.658 (16)	.885 (21)	.903 (19)	.833 (20)	.882 (19)	.923 (20)
ITPA	.820 (19)	.569 (16)	.735 (21)	.882 (19)	.840 (18)	.857 (19)	.779 (19)
LOGMI	.626 (19)	.564 (16)	.555 (21)	.623 (19)	.459 (18)	.710 (19)	.302 (19)
MI	.744 (19)	.599 (16)	.550 (21)	.652 (19)	.421 (18)	.714 (19)	.381 (19)
Enumeration Pointing & Touching	.530 (19)	.318 (17)	.807 (20)	.620 (18)	.906 (20)	.499 (19)	.0276 (20)
Counting	.946 (19)	.496 (16)	.690 (20)	.801 (18)	.878 (18)	.700 (19)	.636 (19)
Same No.	.382 (19)	.726 (17)	.521 (20)	.0358 (18)	.833 (20)	.847 (19)	.0510 (20)
Same Order	.608 (19)	.484 (17)	.132 (20)	.242 (18)	.575 (20)	.304 (19)	.261 (20)
Same Total	.487 (19)	.737 (17)	.538 (20)	.108 (18)	.807 (20)	.790 (19)	.498 (20)

counting subtest, the estimates vary greatly from one subtest to another, from one group to another. Because of the relatively small sample sizes, interpretation of these variations is perhaps inappropriate. Nevertheless, it is quite clear that the seeming unreliability of the subtests make them inadequate for our uses.

CHART 1

SCORING PROCEDURES

PSI

All items on which a child is correct, either verbally or non-verbally, is scored 1. The test score consists of the sum of correct items (range = 0-32).

ITPA

The test score consists of the sum of all the number of times each category occurs for all objects.

Motor Inhibition

The test score is the slow time for the winding up on the Tow Truck Task. The time is measured in 1/10 secs. The "Log of the Motor Inhibition" is simply the natural log of the slow time for the tow truck. A child's slow score is used only if the child passed two out of the three pre-test tasks.

ETS Enumeration:

Counting Subtest: (Items 2A, 3A, 4A): Range (0-6) A child receives 1 point for correctly counting each item (maximum = 3 points). A child receives 1 point for telling how many points there are, either by giving the correct number (irregardless of whether or not he previously counted to that number), or by giving a single incorrect number which is the same number he just previously counted to (maximum = 3 points).

Touching Subtest: (Items 6B-11B): Range (0-6) A child receives 1 point for each correct item.

Same Number Matching: (Items 13C-20C): Range (0-8) A child receives 1 point for each correct item.

Same Order Matching: (Items 22C-27C): Range (0-6) A child receives 1 point for each correct item.

Same Total Subtest: This is the sum of the same number matching and same order matching subtests.

APPENDIX B

Eight-Block Sort Reliability Study

In the fall of 1971 in one site (Kansas City) Huron Institute and SRI conducted an inter-observer reliability study of the Eight-Block Sort observations using the observer form included in both the Fall 1971 and Spring 1972 batteries.[1] In this study two observers (paraprofessionals) simultaneously watched 20 children and three observers (two paraprofessionals and one expert trainer) simultaneously observed 8 children.

Definition of Variables

The variables used in this reliability study are displayed in Tables 1 and 2. The components of the Eight-Block Sort scoring sheet which constitute the variables outlined in Table 1 are numbered on the sample scoring sheet in Table 2. The SRI Spring 1972 scoring procedures manual for Eight-Block Sort observers is attached at the end of this study (Chart 1) to give more information about the meaning of these variables.

Some of the variables with their appropriate subcategories that one might like to obtain from the Eight-Block Sort observations are listed below:

---

[1] The observer forms used in the Fall 1969, Spring 1970, and Spring 1971 batteries were different.

TABLE 1

VARIABLES FOR ANALYZING INTER-OBSERVER RELIABILITY OF THE EIGHT-BLOCK SORT SCORING SHEET USED IN FALL 1971 AND SPRING 1972 OF THE HSPV STUDY

Variable #	Name of Variable[1]	Components of Score Sheet[2]
1	Orientation Time	2
2	Training Time	3
3	Mother's Training Time	1
4	Mom Indicates Future (0)	4
5	Mom Indicates Height (0)	5
6	Mom Indicates Mark (0)	6
7	Mom Indicates Ht.& Mk.(0)	7
8 = (4+5+6+7)	Mom Indicates Total (0)	4+5+6+7
9	Mom Reason (0)	10+11+12
10	Mom Praise (0)	8+9+18
11	Mom Blame (0)	13+14+15+17
12	Child Talk Height (0)	19
13	Child Talk Mark (0)	20
14	Child Talk Ht.& Mk. (0)	21
15 = (12+13+14)	Child Talk Total (0)	19+20+21
16	Child Non-Work (0)	22+23+24+26
17	Child Observe (0)	25
18	Mom Place Height (T)	27
19	Mom Place Mark (T)	28
20	Mom Place Ht.& Mk. (T)	29
21 = (18+19+20)	Mom Place Total (T)	27+28+29
22	Mom Talk Height (T)	30
23	Mom Talk Mark (T)	31
24	Mom Talk Ht.& Mk.(T)	32
25 = (22+23+24)	Mom Talk Total (T)	30+31+32
26 = (21+25)	Mom Train Total (T)	27+28+29+30+31+32
27	Mom Reason (T)	35+36+37
28	Mom Praise (T)	33+34+43
29	Mom Blame (T)	38+39+40+42
30	Child Place (T)	44
31	Child Goof (T)	45
32	Child Talk Height (T)	46
33	Child Talk Mark (T)	47
34	Child Talk Ht.& Mk. (T)	48
35 = (32+33+34)	Child Talk Total (T)	46+47+48
36 = (30+31)	Child Work (T)	44+45
37	Child Non-Work (T)	49+50+51+53
38	Child Observe (T)	52
39	Success Ht.Placement (0-2pts.)	54
40	Success Mk.Placement (0-2pts.)	55

TABLE 1 (cont.)

Variable #	Name of Variable[1]	Components of Score Sheet[2]
41 = (39+40)	Success Placement (0-2pts.)	54+55
42	Success Ht. Reason (0-2pts.)	56
43	Success Mk. Reason (0-2pts.)	57
44 = (42+43)	Success Reason (0-4pts.)	56+57
45 = (41+44)	Success Total (0-8pts.)	54+55+56+57

[1] Abbreviations used include (0) - Orientation period, (T) - training period, Ht. - height, Mk. - Mark, Pts. - points.

[2] Numerals represent which parts of the scoring sheet (see Table 2) are used for each variable

TABLE 2

EIGHT BLOCK SORT SCORING SHEET WITH NUMERALS IN SPACES USED TO FORM VARIABLES FOR INTER-OBSERVER RELIABILITY STUDY

Adult's Name _____ Date _____

Relationship to Child _____

Child's Name _____

Child's Ethnic Background _____

Teacher _____

School/Center _____

Community/City _____

State _____

Observer _____

Trainer _____

Comments (absenteeism, refusals, etc.) _____

Where was task administered and under what conditions? _____

PART I: TRAINER TEACHING MOTHER

MOTHER'S PERFORMANCE IN TRAINING SESSION

Time Started _____

	NUMBER OF TRIALS	DID SHE SUCCEED ON FINAL TRIAL?	ANY ERRORS TRAINER MADE? (Please specify)
CYCLE 1: HEIGHT	_____	Yes___ No___	_____
CYCLE 2: MARK	_____	Yes___ No___	_____
CYCLE 3: HEIGHT & MARK	_____	Yes___ No___	_____
CYCLE 4: SORT 8-BLOCKS	_____	Yes___ No___	_____

Time Ended _____

Time: trainer teaching mother - 1

Time Started _____

PART II: MOTHER TEACHING CHILD

ORIENTATION PERIOD Orientation Time - 2

MOTHER		CHILD	
INDICATES verbally or non-verbally:		SPECIFICALLY TALKING about:	
FUTURE TASK	4		
HEIGHT	5	HEIGHT	19
MARK	6	MARK	20
HEIGHT & MARK	7	HEIGHT & MARK	21
DIRECT REQUEST		DIRECT REQUEST	
RESPOND		RESPOND	
COMMENTS, PLAY		COMMENTS, PLAY	22
PRAISE	8	PRAISE	
ACKNOWLEDGE	9	ACKNOWLEDGE	
BEHAVIOR MODIFICATION		BEHAVIOR MODIFICATION	
reason	10	reason	
question	11	question	
firm	12	firm	
threaten, demean	13	threaten, demean	
punish	14	punish	
"I DON'T KNOW" IGNORE, NO RESPONSE		"I DON'T KNOW" IGNORE, NO RESPONSE	23
REFUSE, REJECT	15	REFUSE, REJECT	24
OBSERVE	16	OBSERVE	25
BRIBE	17	BRIBE	
ENCOURAGE	18	ENCOURAGE	
TASK IRRELEVANCY		TASK IRRELEVANCY	26

Time Started _____ Time Ended _____

TRAINING PERIOD

MOTHER

REQUESTS PLACING:

HEIGHT 27

MARK 28

HEIGHT & MARK _____

REQUESTS TALKING:

HEIGHT 30

MARK 31

HEIGHT & MARK 32

DIRECT REQUEST _____

RESPOND _____

TEACH _____

COMMENTS, PLAY _____

PRAISE 33

ACKNOWLEDGE 34

BEHAVIOR MODIFICATION _____

reason 35

question 36

firm 37

threaten, demean 38

punish 39

"I DON'T KNOW"
IGNORE, NO RESPONSE _____

REFUSE, REJECT 40

OBSERVE 41

BRIBE 42

ENCOURAGE 43

TASK IRRELEVANCY _____

CHILD

PLACING BLOCKS:

44

_____

GOOFING AROUND: 45

_____

SPECIFICALLY TALKING about:

HEIGHT 46

MARK 47

HEIGHT & MARK 48

DIRECT REQUEST _____

RESPOND _____

TEACH _____

COMMENTS, PLAY 49

PRAISE _____

ACKNOWLEDGE _____

BEHAVIOR MODIFICATION _____

reason _____

question _____

firm _____

threaten, demean _____

punish _____

"I DON'T KNOW"
IGNORE, NO RESPONSE 50

REFUSE, REJECT 51

OBSERVE 52

BRIBE _____

ENCOURAGE _____

TASK IRRELEVANCY 53

CHILD REQUESTED TO PLACE BLOCKS AND SAY WHY

	CHILD PLACED BLOCK WITH?	WHAT DID CHILD SAY?	MOTHER RESPONDED WITH → Rejection / Dissatisfaction / Neutrality / Nonverbal support / Verbal Support / Parent Answered
SHORT O		Q1	(Circle as many as appropriate)
		Q2	1 2 3 4 5 6
		Q3	
		Q4	
TALL X		Q1	1 2 3 4 5 6
		Q2	
		Q3	
		Q4	

FOR OFFICE USE ONLY

	Correctly Placed?		Correct Words?	
	Height	Mark	Height	Mark
SHORT O	- +	- +	O NV V	O NV V
	54	55	56	57
TALL X	- +	- +	O NV V	O NV V

Successful height placement - 56

Successful mark placement - 57

This Booklet was prepared by Stanford Research Institute, Menlo Park, California
for use under Office of Child Development, HEW, Contract No. HEW-OS-70-134

Aug 1971
8071

-- Mother's verbal communication to child
    a. positive
    b. negative
    c. neutral
    d. total

-- Task orientation and presentation to the child
    a. verbal
    b. non-verbal
    c. ordering of presentation

-- Child's responses to the mother
    a. work
    b. non-work
    c. verbal
    d. non-berbal

-- Mother's control system of child (i.e. motivational techniques)
    a. positive (i.e., praise, approval, support, etc.)
    b. negative (i.e., blame, criticism, disapproval, etc.)

-- Child's success during testing period
    a. placement success
    b. reason success

Unfortunately, only a few of these variables can be obtained from the observational sheets used in 1971-72. Some of the constraints imposed by the score sheets in defining variables are as follows:

-- In the orientation period, mother's verbal and non-verbal responses are recorded together. This is also true in both the orientation and training periods for categories such as "praise", "acknowledge", "threaten", "refuse", "bribe" and "encourage". Thus, it is impossible to get an accurate comprehensive verbal variable for the mother. The only valid verbal variable that can be specifically defined is mother's verbalness in task directions during training.

-- Behaviors observed are recorded in only one place (except for the behavior modification behaviors). For instance, if the mother gives an instruction about the heights of the blacks during the training period it is placed under "requests placing -- height" or "teach". It should be placed under the former since the general rule given to observers is to code under the more specific category. In many cases, the behaviors coded are confounded. For instances, if a mother says quickly, "I'll get you an ice cream cone if you place this block correctly", a mark is placed under "requests placing" and not "bribe". A request such as "put this where it belongs and later I'll get you an ice cream cone" is scored first under "requests placing" and then under "bribe".[1] Thus, it is probable that some of the control system variables (i.e., "bribe", "threaten", "praise", etc.) are listed with the "requests placing" category above the double line on the form, and that the behaviors recorded below the double line are most likely conservative estimates, while the categories above the double line are more valid.

-- Orientation is defined as "ended" when the mother gives her first instruction to the child to place a

---

[1] Information for coding obtained from an SRI trainer.

block. Thus, some orientation behaviors for some mothers are recorded under training. The mother who explains all dimensions of the blocks to her child before requesting placement of one block will probably appear to have a longer orientation period than the mother who explains one dimension and asks the child to place a block on that dimension before explaining other dimensions. Therefore, even though both of these mothers may use the same amount of orienting behaviors (although in a different sequence), the former mother will have more behaviors and time marked under orientation on the present form.

-- An accurate conception of how the mother teaches the child is hard to get since the sequencing of behaviors is not noted on the score sheet. The child's behaviors are recorded irregardless of what the mother requests or demands. Specific responses to the mother's requests are not known.

-- During the testing period of the child, the mother is instructed to be quiet -- i.e., neutral. If she is not neutral, the observer is required to identify the mother's behavior as "rejection, dissatisfaction, non-verbal support, verbal support, or parent answered." Since these variables are complicated by the fact that

the mother is instructed to be quiet, they can not be used as part of a larger variable such as "mother's verbalness". For example, a normally "rejecting" mother may not be "rejecting" during the testing period since she is told to be quiet.

In spite of these constraints the group of variables listed in Table 1 were chosen for preliminary analysis. In light of the problems outlined above, more confidence we have, more can be placed in those variables formed from categories above the double line for the orientation period (i.e., variable #s 4, 5, 6, 7, 8, 12, 13, 14, 15) and the training period (i.e., variable #s 18, 19, 20, 21, 22, 23, 24, 25, 26, 30, 31, 32, 33, 34, 35, 36), and in the success and time variables (i.e., variable #s 1, 2, 3, 39, 40, 41, 42, 43, 44, 45).

Another problem encountered in determining meaningful variables is the lack of time limits for the orientation and training periods. From a child development point of view, it can be argued that the percentage of the total time a mother or child engages in a particular behavior during the orientation and training period is a more satisfactory estimate of the behavior than just absolute frequency counts. For example, a mother who has ten tallies under "requests talking" during a five minute training session may be teaching her child quite differently from the mother who had ten tallies under the same category during a fifteen minute

training session. One of these two mothers is probably more verbal and/or more instructive; it is difficult to say that the two mothers are equivalent because they have the same number of tallies. Since it is not certain which of the two units af analysis (frequency counts vs. frequency per minute) is best for analyzing this observational data, both units were used in the reliability study for variables describing the orientation and training process. Only frequency counts were used for the success and time variables.

Item and Score Characteristics

Table 3 lists basic statistics (mean, standard deviation, skewness and kurtosis) for most of the 45 variables, using frequency count as the unit of analysis (Part A) and for several of the variables, using frequency per minute as the unit of analysis (Part B). These statistics are computed for each observer across twenty children. From these statistics using the frequency counts, it can be seen that several of the variables occur very infrequently: 4-Mom Indicates Future (0), 9-Mom Reason (0), 10-Mom Praise (0), 11-Mom Blame (0), 14-Child Talk Ht. & Mk. (0), 16-Child non-work (0), 29-Mom Blame (T), 31-Child Goof (T), and 34-Child Talk Ht. & Mk. (T).

Several of the variables are positively skewed; in most cases these are the same variables that occur very infrequently.

TABLE 3

BASIC STATISTICS (MEAN, S.D., SKEWNESS AND KURTOSIS) FOR SEVERAL EIGHT-BLOCK SORT VARIABLES FOR EACH OBSERVER (01, 02)[1]

Part A - Unit of Analysis = Frequency Count

Variable #/Name	Observer	Mean	S.D.	Skewness[2]	Kurtosis[2]
1-Orientation Time	01	1.400	1.319	.674	.778
	02	1.50	.910	.099	-1.111
2-Training Time	01	12.050	6.946	-.123	-1.337
	02	11.700	6.922	.121	-1.096
3-Mother Training Time	01	6.150	2.151	2.458***	6.141***
	02	6.800	2.272	2.167***	5.777***
4-Mom Indicate Future (0)	01	.600	.583	.363	-.723
	02	1.050	.865	.368	.689
5-Mom Indicate HT. (0)	01	3.350	2.903	1.086*	1.468
	02	2.400	2.223	.775	-.225
6-Mom Indicate Mk. (0)	01	3.050	2.376	.254	-.853
	02	2.950	2.459	.239	-.997
7-Mom Indicate Ht. & Mk. (0)	01	1.850	2.007	.614	-1.158
	02	1.250	1.577	.731	-1.167
8-Mom Indicate Total (0)	01	8.850	4.993	-.027	-.626
	02	7.650	4.819	.301	-.723
9-Mom Reason (0)	01	.200	.678	3.577***	11.628***
	02	.100	.300	2.667***	5.111***
10-Mom Praise (0)	01	1.900	2.567	1.636**	2.482*
	02	.950	1.857	2.088***	3.417**
11-Mom Blame (0)	01	.050	.218	4.129***	15.053***
	02	.100	.300	2.667***	5.111***
12-Child Talk Ht. (0)	01	2.750	2.605	.461	-1.076
	02	2.150	2.330	.435	-1.526
13-Child Talk Mk. (0)	01	2.050	2.132	.739	-.473
	02	1.800	2.337	.994	-.466
14-Child Talk Ht. & & Mk. (0)	01	.050	.218	4.129***	15.053***
	02	.100	.300	2.667***	5.111***
15-Child Talk Total (0)	01	4.850	4.304	.332	-1.185
	02	4.050	4.318	.745	-.733
16-Child Non-Work (0)	01	.300	.714	2.784***	7.379***
	02	1.800	6.735	4.060***	14.671***

[1] N = 19-20

[2] Significance levels * = .05
** = .01
*** = .001

TABLE 3 (con't.)

Variable #/Name	Observer	Mean	S.D.	Skewness[2]	Kurtosis[2]
18-Mom Place Ht. (T)	01	3.950	4.477	1.310*	.789
	02	3.300	3.648	1.180*	.494
19-Mom Place Mk. (T)	01	3.250	4.918	2.704***	7.694***
	02	3.400	5.132	2.370***	5.826***
20-Mom Place Ht. & Mk. (T)	01	15.800	15.045	1.212*	.333
	02	15.450	15.138	1.364*	.915
21-Mom Place Total (T)	01	23.000	17.697	1.247*	.668
	02	22.150	16.912	1.211*	.683
22-Mom Talk Ht. (T)	01	10.300	8.984	.619	-1.034
	02	8.900	8.185	.998	.018
23-Mom Talk Mk. (T)	01	7.550	5.500	.979	.869
	02	6.250	4.700	1.273*	2.732*
24-Mom Talk Ht. & Mk. (T)	01	8.900	7.981	1.151*	.792
	02	10.700	9.198	1.104*	.592
25-Mom Talk Total (T)	01	26.750	16.226	.416	-.685
	02	25.850	15.278	.509	-.518
26-Mom Train Total (T)	01	49.750	28.133	.741	.898
	02	48.000	27.631	.625	.409
27-Mom Reason (T)	01	4.000	4.940	1.603**	1.657
	02	1.800	2.088	1.387*	.738
28-Mom Praise (T)	01	18.250	9.762	.305	-.594
	02	14.650	9.551	.559	-.350
29-Mom Blame (T)	01	.350	.953	2.948***	7.772***
	02	.350	.792	2.310***	4.279***
30-Child Praise (T)	01	36.250	23.343	.811	.250
	02	35.500	25.463	.672	-.294
31-Child Goof (T)	01	.150	.477	3.173***	8.829***
	02	.150	.477	3.173***	8.829***
32-Child Talk Ht. (T)	01	11.450	7.046	.062	-1.473
	02	10.350	7.227	.407	-1.343
33-Child Talk Mk. (T)	01	10.550	7.652	.229	-1.078
	02	9.150	6.966	.482	-.696
34-Child Talk Ht. & Mk. (T)	01	1.000	2.280	2.681***	5.757***
	02	1.450	2.765	2.011***	2.796*
35-Child Talk Total (T)	01	23.000	13.539	-.066	-1.312
	02	20.950	12.706	.075	-1.249
36-Child Work (T)	01	36.400	23.427	.792	.193
	02	36.650	25.558	.658	-.334

[1] N = 19-20

[2] Significance levels * = .05
** = .01
*** = .001

TABLE 3 (con't.)

Variable #/Name	Observer	Mean	S.D.	Skewness[2]	Kurtosis[2]
37-Child Non-Work (T)	01	6.200	11.634	3.289***	10.213***
	02	5.800	9.750	3.197***	10.127***
39-Success Ht. Placement	01	1.789	.521	-2.443***	4.901***
	02	1.789	.521	-2.443***	4.901***
40-Success Mk. Placement	01	1.789	.521	-2.443***	4.901***
	02	1.789	.521	-2.443***	4.901***
41-Success Placement	01	3.579	.936	-2.158***	8.125**
	02	3.579	.936	-2.158***	3.125**
42-Success Ht. Reason	01	.842	.987	.320	-1.898
	02	.842	.987	.320	-1.898
43-Success Mk. Reason	01	1.158	.933	-.318	-1.775
	02	1.158	.933	-.318	-1.775
44-Success Reason	01	2.000	1.806	.054	-1.804
	02	2.000	1.806	.054	-1.804
45-Success Total	01	5.579	2.369	-.419	-.999
	02	5.579	2.369	-.419	-.999
Part B - Unit of Analysis = Frequency Per Minute					
8-Mom Indicate Total (0)	01	6.768	3.124	.347	-.470
	02	6.607	4.036	.496	.826
9-Mom Reason (0)	01	.036	.129	3.328***	9.077***
	02	.071	.175	2.041**	2.167
10-Mom Praise (0)	01	1.583	1.624	1.290*	1.317
	02	.810	1.285	1.612*	1.185
11-Mom Blame (0)	01	.071	.258	3.238***	9.077***
	02	.107	.279	2.494***	4.798
15-Child Talk Total (0)	01	3.708	2.545	.711	.379
	02	3.714	3.293	.772	-.372
16-Child Non-Work (0)	01	.286	.452	.949	-1.100
	02	1.286	3.963	3.271***	8.832***
21-Mom Place Total (T)	01	2.189	1.146	.811	-.289
	02	2.029	1.013	1.196*	.944
25-Mom Talk Total (T)	01	2.890	1.828	.553	-.417
	02	2.820	2.162	1.646**	2.234*

[1]N = 19-20

[2]Significance levels * = .05
** = .01
*** = .001

TABLE 3 (con't.)

Variable #/Name	Observer	Mean	S.D.	Skewness[2]	Kurtosis[2]
26-Mom Train Total (T)	01	5.078	2.503	1.140*	.525
	02	4.850	2.727	1.486**	1.928
27-Mom Reason (T)	01	.365	.366	.829	-.697
	02	.148	.157	.924	-.510
28-Mom Praise (T)	01	2.266	1.945	2.109***	4.578***
	02	1.748	1.518	1.346*	.560
29-Mom Blame (T)	01	.028	.072	2.648***	5.939***
	02	.038	.086	2.370***	4.645***
30-Child Place (T)	01	3.529	1.724	1.023	.230
	02	3.168	1.524	.865	2.263*
31-Child Goof (T)	01	.010	.032	3.535***	11.343***
	02	.009	.030	3.690***	12.374***
35-Child Talk Total (T)	01	2.686	1.858	.366	-1.124
	02	2.617	2.475	1.696**	2.273*
36-Child Work (T)	01	3.538	1.721	1.012	.225
	02	3.177	1.522	.852	2.267*
37-Child Non-Work (T)	01	.537	.814	3.153***	9.780***
	02	.380	.635	3.439***	11.504***

[2]Significance levels * = .05
** = .01
*** = .001

In addition, if the variable has a skewed distribution using frequency counts as the unit, it tends to also have a skewed distribution using frequency per minute as the unit. Variables which seem to be both infrequent and/or positively skewed (with both units of analysis) are 9, 10 and 11 (Mom's Reason, Praise and Blame during orientation), 29 (Mom's Blame during training), 14 and 16 (Child Talks Ht. & Mark and Child Non-work during orientation,) and 31, 34 and 37 (Child's Goofing, Non-work, and Talking Ht. & Mk. during training). Other variables which appear to have positively skewed distributions (based on frequency counts) are 3 (Mother's Training Time) and 19 (Mom Place Mk. during training).

The distribution of all the successful placement variables (39, 40, 41) are negatively skewed. There is a ceiling effect for these scores for both observers for every variable. Out of a total possible score of four points for successful placement, the mean for each observer for the twenty children was 3.579 (S.D. = .936).

Inter-observer Reliability

Reliability estimates for the two paraprofessional observers were calculated in two ways. One estimate of the observers' agreement is the Pearson product-moment correlation coefficient. These correlations for thirty variables using frequency counts as the unit of analysis are listed in Table 4. The coefficients (which range from .147 for Mom Reason (0) to 1.000 for the

TABLE 4

INTER-OBSERVER RELIABILITY COEFFICIENTS FOR TWO OBSERVERS FOR SEVERAL EIGHT-BLOCK SORT VARIABLES WITH FREQUENCY COUNTS USED AS UNIT OF ANALYSIS

Variable #/Name	Correlation Coefficient	R's from ANOVA	Significance F Observers[1]
1-Orientation Time	.783	.724	N.S.
2-Training Time	.936	.938	N.S.
Total Time (1 + 2)	.939	.938	N.S.
8-Mom Indicate Total (0)	.804	.787	N.S.
9-Mom Reason (0)	.147	.692	N.S.
10-Mom Praise (0)	.807	.124	N.S.
11-Mom Blame (0)	.688	.703	.025
12-Child Talk Ht. (0)	.929	.900	.025
13-Child Talk Mk. (0)	.865	.862	N.S.
14-Child Talk Ht. & Mk.(0)	.688	.655	.025
15-Child Talk Total (0)	.945	.931	.026
16-Child Non-Work (0)	.241	.939	N.S.
17-Child Observe (0)	.694	.051	N.S.
21-Mom Place Total (T)	.982	.981	N.S.
25-Mom Talk Total (T)	.958	.957	N.S.
26-Mom Train Total (T)	.987	.986	N.S.
27-Mom Reason (T)	.800	.976	N.S.
28-Mom Praise (T)	.884	.655	N.S.
29-Mom Blame (T)	.954	.829	.010
30-Child Place (T)	.940	.939	N.S.
31-Child Goof (T)	1.000	1.000	N.S.
32-Child Talk Ht. (T)	.908	.901	N.S.
33-Child Talk Mk. (T)	.935	.917	.025
34-Child Talk Ht.&Mk. (T)	---	.678	N.S.
35-Child Talk Total (T)	.963	.952	.025
36-Child Work (T)	.941	.940	N.S.
37-Child Non-Work (T)	.960	.948	N.S.
38-Child Observe (T)	.443	.787	N.S.
41-Success Placement	1.000	---	---
44-Success Reason	1.000	---	---
45-Success Total	1.000	---	---

[1] F test for observers equals $MS_{observers} \div MS_{observer \times units}$

N = 20 children observed

success scores and Child Goof (T) are quite adequate. Sixty percent of the r's are greater than .900. Correlation coefficients for the same two observers on a smaller group of variables using frequency per minute as a unit of analysis are listed in Table 5. These are not as high as those using frequency counts as a unit of analysis. They range from .164 for Child Work (T) to .998 for Child Goof (T). Only 12% of the variables have r's greater than .90. Thus, higher inter-observer agreement is obtained using frequency counts as the unit of analysis.

The other estimate of reliability was calculated from the sums of squares of a one-way repeated measures analysis of variance between the observers (with the observers' scores used as repeated measures) was completed for a particular variable, an estimate of reliability was calculated using the following formula:[1]

$$r = \frac{\frac{1}{h} MS_{unit} - MS_{within}}{\frac{1}{h} MS_{unit} - MS_{within}} + MS_{within}$$

where unit = observer's scores

h = # of observers

$$MS_{unit} = \frac{SS_{observer}}{df_{observer}}$$

$$MS_{within} = \frac{SS_{observer} + SS_{observer \times unit}}{df_{observer} + df_{observer \times unit}}$$

[1] B.J. Winer, Statistical principles in experimental design. New York, McGraw Hill, 1962.

TABLE 5

INTER-OBSERVER RELIABILITY COEFFICIENTS FOR TWO OBSERVERS FOR SEVERAL EIGHT-BLOCK SORT VARIABLES WITH FREQUENCY PER MINUTE AS UNIT OF ANALYSIS

Variable #/Name	Correlation Coefficient	R's From ANOVA	Significance F Observers[1]
21-Mom Place Total (T)	.465	.176	N.S.
25-Mom Talk Total (T)	.536	.995	N.S.
26-Mom Train Total (T)	.363	.307	.010
27-Mom Reason (T)	.660	.382	N.S.
28-Mom Praise (T)	.337	.174	N.S.
29-Mom Blame (T)	.544	.571	N.S.
30-Child Place (T)	.167	.321	N.S.
31-Child Goof (T)	.998	.554	N.S.
35-Child Talk Total (T)	.577	.474	N.S.
36-Child Work (T)	.164	.858	N.S.
37-Child Non-Work (T)	.900	.214	N.S.
8-Mom Indicate Total (0)	.720	---	---
9-Mom Reason (0)	.674	---	---
10-Mom Praise (0)	.792	---	---
11-Mom Blame (0)	.887	---	---
15-Child Talk Total (0)	.827	---	---
16-Child Non-Work (0)	.417	---	---

N = 20 children observed

[1] F test for observers equals $MS_{observer} \div MS_{observer \times units}$

The reliability coefficients calculated from the ANOVA design for the variables using frequency counts as the unit of analysis (Table 4) are very similar to the product-moment correlation coefficients. They range from .051 for Child Observe (O) to 1.000 for Child Goof (T), with 57% of the r's being greater than .900. Seven out ot 28 F tests for observer effects are significant, meaning that for these scores there is a significant component of variance due to differences in observers. This can be attributable to differences in observers' concentration and attention to the children observed or to actual differences in the frame of references the observers used to judge the variables.

The reliability estimates calculated from the ANOVA design for the variables using frequency per minute as the unit of analysis (Table 5) are often not similar to the product-moment correlations. These estimates range from 1.76 for Mom Place Total (T) to .995 for Mom Talk Total (T) with one out of 11 (9%) being greater than .900. Only one of the F tests for observer effects is significant.

Finally, estimates of inter-observer reliability for the three observers (N = 8 children) were calculated from the ANOVA design for ten selected variables using frequency counts as the unit of analysis (Table 6). All of these coefficients are quite adequate, ranging from .406 for orientation time to .988 for Child Talk Total (O). One-half of the r's is greater than .900; 90% is greater than .800. Four

TABLE 6

INTER-OBSERVER RELIABILITY CALCULATED FROM ANOVA FOR THREE OBSERVERS FOR SEVERAL EIGHT-BLOCK SORT VARIABLES WITH FREQUENCY COUNTS AS UNIT OF ANALYSIS

Variable #/Name	R from ANOVA	Significance[1]
8-Mom Indicate Total (0)	.942	N.S.
1[illegible]-Child Talk Total (0)	.988	N.S.
21-Mom Place Total (T)	.890	.003
25-Mom Talk Total (T)	.834	.011
26-Mom Train Total (T)	.843	.001
35-Child Talk Total (T)	.920	.015
36-Child Work (T)	.873	N.S.
1-Orientation Time	.406	N.S.
2-Training Time	.911	N.S.
Total Time (1 + 2)	.914	N.S.

N = 8 children observed.

[1] F test for observers equals $MS_{observer} \div MS_{observer \times units}$

of the F tests (40%) for observer effects are significant. After looking at the raw data, this seems to be attributable to the fact that the expert observer was using a different frame of reference for rating than the two paraprofessionals.

Variable Intercorrelations

The intercorrelations of most of the Eight-Block Sort variables listed in Table 1 using frequency counts as the unit of analysis are listed in Table 7 for each of the two observers. The first number in the appropriate space is the correlation for the two variables for observer 1 based on 20 children, while the second number in the space is the same correlation for observer 2. In most cases, the correlations for both observers are very similar. The intercorrelations among the success scores (variable #s 39-45) are exactly the same for both observers.

In general, many correlations are low. Some of the more interesting higher correlations between variables are as follows:

-- Mom Indicate Total (O) correlates .82/.76 with Child Talk Total (C).

-- Training Time correlates .66/.70 with Mom Place Total (T) .62/.66 with Mom Talk Total (T), .77/.80 with Mom Train Total (T), and .82/.83 with Child Work (T).

-- Mom Train Total (T) correlates -.52/-.49 with Child Talk Total (O), and .60/.52 with Mom Praise (T).

-- Child Work (T) correlates .73/.76 with Mom Place HT. (T) and .56/.60 with Mom Place Total (T).

TABLE 7

Inter-correlations of Several Eight-Block Sort Variables Using Frequency Counts as Unit of Analysis for Two Observers[1]

Variable #/Name[b]	1[c]	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	39	40	41	42	43	44
2	-.13																																									
Training Time	-.19																																									
3	-.16	-.02																																								
Mom Training Time	-.06	+.39																																								
4	-.05	+.07	-.35																																							
(C)Mom Indicate Future	.05	-.04	-.17																																							
5	.55	-.45	-.24	.35																																						
(C)Mom Indicate Ht.	.32	-.46	.14	.59																																						
6	.33	.16	-.18	-.06	.33																																					
(C)Mom Indicate Mk.	.20	-.20	-.09	.21	.53																																					
7	.09	-.25	-.41	-.22	.07	.06																																				
(C)Mom Indicate Ht.& Mk.	.32	-.23	-.24	-.09	-.03	.06																																				
8	.50	-.29	-.44	.20	.81	.69	.45																																			
(C)Mom Indicate Total	.41	-.42	-.09	.53	.83	.62	.33																																			
9	-.20	-.01	-.16	.08	-.16	-.16	-.02	-.17																																		
(C)Mom Reason	.31	-.09	.15	.17	.24	.14	-.02	.20																																		
10	.51	-.47	-.33	.37	.70	.15	.11	.57	-.22																																	
(C)Mom Praise	.45	-.24	.14	.56	.56	.35	.00	.57	.46																																	
11	-.07	-.30	-.02	.16	.13	.19	.07	.19	-.07	-.08																																
(C)Mom Blame	.13	-.35	-.19	-.21	-.14	.28	.26	.13	-.11	-.17																																
12	.61	-.43	-.37	-.05	.71	.33	.36	.70	-.17	.68	.02																															
(C)Child Talk Ht.	.53	-.59	.09	.15	.69	.43	.32	.60	-.43	.41	.12																															
13	.22	-.50	-.39	.10	.65	.45	.41	.29	-.15	.61	+.01	.64																														
(C)Child Talk Mk.	.32	-.50	.01	.35	.69	.56	.24	.29	.24	.42	.03	.72																														
14	.38	.10	-.12	.16	.05	-.20	.47	.15	-.07	.10	-.05	.02	+.01																												
(C)Child Talk Ht.& Mk.	.50	.26	-.12	-.02	-.29	-.33	.26	-.22	.44	.19	-.11	.05	-.19																												
15	.50	-.51	-.42	.02	.77	.43	.44	.82	-.18	.72	.01	.92	.88	.06																											
(C)Child Talk Total	.49	-.57	.04	.28	.72	.50	.37	.76	-.34	.49	.07	.93	.92	.00																											
16	.10	.41	.23	-.19	-.17	.41	-.11	.08	-.12	-.20	-.10	-.23	-.24	-.10	-.26																											
(C)Child Non-Work	.25	-.15	-.18	-.12	-.24	.21	.40	.07	-.09	-.12	.08	.02	-.12	-.09	.06																											
18	-.11	.55	.17	-.18	-.46	-.21	-.51	-.52	-.21	-.34	.20	-.41	-.48	.05	-.48	.11																										
(I)Mom Place Ht.	-.27	.51	-.20	-.23	-.37	-.17	-.27	-.39	-.32	-.14	-.21	-.44	-.33	-.07	-.42	-.05																										
19	.00	.35	.09	-.07	-.22	-.10	-.21	-.18	-.20	-.01	-.15	.01	-.18	.22	-.07	.31	.19																									
(I)Mom Place Mk.	.05	.31	-.25	.19	-.22	-.11	.13	-.08	-.22	-.00	-.06	-.23	-.16	.07	-.20	.07	.24																									
20	-.41	.46	-.21	.35	-.35	.03	-.12	-.33	.21	-.35	-.13	-.42	-.23	-.15	-.41	-.08	.24	.02																								
(I)Mom Place Ht.& Mk.	-.53	.51	-.45	-.01	-.42	-.20	-.33	-.37	-.09	-.33	-.23	-.42	-.22	-.08	-.38	-.22	.23	-.11																								
21	-.37	.66	-.12	.24	-.50	-.06	-.43	-.46	-.09	-.40	-.21	-.50	-.37	-.06	-.49	.04	.51	.34	.91																							
(I)Mom Place Total	-.52	.70	-.52	-.09	-.63	-.25	-.43	-.54	-.19	-.53	-.27	-.59	-.32	-.06	-.49	-.19	.49	.26	.91																							
22	-.23	.64	.06	-.12	-.36	.06	-.09	-.22	-.16	-.36	-.19	-.50	-.36	-.02	-.48	.42	.54	.02	.14	.26																						
(I)Mom Talk Ht.	-.16	.56	-.15	.15	-.64	.16	-.01	.17	-.33	-.03	-.22	-.43	-.23	-.14	-.36	-.11	.29	.21	.05	.23																						
23	-.25	.24	-.39	.23	-.25	-.09	.21	-.04	-.03	-.38	-.11	-.25	-.26	-.02	-.29	.69	-.09	.07	.37	.31	.11																				
(I)Mom Talk Mk.	-.09	.15	-.41	-.29	-.33	-.10	.14	-.14	-.02	-.30	-.02	-.15	-.25	-.02	-.22	.05	-.06	.27	.22	.26	.02																				
24	-.08	.39	-.28	.10	-.03	.54	.12	.30	.02	-.24	.00	-.14	-.16	-.06	-.01	.13	.00	-.32	.40	.25	.45	.15																				
(I)Mom Talk Ht.& Mk.	-.07	.52	-.37	.14	-.11	.05	.05	.03	.26	-.12	-.08	-.19	-.08	.24	-.13	-.09	.08	-.37	.55	.44	.32	.11																				
25	.24	.62	-.31	.05	-.50	.26	.08	.08	.20	-.31	-.14	-.43	-.21	-.03	-.37	.32	.22	-.12	.40	.37	.81	.47	.79																			
(I)Mom Talk Total	-.16	.66	-.43	.02	-.39	.20	.13	.03	-.10	-.19	-.17	-.40	-.23	-.09	-.33	-.10	.30	.09	.42	.47	.34	.19	.81																			
26	-.37	.77	-.21	.19	-.49	.13	-.22	-.29	-.36	-.50	-.21	-.52	-.35	-.04	-.52	.23	.47	.14	.81	.85	.69	.47	.61	.81																	
(I)Mom Train Total	-.40	.80	-.56	.05	-.37	-.20	-.45	-.26	-.17	-.33	-.26	-.59	-.32	-.01	-.49	-.17	.46	.21	.79	.87	.55	.32	.72	.84																	
27	.15	.33	-.18	-.14	-.02	.11	-.12	-.01	.43	-.23	-.14	-.39	-.16	-.39	-.03	-.09	.08	.07	.26	.27	-.25	.37	.17	.07	.21																
(I)Mom Reason	.12	.48	-.09	-.22	-.03	-.19	-.43	-.22	.32	-.23	-.11	-.09	-.26	-.07	-.08	-.05	.21	-.13	.23	.21	.08	.09	.47	.35	.32																
28	-.38	.56	-.14	.33	-.17	.01	-.09	-.09	.06	-.21	-.12	-.33	-.25	-.11	-.32	.23	.02	.36	.56	.58	.33	.37	.18	.40	.60	.02															
(C)Mom Praise	-.06	.32	-.40	.13	-.14	-.34	.04	-.20	-.12	-.13	-.22	-.34	-.28	-.33	-.31	-.31	.02	.28	.51	.24	.14	.22	.33	.34	.52	.16															
29	.36	.13	-.22	-.20	-.06	-.01	-.10	-.10	-.13	.02	-.08	.21	-.10	-.08	.05	-.08	.42	.03	.04	.15	-.04	-.05	.10	.02	.11	.46	-.40														
30	.13	.23	-.13	-.17	.01	.14	-.40	-.02	.06	.11	.06	.22	.09	-.15	.16	.26	.40	-.10	.07	.12	-.10	-.01	.06	-.03	.06	.41	-.46														
31	-.13	.42	.04	-.21	-.50	.11	-.25	-.34	-.13	-.37	-.28	-.31	-.35	-.03	-.37	.30	.72	.38	.32	.56	.46	-.01	.07	.29	.52	.19	.24	.17													
(I)Child Place	-.34	.83	-.29	-.06	-.34	-.03	-.25	-.37	-.09	-.39	-.41	-.43	-.41	.79	-.45	-.31	.75	.38	.37	.60	.49	.00	.19	.38	.57	.37	.22	.05													
32	.38	.19	-.17	-.14	-.21	-.02	-.19	-.10	-.03	-.03	-.07	.23	-.20	-.07	.04	-.13	.45	.03	.06	.19	-.01	-.01	.10	.04	.14	.51	-.35	.92	.17											
(I)Child Goof	-.06	.32	-.11	-.26	-.06	-.03	-.25	-.17	-.11	-.07	-.11	.16	-.11	-.11	.02	.01	.46	-.07	.10	.17	-.04	.05	.04	.02	.11	.48	-.33	.92	.19											
33	-.27	.46	-.15	-.14	-.28	-.10	-.33	-.00	-.09	-.30	-.11	-.39	-.13	-.21	-.29	.22	.36	-.04	.05	.12	.57	.29	.54	.45	.56	-.22	.25	-.08	.30	-.08										
(I)Child Talk Ht.	-.09	.33	-.21	-.17	-.06	-.03	-.21	-.11	-.32	-.04	-.02	-.26	-.08	.05	-.18	.04	.41	.20	-.17	.09	.83	.06	.25	.41	.34	-.08	.07	-.20	.22	-.16										
34	-.22	.39	-.26	-.16	-.15	-.24	-.21	-.13	-.32	-.29	-.08	-.36	-.07	-.16	-.25	.26	-.12	-.14	.36	.24	.61	.50	.69	.84	.64	-.08	.52	-.20	-.02	-.19	.72									
(C)Child Talk Mk.	-.02	.38	-.37	-.29	-.06	-.19	-.31	-.25	-.25	-.04	-.08	-.28	-.02	-.09	-.15	-.10	.02	.00	.29	.22	.62	.15	.75	.83	.62	.05	.37	-.15	.03	-.13	.62									
35	-.01	-.36	-.29	-.01	.00	-.01	.40	-.03	-.03	.21	-.10	.23	.40	-.10	.33	.09	-.17	-.06	-.29	-.30	-.23	-.08	-.06	-.18	-.30	-.21	-.29	.05	-.06	-.14	-.09	-.14								
(I)Child Talk Ht.& Mk.	.17	.23	-.05	-.04	-.02	.01	.25	.06	-.05	-.11	.25	.08	.06	.26	-.18	.17	-.24	-.15	.03	-.07	-.29	.11	-.04	-.15	-.15	-.22	-.11	.02	-.08	-.13	-.17	-.06								
36	-.30	.42	-.28	.01	-.23	.19	.36	.19	.19	-.33	-.12	-.32	-.04	.19	-.23	.26	.09	-.11	.18	.15	.76	.42	.66	.89	.60	-.20	.38	-.15	.13	-.17	.91	.02	.04								
(I)Child Talk Total	-.01	.34	-.34	-.20	-.01	.09	.18	.21	.23	.02	.00	-.29	-.00	.04	-.15	.07	.19	.08	.07	.13	.55	.14	.55	.77	.51	-.04	.22	-.19	.08	-.19	.87	.09	.09								
37	-.12	.82	.04	-.21	-.50	.11	-.28	-.38	-.13	-.37	-.28	-.32	-.35	-.03	-.36	.29	.23	.38	.32	.56	.46	-.01	.07	.29	.52	.20	.23	.18	1.00	.19	.29	-.02	-.06	.13							
(C)Child Work	-.34	.83	-.21	-.16	-.34	-.16	-.36	-.37	-.09	-.29	-.41	-.43	-.41	.78	-.45	-.31	.76	.24	.37	.60	.48	.00	.19	.39	.57	.37	.26	.07	1.00	.21	.21	.02	-.29	.07							
38	.44	.24	-.17	-.31	.07	.04	-.04	.01	.01	-.02	-.10	.35	-.12	-.00	.13	-.10	.33	.11	-.10	.03	-.13	.03	-.01	-.07	-.02	.72	-.33	.86	.23	.89	-.12	-.27	-.16	-.24	.24						
(I)Child Non-Work	.16	.41	-.07	-.24	-.11	-.23	-.29	-.33	-.03	-.07	-.20	.11	-.32	.09	-.11	-.05	.51	.14	-.05	.10	.00	.08	-.12	-.05	.04	.59	-.17	.22	.38	.87	-.11	-.27	-.17	-.25	.40						
39	-.39	-.22	.16	.27	-.07	-.05	.09	.01	-.02	.09	.10	-.41	.17	.10	-.16	.04	-.11	-.38	.07	-.08	.35	-.10	.16	.25	.09	-.79	.16	-.67	-.19	-.70	.37	.40	.18	.46	-.20	-.90					
Success Ht. Placement	-.24	-.24	-.04	.24	.10	.16	.21	.27	-.12	.11	.14	-.26	.28	-.19	.09	.02	-.15	-.27	.11	-.02	.27	-.21	.10	.15	.07	-.65	.11	-.57	-.24	-.70	.38	.19	.22	.49	-.25	-.86					
40	-.16	-.18	.16	.27	-.03	-.14	-.06	-.09	-.17	.20	.10	-.21	.03	.06	-.11	.18	-.02	.26	-.11	-.05	.39	-.16	-.07	.14	.04	-.81	.08	-.26	-.11	-.28	.30	.29	.18	.36	-.11	-.60	.61				
Success Mk. Placement	-.13	-.17	-.04	.39	.24	.35	.04	.14	-.02	.26	.14	-.12	.32	-.02	.05	.02	-.01	.21	-.17	-.09	.34	-.21	-.13	.04	-.03	-.69	-.18	-.19	-.18	-.28	.37	.35	.22	.46	-.18	-.45	.61				
41	-.31	-.22	.18	.39	-.06	-.19	.02	-.01	-.11	.11	.11	-.34	.11	.11	-.15	.12	-.07	-.12	-.02	-.07	.41	-.14	.05	.21	.07	-.90	.14	.52	-.17	-.55	.37	.38	.20	.46	-.18	-.83	.90	.90			
Success Placement	-.21	-.23	-.05	.45	.22	.38	.16	.30	-.30	.15	.15	-.24	.34	-.40	.05	.05	-.09	-.03	-.03	-.06	.34	-.24	-.02	.10	.02	-.75	-.04	-.42	-.23	-.55	.42	.41	.24	.53	-.24	-.73	.90	.90			
42	-.05	-.43	-.26	.36	.40	-.03	.50	.49	-.26	.21	.28	.12	.39	.28	.28	-.08	-.35	-.13	-.41	-.48	-.03	-.05	.12	-.01	-.32	-.46	-.12	-.21	-.48	-.28	.20	.17	.44	.29	-.48	-.30	.35	.35	.38		
Success Ht. Reason	.28	.42	-.00	.22	.33	.14	.52	.44	.06	.23	.40	.27	.40	.06	.38	.26	-.36	-.05	-.40	-.45	-.06	.03	-.01	-.04	-.31	-.21	-.04	-.13	-.57	-.28	.29	.27	.18	.36	-.57	-.38	.35	.35	.38		
43	.03	-.45	-.20	.49	.52	-.09	.47	.31	-.38	.15	.21	.03	.36	.21	.21	-.15	-.21	-.09	-.07	-.14	-.15	-.14	.19	-.04	-.11	-.47	-.12	-.12	-.48	-.17	.05	.07	.31	.12	-.48	-.37	.39	.39	.44	.77	
Success Mk. Reason	.15	-.32	-.26	.24	.24	.19	.32	.35	-.09	.06	.31	.12	.51	-.06	.37	.18	-.25	-.02	-.05	-.11	-.11	-.03	.06	-.03	-.09	-.26	-.00	-.01	-.50	-.17	.10	.22	.05	.19	-.50	-.43	.39	.39	.44	.77	
44	-.02	-.47	-.24	.40	.39	-.09	.41	.43	-.34	.18	.26	.08	.40	.26	.26	-.12	-.30	-.12	-.26	-.33	-.13	-.10	.16	-.03	-.23	-.50	-.13	-.18	.51	-.24	.14	.13	.40	.22	-.51	-.36	.39	.39	.44	.94	.94
Success Reason	.23	-.40	-.13	.27	.30	.16	.44	.42	-.00	.15	.38	.24	.48	.00	.39	.24	-.32	-.03	-.25	-.30	-.10	.00	.03	-.04	-.21	-.25	-.03	-.07	-.57	-.24	.21	.26	.12	.30	-.57	-.43	.39	.39	.44	.94	.94
45	-.14	-.44	-.11	.40	.27	-.09	.32	.31	-.30	.18	.24	-.08	.35	.24	.14	-.05	-.26	-.14	-.21	-.28	.06	-.13	.14	.07	-.15	-.33	-.04	-.34	-.45	-.40	.25	.25	.39	.35	-.46	-.60	.65	.65	.73	.87	.89
Success Total	.09	-.39	-.12	.39	.31	.24	.43	.49	-.16	.18	.35	.09	.50	-.16	.31	.20	-.28	-.04	-.20	-.25	.06	-.09	.01	.01	-.16	-.49	-.04	-.22	-.53	-.40	.33	.36	.19	.44	-.53	-.62	.65	.65	.73	.87	.89	.94

[a] The 1st correlation is for observer 1; the 2nd correlation is for observer 2, N=19 or 20.

[b] See Tables 1 and 2.

[c] Variable 1 = orientation time

-- Child Goof (T) correlates .97/.92 with Mom Blame (T).

-- Child Talk Total (T) correlates .89/.77 with Mom Talk Total (T).

-- Child Non-Work (T) correlates .72/.59 with Mom Reason (T) and .86/.72 with Mom Blame (T).

Successful height placement is negatively correlated (60's to 80's range) with Mom Reason (T), Mom Blame(T), Child Goof (T) and Child Non-Work (T). Successful mark placement is negatively correlated with Mom Reason (-.81/-.69) and Child Non-Work (-.60/-.45) during training. Successful placement (total) is negatively correlated with Mom Reason (-.90/-.75) and Child Non-Work (-.83/-.73) during training. The largest correlation for Total Success score was with Child Non-Work during training (-.60/-.62).

Intercorrelations for the two observers on some of the variables using frequency per minute as a unit of analysis are listed in Table 8. There are more discrepancies between the two observers' correlations for two variables using this unit of analysis than there are when frequency counts is the unit of analysis (Table 7). In addition, there are only a few interesting large correlations: Child Talk Total (T) correlates .95/.96 with Mom Talk Total (T); Child Non-Work (T) correlates .88/.92 with Child Goof (T). Of the three success scores used in the analysis, only the successful placement score correlates with a particular behavior in training in the .50's or above for both observers: -.66/-.58 with Mom Reason (T) and -.59/-.60 with Child Goof (T).

TABLE 8

INTER-CORRELATIONS OF SEVERAL EIGHT-BLOCK SORT VARIABLES USING FREQUENCY PER MINUTE AS UNIT OF ANALYSIS[a] FOR TWO OBSERVERS

Variable #[c]/name[b]	1	2	3	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	41	44
	-.13																								
Training Time	-.19																								
3	-.16	-.02																							
Mom Training Time	-.06	-.39																							
18	-.23	.30	.43																						
(T)Mom Place Ht.	-.25	.48	-.08																						
19	-.13	.05	.26	.30																					
(T)Mom Place Mk.	.04	-.02	-.17	.06																					
20	.10	-.44	-.09	-.35	-.44																				
(T)Mom Place Ht.& Mk.	-.38	-.29	-.25	-.31	-.30																				
21	.02	-.42	.06	-.09	-.17	.94																			
(T)Mom Place Total	-.45	-.23	-.34	-.12	.04	.93																			
22	.08	-.15	-.07	-.22	.22	.22	.14																		
(T)Mom Talk Ht.	-.01	-.22	-.11	-.15	.39	.06	.17																		
23	.17	-.64	-.38	-.51	-.37	.60	.46	.26																	
(T)Mom Talk Mk.	.07	-.56	-.26	-.33	.02	.40	.38	.13																	
24	.22	-.54	-.32	-.51	-.87	.41	.25	-.05	.53																
(T)Mom Talk Ht.& Mk.	.05	-.40	-.15	-.53	.11	.41	.38	.53	.50																
25	.23	-.64	-.37	-.60	-.46	.57	.39	.47	.81	.81															
(T)Mom Talk Total	.05	-.48	-.20	-.46	.22	.37	.40	.73	.63	.93															
26	.18	-.66	-.24	-.48	-.41	.64	.74	.41	.80	.70	.91														
(T)Mom Train Total	-.13	-.46	-.29	-.41	.19	.64	.69	.64	.64	.88	.94														
27	.21	-.16	-.11	-.08	-.19	.13	.08	-.60	.24	.26	.01	.04													
(T)Mom Reason	.33	.06	.01	.05	-.30	-.18	-.29	-.35	-.06	-.04	-.17	-.24													
28	.28	-.67	-.17	-.54	-.31	.68	.56	.10	.66	.82	.80	.84	.19												
(T)Mom Praise	.02	-.55	-.10	-.33	.07	.57	.59	.38	.60	.77	.76	.82	-.01												
29	.33	.03	-.26	.39	.17	-.24	-.14	-.17	-.14	-.10	-.18	-.20	.26	-.32											
(T)Mom Blame	.31	-.15	.08	.02	-.13	-.17	-.23	-.32	-.07	-.09	-.19	-.24	.40	-.33											
30	.38	-.52	.09	-.03	.00	.40	.44	.11	.32	.27	.33	.45	-.02	.43	.09										
(T)Child Place	-.18	-.15	.21	-.15	-.04	.09	.12	.11	.11	.15	.16	.17	.22	.53	-.36										
31	.41	.15	-.16	.41	.02	-.19	-.12	-.17	-.18	-.16	-.23	-.22	.41	-.29	.87	-.08									
(T)Child Goof	.13	.24	-.06	.40	-.11	-.24	-.22	-.19	-.13	-.20	-.20	-.26	.35	-.30	.45	-.07									
32	.24	-.65	-.20	-.47	-.29	.48	.36	.73	.68	.50	.86	.79	-.30	.59	-.16	.44	-.25								
(T)Child Talk Ht.	.02	-.53	-.06	-.33	.54	.26	.29	.30	.50	.64	.94	.85	-.31	.69	-.19	.15	-.22								
33	.02	-.59	-.38	-.68	-.37	.45	.26	.50	.73	.65	.88	.76	-.11	.63	-.26	.07	-.35	.81							
(T)Child Talk Mk.	.04	-.47	-.15	-.47	.23	.37	.39	.72	.52	.94	.97	.92	-.26	.76	-.20	.12	-.23	.94							
34	.20	-.50	-.28	-.20	.00	.15	.13	-.27	.40	.68	.47	.40	.25	.58	.14	.58	-.13	.22	.28						
(T)Child Talk Ht.& Mk.	.31	-.38	.01	-.14	-.09	-.05	-.09	-.21	.28	.10	.06	.01	.44	.19	.28	.14	-.14	-.01	.01						
35	.19	-.73	-.35	-.58	-.30	.48	.33	.49	.77	.93	.95	.84	-.12	.74	-.15	.43	-.31	.91	.91	.54					
(T)Child Talk Total	.08	-.56	-.10	-.42	.28	.28	.33	.72	.56	.90	.96	.89	-.21	.75	-.15	.16	-.25	.97	.97	.17					
36	.39	-.52	.09	-.02	.00	.30	.44	.11	.32	.26	.33	.44	-.01	.43	.11	1.00	-.06	.43	.06	.58	.42				
(T)Child Work	-.18	-.14	.21	.16	-.04	.09	.11	.10	.10	.15	.16	.17	.22	.53	-.35	1.00	-.05	.15	.12	.14	.16				
37	.61	-.05	-.16	.19	-.02	-.06	-.03	-.19	-.01	.00	-.08	-.07	.55	-.06	.78	.13	.88	-.07	-.25	.02	-.14	.14			
(T)Child Non-Work	.20	.31	-.06	.43	-.07	-.29	-.26	-.23	-.20	-.35	-.35	-.37	.35	-.36	.31	-.04	.92	-.33	-.36	-.21	-.38	-.03			
41	-.31	-.22	.18	-.13	-.06	.21	.20	.49	.22	.25	.45	.41	-.66	.31	-.48	.15	-.59	.47	.55	.20	.53	.14	-.74		
Success Placement	-.21	-.23	-.05	-.15	.09	.14	.15	.38	.14	.21	.31	.31	-.58	.20	-.18	-.02	-.60	.37	.36	.24	.41	-.03	-.73		
44	-.02	-.47	-.24	-.16	.06	.08	.07	.06	.38	.52	.51	.39	-.22	.38	-.09	.19	-.25	.48	.49	.47	.60	.19	-.23	.44	
Success Reason	.23	-.40	-.13	-.19	.02	-.04	-.07	-.05	.44	.19	.22	.15	.07	.21	.18	-.20	-.26	.24	-.24	.33	.30	-.20	-.41	.44	
45	-.14	-.44	-.11	-.17	.02	.14	.13	.24	.38	.49	.57	.46	-.43	.42	-.26	.21	-.42	.55	.59	.44	.67	.20	-.46	.73	.94
Success Total	.09	-.39	-.12	-.20	.04	.03	.01	.12	.39	.23	.29	.24	-.16	.24	.07	-.16	.43	.33	.32	.34	.39	-.17	-.60	.73	.94

[a] The first correlation is for observer1; the second correlation is for observer 2; N = 19 or 20.

[b] See Tables 1 and 2.

[c] Variable 1 = orientation time.

In conclusion, most of the intercorrelations among the Eight-Block Sort variables selected for analysis in this study are low and fairly similar for both paraprofessional observers. Even though some of the large correlations may be misleading, since a number of the variables involved occur infrequently (see Table 3), a few suggest some interesting relationships that may exist between a mother's and/or a child's behaviors that could be explored further in analysis of a larger sample.

CHART 1

SRI EIGHT-BLOCK SORT TASK SCORING PROCEDURES MANUAL FOR OBSERVERS

Materials needed:

Score Form
Watch or clock with second hand
Several #2 pencils

General description:

The 8-Block Sort task requires the cooperation of you as an observer, your trainer who administers the task, and a mother/child team. Your efforts in the entire procedure are especially important because you will *observe and record* the interactions during:

PART I: TRAINER TEACHING MOTHER
PART II: MOTHER TEACHING CHILD
PART III: TRAINER WORKING WITH CHILD

It is your responsibility to teach your trainer how to administer the task and how to work with the child after the mother has taught her child the task. Two copies of the *Administration Manual for Trainers* are provided so you can each have a copy.

The following pages provide a step-by-step description of the use of each portion of the score form. Tally marks are used on the score form to record the verbal and non-verbal interactions that take place during the task.

Please make sure the forms are completed properly before returning them to your Site Coordinator.

SCORING

Instructions for completing top of score form cover:

For each child who should get the 8-Block Sort Task, fill in the top of the scoring sheet. Fill it in regardless of whether or not the 8-Block Sort Task was administered.

If the task was *not* given, note the reason on the line labeled "Comments". For example:

"The child moved," or "No longer in Head Start," or "Mother could not come," etc.

If the task *was* given, fill in at that time the line, "Where was task administered and under what conditions?" This information may have to do with the mother, the child, the task, or the physical arrangements under which the task was given. For example:

"In nurse's office, phone rang constantly," or "Mother brought several other children, they remained in room," or "Small office, administered task on floor," or "Mother could speak very little English."

It is better to record too much than not enough. Provide us with as complete a picture as possible.

Part I: TRAINER TEACHING MOTHER

Record the *starting time* of the mother's training session to the nearest minute.

For each section note the number of trials (from 1 to 3) needed for the mother to learn the task.

Indicate whether or not she succeeded on her final trial for each section.

We suggest that you and your trainer decide on some signal or cue to indicate, and cut down on, trainer errors. For example:

"verbal", "short", "X", "place", etc.

Please write out what the errors were, *unless* she corrects herself.

Record the *ending time* of the mother's training session to the nearest minute.

Part II: *MOTHER TEACHING CHILD*

ORIENTATION PERIOD

Record the *starting time* of the child's ORIENTATION PERIOD to the nearest minute.

The ORIENTATION PERIOD includes everything that happens from the time the mother starts teaching her child until she verbally requests the child to *place a block on the board*. As soon as the mother requests a block placement, move immediately into the TRAINING PERIOD.

The information we are asking you to tally above the double line during ORIENTATION PERIOD is the more specific information. If you feel that something could be tallied both above and below the double line, tally it above the line *only*. For example:

> "See these tall blocks?" could be tallied under HEIGHT and also under DIRECT REQUEST. Please tally it under HEIGHT because this gives us more specific information.

We are interested in how the mother familiarizes her child with the task prior to requesting the child to place blocks. Please tally *each time* the mother indicates either *verbally* or *non-verbally* (pointing, gesturing with her hands).

On the following pages there are examples of the various types of interactions between the mother and child and where they should be tallied.

MOTHER

INDICATES verbally or non-verbally:

FUTURE TASK	"We're going to play a game." "I want to teach you something."
HEIGHT	"These are all blocks." "Look at the baby blocks and the poppa blocks." "How are these alike?" (If the blocks are arranged so there is no doubt that mother is referring to size.) Any words or gestures the mother chooses to use that distinguish between different sizes are acceptable.
MARK	"These are flowers and these are cherries." "Do you know what these letters are?" Any words or gestures the mother chooses to use that distinguish between different marks are acceptable.
HEIGHT & MARK	"These are tall and have an X." "Tell me how these blocks are alike." (If the blocks are arranged in 4 groups by height and mark.) Whenever the board if set up with 4 groups or arranged in such a way that you are unable to determine whether the mother is referring to height or mark, tally under HEIGHT & MARK.

ORIENTATION PERIOD (cont'd)

CHILD

SPECIFICALLY TALKING about:

HEIGHT	"These are tall." "Big red block." "Baby block."
MARK	"Looks like a cheerio." "It's a circle." "Airplanes." "They're flowers."
HEIGHT & MARK	"Tall X." "Little flowers." "Big cheerios."

In order to score above the line on the child side of the score sheet the child must *say words that distinguish height and/or mark.*

If the child points rather than talking you should tally under RESPOND.

Do *not* tally phrases like "same size," "they're alike," under SPECIFICALLY TALKING; these remarks should be tallied below the line under RESPOND.

The categories below the double line are defined below through the use of examples. Your most immediate, simplest understanding of the category is more than likely correct. When you think that something the mother or child has said or done could be tallied in more than one category, always tally it in the more specific category only. Never double tally except under BEHAVIOR MODIFICATION. Examples for the categories below the double line are:

DIRECT REQUEST	"Is that chair high enough?" "Please come in and sit over here." "Can you see all right from there?"
RESPOND	Child sits down where told. Child points to blocks. "Blocks are all the same size." An answer to "Is that chair high enough?" or "Can you see all right from there?" should be tallied under RESPOND.
COMMENT, PLAY:	"It's hot in here." "I can't remember what I'm supposed to do next." Child begins handling the blocks or building towers, either on the board or off the board.

ORIENTATION PERIOD (cont'd)

PRAISE (of response)	"That's perfect!" "Good!" "You did that so quickly."
ACKNOWLEDGEMENT (of response)	"O.K." "That's fine." "That's right, that's an X." (Repeating words.) Nodding (non-verbal)

BEHAVIOR MODIFICATION:

One person attempts to change another person's behavior (usually the mother correcting her child).

The most often heard behavior modification is simply "No" (following an incorrect placement by the child).

Five sub-categories have been placed underneath BEHAVIOR MODIFICATION in order to give us more detailed information.

When you see or hear behavior modification, always tally it in BEHAVIOR MODIFICATION, and then, if it fits in one of the five sub-categories listed below, also tally it there.

reason:	"No, it goes here because it is little."
question:	"No, that isn't right. Don't you see those are all big?"
firm:	"STOP PLAYING WITH THE BLOCKS!" "NOW WATCH WHAT I AM DOING!"
demean, threaten:	"I don't know why you can't do it right!" "If you don't sit up and listen I'm going to spank you."
punish:	Mother shakes the child's arm (physical contact, in an unfriendly manner!)
"I DON'T KNOW" IGNORE, NO RESPONSE	Mother asks "What is this mark?" Child says "I don't know" or doesn't say anything or shrugs his shoulders. Child says "I want a drink of water." Mother goes right on with "What is this mark?"

ORIENTATION PERIOD (cont'd)

REFUSE, REJECT:	Mother says to point to the big blocks, and child says "No," or shakes his head. "I don't want to play with these blocks."
OBSERVE:	The mother sits and watches the child, but says or does nothing. The child sits and watches the mother, but says or does nothing. (If this continues for more than a few seconds, tally about every 5 seconds.)
BRIBE:	"If you do it right we'll have some ice cream when we get home."
ENCOURAGE:	"Keep trying. I know you can get it."
TASK IRRELEVANCY: (Mother only)	"These blocks are red." "Point to the square blocks."

Part II: MOTHER TEACHING CHILD

TRAINING PERIOD

The mother's first verbal request for block placement starts the training period.

Record the *starting time* of the training period to the nearest minute.

MOTHER REQUESTS PLACING AND/OR TALKING

Requests *must be verbal* to be tallied under REQUESTS PLACING or REQUESTS TALKING.

Do *not* tally pointing, gesturing, or holding a block up for the child to identify under REQUESTS PLACING or REQUESTS TALKING. Rather, they should be tallied under DIRECT REQUEST.

When the mother rewords a request for placement or talking, make one tally. For example:

> "Put this block with the other *big* ones. Put it over here with the *tall* ones." (One tally under REQUESTS PLACING by HEIGHT.)

Several requests tied together are to be tallied as one request. For example:

> "Put the little X's here, and the big O's here, and the big X's over there and the little O's over there." (However, if the mother waits after each request for the child to place the blocks, these would be treated as four separate REQUESTS FOR PLACING by HEIGHT & MARK.)

When the mother requests both placing and talking, put one tally under REQUESTS TALKING and one tally under REQUESTS PLACING. For example:

> "Tell me what this is (points to X on block) and put it with the others that have the same mark." (One tally under REQUESTS TALKING by MARK and one tally under REQUESTS PLACING by MARK.)

When the mother has the blocks arranged in such a way, or asks a question in such a way, that you aren't sure whether she is referring to size or mark, tally under HEIGHT & MARK. For example:

> Mother has groups set up in no particular pattern, and says "Put this one where it belongs." (One tally under REQUEST PLACING by HEIGHT & MARK because you can't tell where she wanted it placed.)

On the other hand, when the mother has the blocks arranged in the four groups and asks the child to tell her why the block was placed in the proper group it would have to be a REQUEST TALKING by HEIGHT & MARK. For example:

> "Because it is big and has an X on it."
> (One tally under CHILD SPECIFICALLY TALKS about HEIGHT & MARK.)

TRAINING PERIOD (cont'd)

CHILD PLACING BLOCKS

Remember that the mother's first request for block placement starts the training period; *from then on* you tally each time a child places a block on the board, *regardless of whether or not the mother has requested it.* For example:

> Mother says "Place all the big blocks on the board."
> Child picks up six blocks, one at a time, and places each one on the board.
> (Make six tallies under PLACING BLOCKS.)

When two or more blocks are placed at one time with one hand, put one tally under child PLACING BLOCKS.

When a child moves a block around before finally leaving it on the board, make one tally under PLACING BLOCKS.

If a child reponds with placement and talks *at the same time*, put one tally under PLACING BLOCKS and one tally under SPECIFICALLY TALKS.

GOOFING AROUND

When the child is playing with the blocks, building towers, etc., each time he places a block *on the board* put a tally under GOOFING AROUND. For example:

> Mother says to put a block with the others just like it and child puts it on top of the others. This is still considered a placement on the board, but in a "funny" way.

CHILD SPECIFICALLY TALKING about:

(Same as ORIENTATION PERIOD, see page 6)

Most of the categories below the double line are the same for TRAINING PERIOD as for ORIENTATION PERIOD, see pages 6 thru 8. The following are exceptions:

DIRECT REQUEST:	"Point to the X." Mother holds up a block for the child to identify but doesn't say anything. (Requests must be verbal during *training period* in order to go above the line.)
TEACH:	"There are 8 blocks here, Johnny, and they are different heights." (Any task related information that doesn't ask the child to place a block or say something about the height or mark.)
COMMENTS, PLAY:	When the child plays with the blocks *off the board*. (On the board is GOOFING AROUND.)

Part III: CHILD REQUESTED TO PLACE BLOCKS AND SAY WHY

We are interested in how the mother reacts as well as what the child says and does. So watch both and record responses in appropriate places.

The child will first be given the SHORT O block to place and say why he placed it there. Record where the child placed it under CHILD PLACED BLOCK WITH:

Watch the mother's reaction. Circle anything you note. Circle at least *one* response.

The trainer will now ask the child up to 4 questions in order to get the child to say "small O" (or any words meaning small and O). After each question write down everything the child says or does. If child moves block after being questioned, *record final location*, and again ask up to 4 questions. For example:

Trainer says "Put this block where it belongs."
Child places SHORT O with other short O blocks.
Trainer says "Why does it go there?"
Child says "It's little."
Trainer says "What else can you tell me about why it goes there?"
(You may use up to 3 probes – such as "Tell me more about why it goes there." or "What is another reason why you put it there?")

You may use the following abbreviations:

DK	"I don't know."
NR	No response
NV	Non-verbal (gesturing height, etc.)

The trainer will then repeat the above procedures for TALL X block.

Be as complete as possible. Distinguish clearly between verbal and non-verbal answers. Please indicate NV in front of all non-verbal answers so we have a complete picture of what the child *did* as well as what he said.

Circle as many of mother's reactions as appropriate (at least one). Following is a description of each code:

Rejection:	Mother blamed the child for failure or made derogatory remarks about him.
Dissatisfaction:	Mother scowled, frowned, showed impatience, but did not openly blame or accuse him.
Neutral:	Mother did not get involved. She watched the interaction between the child and trainer, but did not reveal her reactions.
Nonverbal support:	Communicated sympathy, confidence and/or support by small expressons (without saying anything).
Verbal support:	Mother praised the child if he succeeded, reassured him if he failed.
nt answered:	Mother answered for the child or gave him information or answers.

APPENDIX C

Classroom Behavior Inventory Test-Retest Reliability Study

A test-retest reliability study of the Classroom Behavior Inventory was conducted in the fall of 1971 by the Huron Institute and SRI. Teachers in four sites (Kansas City, Des Moines, Lafayette, and Greeley) rated the children in their classes on the Classroom Behavior Inventory in both the seventh and ninth weeks of the school year. The Classroom Behavior Inventory is a 15 item, seven-point rating scale, which measures three basic trait categories: task orientation, extraversion, and hostility. (For a complete test description with a list of the items, see Part II.)

Factor Analysis

A principal components analysis followed by a varimax rotation (i.e., rotating until the loadings on each factor have a maximum variance) was done on the pooled seventh and ninth week scores for the total sample (n = 464). This analysis revealed the existence of the three factors named by Schaefer in developing the test: Factor 1--Extraversion; Factor II--Hostility; Factor III--Task Orientation (See Table 1). The eigenvalues for these factors (6.614, 3.506, and 1.943) were the only ones greater than 1.000. Altogether the three factors explained 80.4% of the total variance.

TABLE 1

FACTOR LOADINGS OF ITEMS ON CLASSROOM BEHAVIOR INVENTORY AFTER A VARIMAX ROTATION

Item Number and Description	Factor I	Factor II	Factor III
1. Pays attention to what he's doing when other things are going around him.	.173	-.176	.875
2. Tries to be with another person or group of people.	.887	.011	.179
3. Gets impatient or unpleasant if he can't get what he wants when he wants it.	-.009	.905	-.219
4. Stays with a job until he finishes it.	.169	-.136	.882
5. Likes to take part in activities with others.	.826	-.100	.267
6. Slow to forgive when offended.	-.065	.690	-.154
7. Becomes very absorbed in what he is doing.	.217	-.144	.871
8. Enjoys being with others.	.875	-.107	.197
9. Stays angry for a long time after a quarrel.	-.054	.727	-.107
10. Works earnestly at his classwork. Doesn't take it lightly.	.248	-.183	.862
11. Seeks contact with others.	.922	-.012	.149
12. Complains or whines if he can't get his own way.	-.032	.911	-.192
13. Watches carefully when a teacher or classmate is showing how to do something.	.231	-.192	.847
14. Does not wait for others to approach him, but makes the first friendly move.	.824	.068	.182
15. Angry when he has to wait his turn or share with others.	-.013	.907	-.182
% of Total Variance	44.1	23.4	13.0

Test-retest Estimates

Results of test-retest reliability estimates calculated in two ways are listed in Table 2. One estimate is the correlation coefficient between the seventh and ninth week scores for each subtest. The correlation coefficients, ranging from .589 to .830, were adequate for test-retest estimates of subtests of a rating scale. The other estimate of reliability was calculated from the sums of squares of a one-way repeated measures analysis of variance design. After an analysis of variance between subject's scores with the two times as repeated measures was completed for each subtest at each site, an estimate of reliability was calculated using the following formula from Winer[1]:

$$r = \frac{\frac{1}{h}\left[MS_{unit} - MS_{within}\right]}{\frac{1}{h}\left[MS_{unit} - MS_{within}\right] + MS_{within}}$$

where unit = scores

h = number of times rated

$$MS_{unit} = \frac{SS_{scores}}{df_{scores}}$$

$$MS_{within} = \frac{SS_{time} + SS_{time \times score}}{df_{time} + df_{time \times score}}$$

---

[1]B.J. Winer. Statistical principles in experimental design (New York: McGraw-Hill, 1962).

TABLE 2

TEST-RETEST RELIABILITY COEFFICIENTS FOR THE CLASSROOM BEHAVIOR INVENTORY

	Correlation Coefficients	R's from ANOVA	Significance F time[1]
Kansas City (n = 201)			
Task Orientation	.784	.783	N.S.
Extraversion	.724	.722	N.S.
Hostility	.795	.777	.01
Des Moines (n = 81)			
Task Orientation	.806	.803	N.S.
Extraversion	.785	.769	.044
Hostility	.830	.819	N.S.
Lafayette (n = 75)			
Task Orientation	.764	.726	.002
Extraversion	.757	.760	N.S.
Hostility	.658	.651	N.S.
Greeley (n = 107)			
Task Orientation	.589	.557	N.S.
Extraversion	.710	.689	N.S.
Hostility	.591	.520	.001
Total (n = 464)			
Task Orientation	.760	.754	.014
Extraversion	.740	.737	N.S.
Hostility	.726	.704	.001

[1] F test for time effects equals $MS_{time} \div MS_{time \times units}$.

These reliability estimates, ranging from .520 to .819, were very similar to the correlation coefficients.

Several of the F tests for time effects were significant, meaning that for these scores there was a significant component of variance due to change over time. This can be attributed to changes in the actual behavior of children and/or changes in the rating "framework" used by the teachers. The reliability estimates calculated from the ANOVA design can be considered estimates of stability. These stability estimates are generally at the lower bounds of the reliability estimates.

Item and Score Characteristics

Analysis of the distribution of items and subtest scores reveals that there are ceiling and floor effects (see Tables 3, 4, 5). Scores for each subtest range from 7-35. High scores on the Task Orientation and Extraversion subtests and low scores on the Hostility subtest indicate the more "socially desirable" responses.

The distribution of scores for the Task Orientation subtest and the Extraversion subtest were negatively skewed at all sites. There was a definite ceiling effect for the Task Orientation score for both times at only one site--Kansas City. There was a potential ceiling effect for the Extraversion score at two sites--Kansas City and Lafayette.

TABLE 3

STATISTICS FOR TASK ORIENTATION SCORES[1]

	MEAN	SD	N	SKEWNESS	KURTOSIS
7 weeks					
Total	24.196	7.043	464	-0.469***	-0.729**
Kansas City					
Male	25.827	6.514	104	-0.592*	-0.513
Female	27.144	5.941	97	-0.889***	0.385
Des Moines					
Male	20.837	6.675	49	-0.053	-1.192
Female	23.344	6.509	32	-0.173	-1.208
Lafayette					
Male	21.615	7.594	39	-0.165	-1.034
Female	24.861	8.163	36	-0.841*	-0.496
Greeley					
Male	21.593	7.266	59	-0.177	-0.901
Female	23.500	6.147	48	-0.233	-0.925
9 weeks					
Total	24.735	6.344	464	-0.584***	0.121
Kansas City					
Male	26.212	6.845	104	-1.351***	2.607***
Female	27.598	5.586	97	-1.228***	1.572**
Des Moines					
Male	21.184	5.714	49	0.196	-0.396
Female	23.500	6.304	32	-0.107	-1.068
Lafayette					
Male	23.436	6.613	39	-0.366	-0.975
Female	26.917	6.403	36	-0.630	-0.845
Greeley					
Male	22.390	4.832	59	-0.520	-0.214
Female	22.500	4.776	48	-0.664	0.014

[1]Score range (7-35)

Significance levels: * .05
** .01
*** .001

TABLE 4

STATISTICS FOR EXTRAVERSION SCORES[1]

	MEAN	SD	N	SKEWNESS	KURTOSIS
7 weeks					
Total	25.015	6.594	464	-0.609***	-0.383
Kansas City					
Male	25.846	6.703	104	-0.675**	-0.377
Female	25.670	6.509	97	-0.546*	-0.647
Des Moines					
Male	23.878	6.534	49	-0.235	-0.706
Female	22.781	6.559	32	-0.919*	-0.232
Lafayette					
Male	25.564	5.619	39	-0.747	-0.404
Female	27.833	5.950	36	-1.247**	1.142
Greeley					
Male	23.424	6.605	59	-0.311	-0.555
Female	23.937	6.908	48	-0.792*	-0.127
9 weeks					
Total	25.317	6.135	464	-0.749***	0.768***
Kansas City					
Male	25.856	7.387	104	-1.115***	1.325**
Female	26.835	5.755	97	-0.799**	0.492
Des Moines					
Male	25.122	5.566	49	-0.050	-0.354
Female	23.219	5.638	32	-0.624	0.036
Lafayette					
Male	25.333	6.110	39	-0.696	-0.542
Female	27.639	4.969	36	-0.656	-0.313
Greeley					
Male	23.559	5.351	59	-0.633*	0.827
Female	23.083	5.044	48	-1.116**	1.102

[1] Score range (7-35)

Significance levels: * .05
** .01
*** .001

TABLE 5

STATISTICS FOR HOSTILITY SCORES[1]

	MEAN	SD	N	SKEWNESS	KURTOSIS
7 weeks					
Total	12.765	6.506	464	1.201***	0.947***
Kansas City					
Male	13.077	6.344	104	1.179***	0.817
Female	11.536	5.483	97	1.173***	1.293**
Des Moines					
Male	14.367	7.126	49	0.873*	-0.054
Female	12.531	6.877	32	1.463***	2.026*
Lafayette					
Male	11.487	5.703	39	0.909*	-0.081
Female	12.500	7.225	36	1.228**	1.197
Greeley					
Male	14.203	7.712	59	0.959**	-0.385
Female	12.562	5.975	48	1.498***	2.189**
9 weeks					
Total	11.366	6.111	464	1.343***	1.882***
Kansas City					
Male	12.010	7.044	104	1.203***	1.296**
Female	10.113	5.037	97	1.455***	1.892***
Des Moines					
Male	12.878	6.382	49	1.207***	1.348
Female	12.875	5.923	32	0.906*	0.528
Lafayette					
Male	10.000	5.740	39	2.226***	6.232***
Female	11.889	7.230	36	1.283**	0.916
Greeley					
Male	11.051	5.856	59	1.216***	1.375*
Female	11.062	4.965	48	0.762*	0.591

[1]Score range (7-35)

Significance levels: * .05
** .01
*** .001

The distribution of the Hostility subtest scores were always positively skewed. There was a floor effect for these scores at both times at every site. The skewness statistic for males and females at each site for both times was significant. In addition, the median score for each item in the Hostility subtest was 2 (item scale 1 to 7) for each site at each time.

APPENDIX D

Coding Reliability Study

A reliability study was done in fall '71 at Stanford Research Institute on the coding of the Brown IDS Self-Concept Referents Test, the ETS Enumeration Test, and the ITPA Verbal Expression Subtest. Twenty of each test were picked at random and were coded independently by each of the three coders working on the test.

Brown IDS Self-Concept Referents Test: The Brown is potentially difficult to code. The tester is asked to spot code only answers which are verbatim repetitions of one of the choices given in the test. The tester is asked to record other responses, and to indicate repeats. The coders are then responsible for judging whether responses are exact equivalents or not. They are also asked to code the number of repeats, and whether or not the child is smiling.

To check coder reliability, we calculated the number of times each pair of coders disagreed on coding responses, on the number of repeats, and on whether the child was smiling or not. On responses to the questions in the body of the test, there were 16 responses per child, or 320 in all. Coder 1 and 2 disagreed twice; coders 2 and 3 did not disagree, and coders 1 and 3 disagreed twice.

Thus, the percent agreement of the three coders (C1, C2, C3) is as follows:

C1/C2 = 99.4%

C2/C3 = 100.0%

C1/C3 = 99.4%

On repeats, the coders were in exact agreement on the number of repeats the following percentages of the time:

C1/C2 = 13/20 = 65%

C2/C3 = 16/20 = 80%

C1/C3 = 14/20 = 70%

Perfect agreement is a very demanding test. A simple percent also gives no indication of whether the discrepencies were large or small. As another measure of agreement, therefore, one can look at the correlation coefficients between the pairs of coders.

$r_{12} = .966$

$r_{23} = .993$

$r_{13} = .961$

For smiling, there were three possible codes: smiling, not smiling and indeterminant. Coders were in agreement on the following percentages of the tests:

C1/C2 = 15/20 = 75%

C2/C3 = 17/20 = 85%

C1/C3 = 15/20 = 75%

ETS Enumeration Test: On the ETS Enumeration test, coders are responsible for combining the information recorded by the testers into a series of codes expressing both the answers which the child gave and the types of errors which he made. We looked at coding reliability for three sections of the test:

Part A (top) in which the child counts.

Part A (bottom) in which the child tells how many dots there are.

Part B in which the child simply points to data.

On part A (top) there are four questions or 80 responses in all which must be coded. Agreement among the three coders was as follows:

C1/C2 = 77/80 = .96

C2/C3 = 76/80 = .95

C1/C3 = 75/80 = .94

On Part A (bottom) there are also four responses per child. Agreement between coders was as follows:

C1/C2 = 78/80 = .98

C2/C3 = 77/80 = .96

C1/C3 = 77/80 = .96

On Part B, there were six items per child. Agreement among coders was as follows:

C1/C2 = 119/120 = .99

C2/C3 = 119/120 = .99

C1/C3 = 119/120 = .99

ITPA Verbal Expression Subtest: The ITPA Verbal Expression Subtest is the most difficult test in the battery to code. The child is asked to tell "all about" four different objects. The tester records the child's response verbatim. The coder is responsible for transforming this information into a quantitative description of the child's output in a number of categories: name, color, shape, use, etc. The coders use 10 categories for each object, or 40 in all. One measure of agreement, therefore, is the % of categories for which the coders achieve perfect agreement. These percentages for three coders are as follows:

C1/C2 = 784/800 = 98%

C2/C3 = 776/800 = 97%

C1/C3 = 784/800 = 98%

These percentages are deceptively high, however, since well over half of the categories on each protocol are blank (For the twenty protocols in our coding reliability sample, the numbers of categories left blank by all three coders ranged from 26 - 35.)

A second, and more useful, measure of agreement is the correlation between the total numbers of scored responses recorded by each coder:

$$r_{12} = .965$$

$$r_{23} = .968$$

$$r_{13} = .957$$

Structure measures, such as a measure of item by item perfect agreement could, of course, be devised. Nonetheless, the coding reliabilities which we have presented in this section indicate that coding errors have negligible effect on the quality of the data used in the analyses.

APPENDIX E

Classroom Information Form Reliability Study

Demographic data for the HSPV analysis is obtained from the Classroom Information Form (CIF). This form is filled out by Head Start teachers, often from the application forms filled out by parents. For each child, the teacher is asked to list the education, occupation and employment status of both parents; the number of adults and children in the home; the language spoken in the home; and whether the child has had previous Head Start experience.

In order to check on the reliability of this data, a comparison was made of the responses of parents and the responses of teachers on the CIF. Mothers of one-third of the Head Start children in the fall of 1971 were given the Eight-Block Sort Task and filled out a Parent Information Form (PIF). In the fall of 1971, the PIF was designed to elicit demographic data as well as attitudes and participation data. This data was used to check the reliability of the CIF data collected at the same time.

There are two issues that are examined in the analysis of the CIF and the PIF data: 1) the percent response and 2) the level of agreement between the PIF and CIF. The percent response is important because it constitutes an upper limit to the level of agreement. One can have very high agreement (for those cases reported on the CIF), but have a very low response rate. In this situation the

"high agreement" is spurious, and is considerably deflated if we consider those people for whom no information was supplied. In general, there is a higher level of non-response for the CIF than for the PIF. The percent response on the PIF must be viewed as a base line for the percent response on the CIF. Our analysis assumes that the correct response is the one given by the parent. There are several obvious reasons why this may not be true. Although this is in all likelihood a minor contribution to the "noise," it is probably a statistically significant one.

Site Variations:

Although some sites seem to be worse than other sites on particular questions, no one site was uniformly poor. We initially examined both percent response and percent perfect agreement (as well as a series of measures of association) for each of the questions for each site. We then analyzed all of the sites as a group. Next, we deleted what appeared to be the worst six sites, and analyzed the remainder as a group. In general, on these aggregate analyses few differences were found. We interpret this as indicating a homogeneity of response pattern across sites.*

---

* Our convention in reporting information is as follows: Percent perfect agreement is based on all of the data including missing responses. Any statistic reported (e.g., Pearson r, Kendall's Tau) is based upon only those subjects for which a response is recorded on both the PIF and CIF. Thus, in interpreting such statistics, one must also consider the percent response.

Preschool Experience:

Across all sites, there was 79.9% perfect agreement. The response rate was quite good: 97% CIF vs. 99% PIF. Across each site, the percent perfect agreement ranges from 62% to 100% with 15 out of the 23 sites above 86%. On both percent response and percent agreement this is one of the best questions on the CIF. In terms of those who were misclassified, children were more often (about 2 times) classified as having had preschool experience when in fact they had not, than classified in any other combination.

Mother's Education:

Of all the Education and Occupation questions, this had the best results. Across all sites, the response rate was 81.6% for the CIF vs. 99.1% for the PIF. Perfect agreement was 52%. The Pearson r reliability coefficient was .77 and Kendall's Tau was .76.

Looking at sites individually, the response rate ranges from 0 to 100% with 18 out of 23 sites above 83%. The reliability coefficients range from .20 to .98 with 14 out 23 sites above .77.

Father's Education:

Across all sites, the response rate was 51.1% for the CIF vs. 89.5% for the PIF. This is a rather large differential. Perfect agreement was 26%. The Pearson r was .77 and Kendall's Tau was .67.

Looking at sites individually, the percent response ranges from 3% to 86% with 13 out of 23 sites below 52%. For this data, the agreement is fair but the response rate is poor.

Mother's Occupation:

Across all sites, the response rate was 83.2% for the CIF vs. 83.7% for the PIF. Of those missing a response on the PIF, 64% were coded (12 housewife) on the CIF. Of those missing data on the CIF, 66% were coded (15 unemployed) on the PIF. This may indicate a confusion in instructions in the actual recording of the information. Perfect agreement was 24%. The Pearson r was .59; Kendall's Tau was .51.

Looking at sites individually, the percent response ranges from 40% to 97% with 14 out of 23 above 83%. Reliability coefficients range from .03 to .83 with 10 of 23 above .60.

In this situation the percent response is perhaps adequate, but the agreement is not very good.

Mother's Status:

We have reason to believe that considerable confusion existed on the part of parents with regard to the "Status" questions which asked whether the respondant was employed full time, part-time or seasonally or unemployed. For example, on Mother's Status the percent response was lower on the PIF (76%) than on the CIF (92%). Looking at sites individually, it was not unusual to see a high response on occupation and

a low response on status. Percent perfect agreement was 54% which is low considering there are only 4 legitimate responses. Cramer's V was .50 and Kendall's Tau was .48.

Father's Occupation:

Across all sites, the percent response on the CIF was 58% vs. 65% on the PIF. It appears from the PIF information that parents are reluctant to give information on Father's Occupation; either that or there was difficulty comprehending the question. Out of the 384 lacking a response on the CIF, 234 also lack a response on the PIF. It is possible that some of these households had female heads since there was no specific identification of fatherless families.

Perfect agreement was 60% The Pearson r was .53 and Kendall's Tau was .58. The relatively high percent agreement is due to the large percent (23%) missing a response on both PIF and the CIF.

Looking at sites individually, the percent response ranges from 34% to 86% with 10 of 23 above 58%.

Father's Status:

This question has the same problem as Mother's Status. Response rates were 68% PIF vs. 61% CIF. Perfect agreement was 63%, but 23% of this is in a double Non-Response. Cramer's V was .47, and Kendall's Tau was .47.

Across sites individually, the percent response ranged

from 34% to 84% with 11 of 23 above 61%.

Persons Under 18 = Persons over 18:

Both of these questions enjoy a healthy response rate on the CIF: 98.7% - Under 18; 96.9% - Over 18. This response rate is uniform across sites.

	Persons Under 18	Persons Over 18
% Perfect Agreement	70%	64%
Kendall's Tau	.80	.56
Range % Response	90-99%	80-98%
Range % Agreement	40-85%	35-80%

There does appear to be one possible confusion on the question "Persons 18+ over in the House". There is an 11% non-response on the PIF with 90% of this in classes (1) and (2) for the CIF. There may have been some confusion on the PIF as to whether parents were supposed to count themselves in this category. Other than this, the results on the PIF are much like the results of the CIF.

Language Spoken in the Home:

Across sites, the response rate is 99%, which is quite uniform. In general, there is very high perfect agreement (94%). This is because most families (94%) speak English in the home. In almost every case where Spanish was spoken in the home, it was not recorded on the CIF. The

CIF missed 75% of these cases. At one site which had a very large non-English speaking (in the home) population, the percent perfect agreement was only 22%.

APPENDIX F

Parent Information Form Test-retest Reliability Study

A small test-retest reliability study of the Parent Information Form was done in Kansas City in the fall of 1971. Eighteen mothers filled out the same form of the PIF about a month after they had first filled it out.

The form of the PIF used in the fall of 1971 was a shorter form than that used in other years. It contained forty items, most of which were demographic. The areas covered by the form are:

1. Previous Head Start experience
2. Sesame Street viewing
3. Toys and materials in the home
4. Reading in the home
5. Parental aspirations and expectations for child
6. Perinatal experiences
7. Mother's education, employment status and occupation
8. Father's education, employment status and occupation
9. Number of people in the household
10. Language spoken in the home
11. Home items

Table A shows the distributions of answers for all the questions which had only three possible answers: yes, no and don't know. The number of questionnaires for which

TABLE A

DISTRIBUTION OF PARENT INFORMATION FORM (PIF) ITEMS WITH ONLY THREE POSSIBLE RESPONSES FOR 18 MOTHERS IN A TEST RETEST RELIABILITY STUDY

	both yes	both no	1 yes 1 no	% agree	both blank	one blank
1. Pre-school	3	13	2	88.9	0	0
2. Sesame Street	5	10	2	88.2	0	1
5. Toys						
A. Blackboard	7	6	1	92.9	0	4
B. Chalk	7	4	2	84.6	0	5
C. Colored Paper	10	1	2	84.6	0	5
D. Scissors	12	0	3	80.0	0	3
E. Crayons	16	0	1	94.1	0	1
F. Color Books	17	0	0	100.0	1	0
G. Paints	5	4	3	75.0	2	4
H. Clay	1	8	0	100.0	1	8
I. Other Art	4	3	1	87.5	3	7
J. Musical Inst.	6	4	3	75.0	2	3
K. Alphabet & Number Cards	7	5	3	80.0	1	2
L. Games	7	3	4	71.4	2	2
M. Puzzles	8	5	0	100.0	2	3
N. Records	8	3	1	91.7	1	5
6. Ever Read	16	1	1	94.1	0	1
13. Born on Time	13	2	2	88.2	0	1
14. More than month early	0	1	1	50.0	12	4
15. Birth Complications	1	13	3	82.4	0	1

TABLE A

(con't)

	both yes	both no	1 yes 1 no	% agree	both blank	one blank
16. OK first week	17	0	0	100.0	0	1
19. Mother going to school	1	16	1	94.4	0	0
22. Father going to school	0	14	1	93.3	1	2
24. Mother paying job	2	12	2	88.5	0	2
27. Mother looking for job	0	12	3	80.0	0	3
28. Mother had job during year	4	11	2	88.2	0	1
30. Father paying job	13	2	0	100.0	3	0
33. Father looking for job	1	12	1	92.9	4	0
34. Father had job during year	2	10	1	92.3	4	1
39. Other languages	4	12	2	88.9	0	0
40. Home Items						
A. Auto	14	1	1	93.8	1	1
B. TV	13	1	1	93.3	1	2
C. Color TV	7	5	0	100.0	2	4
D. Encyclopedia	6	5	1	91.7	1	5
E. Dictionary	13	2	1	93.8	0	2
F. Washer	10	5	1	93.8	1	1
G. Vacuum Cleaner	11	5	1	94.1	1	0
H. Record Player	15	1	1	94.1	0	1
I. Telephone	16	1	1	94.4	0	0

the response was "yes" for both time 1 and time 2 is given in the first column; the number of both "no" responses in the second, and so on. The percent agreement in the fourth column does not include blanks or don't knows; it is equal to the sum of columns one and two divided by the sum of one, two and three. The percent agreement for those questions which have an adequate response rate is quite high.

Table B shows the distribution of answers for questions which had more than three possible responses. Questions 10, 11, 12 and 17 required the mother to write in an answer. The other questions were multiple-choice, with more than two categories. Question 10, which asked what the parent thought might prevent her child from getting the education he wanted, had both an adequate response and high (83%) agreement. Question 31, whether the father was working full or part-time, also had good agreement. Question 7, on how often the parent read to her child, had good response, but only moderate (62%) agreement. The other questions--on Sesame Street, birth weight, where the child and the mother grew up--seem almost worthless. The low response rates on type of school attended and on mother's job status are obviously explained by the answers to questions 19, 22, 24 and 27. Almost no parent went to school; almost no mother worked.

TABLE B

DISTRIBUTION OF PIF ITEMS WITH MORE THAN THREE POSSIBLE RESPONSES FOR 18 MOTHERS IN A TEST-RETEST RELIABILITY STUDY

	Same Answer	Different Answer	Both DK or Blank	One Blank
3. How often child watches Sesame St.	2	4	10	2
4. How often watch with him	4	3	6	5
7. How often read	10	6	0	2
10. Prevent schooling	15	3	0	0
11. Where child grew up	2	1	10	5
12. Birth weight	13	5	0	0
17. Where mother grew up	0	1	13	4
20. Mother kind of school	1	0	16	1
23. Father kind of school	0	0	17	1
25. Mother full/part time	2	0	13	3
29. Mother why change jobs	3	1	12	2
31. Father full/part time	10	1	4	3
35. Father why change jobs	2	0	15	1

Table C shows the distribution of responses on ten of the more important demographic variables. The questions on mother's and father's occupations had low response rates. It is difficult to judge the reliability of the responses. The data on the number of people in the home seems quite good. Discrepancies in response to this question may, of course, result from real changes in the composition of the household. The data on educational aspirations is moderately consistent (72%) as is the data on educational expectations (67% agreement). The data on mother's and father's education appears to be excellent.

These findings indicate that the PIF is probably a reliable instrument for gathering demographic data of the sort elicited in the fall 1971 short form, especially when questions are asked in a simple yes/no format. Even this modest conclusion must be treated with some scepticism, however, since the reliability study sample was so small. There is no information on the reliability of attitude and participation items of the sort used in other forms of the PIF. The findings on the moderate consistency of the educational aspirations and expectations questions and on the low response rates for many questions make us somewhat dubious about the possibility of gathering good attitude data.

TABLE C

DISTRIBUTION OF SELECTED PIF RESPONSES FOR 18 MOTHERS IN A TEST-RETEST RELIABILITY STUDY

		ASPIRATIONS			EXPECTIONS		
		Time 1	Time 2	# the Same	Time 1	Time 2	# the Same
8 Educational Aspirations	9	0	1	0	0	2	0
9 Educational Expectations	10	0	0	0	0	0	0
	11	0	0	0	0	0	0
	12	1	1	1	1	9	8
	1	0	0	0	1	0	0
	2	1	1	1	0	1	0
	3	0	0	0	0	0	0
	4	8	7	5	5	4	3
	4+	8	8	6	1	2	1
	Blank DK						

TABLE C

(con't)

		MOTHER			FATHER		
		Time 1	Time 2	# the Same	Time 1	Time 2	# the Same
18 Mother's Education	1	0	0	0	0	0	0
21 Father's Education	2	0	0	0	0	0	0
	3	0	0	0	2	2	2
	4	0	0	0	0	0	0
	5	0	0	0	1	1	1
	6	1	1	1	0	0	0
	7	1	1	1	0	0	0
	8	1	1	1	1	1	1
	9	0	0	0	0	0	0
	10	1	1	1	2	2	2
	11	1	1	1	0	0	0
	12	8	8	8	6	6	6
	1	3	3	3	1	0	0
	2	1	1	1	2	3	2
	3	0	0	0	0	0	0
	4	0	0	0	0	0	0
	4+	1	1	1	0	0	0
	Blank DK	0	0	0	3	3	3

TABLE C

(con't)

		MOTHER			FATHER		
		Time 1	Time 2	# the Same	Time 1	Time 2	# the Same
26 Mother's Occupation*	1	1	1	1	0	0	0
32 Father's Occupation	2	0	0	0	0	0	0
	3	0	0	0	0	0	0
	4	0	0	0	0	0	0
	5	0	0	0	1	1	1
	6	0	0	0	2	3	2
	7	0	0	0	0	0	0
	8	1	3	1	3	3	3
	9	0	0	0	0	0	0
	10	0	0	0	0	1	0
	11	1	0	0	5	4	4
	12	0	0	0	0	0	0
	13	0	0	0	0	0	0
	14	0	0	0	0	0	0
	15	0	0	0	0	0	0
	16	0	0	0	0	0	0
Blank	DK	15	14	13	7	6	6

*

Occupations were codes using adaptations of census categories. Category 1 includes professional and technical; category 11 includes laborers; 12 = housewife; 13 = disabled; 14 = student; 15 = unemployed, retired; 16 = no spouse. It should be noted that question 26 was only answered by those mothers who had a paying job.

TABLE C

(con't)

		UNDER 18			OVER 18		
		Time 1	Time 2	# the Same	Time 1	Time 2	# the Same
36 # under 18	1	1	1	1	3	3	3
37 # over 18	2	6	7	6	12	12	11
	3	2	3	2	0	1	0
	4	3	2	2	1	0	0
	5	3	2	2	0	0	0
	6	1	1	1	0	0	0
	7	0	0	0	0	0	0
	8	2	1	1	0	0	0
	8+	0	1	0	0	0	0
	Blank DK	0	0	0	2	2	1

38 Language in home

Both English.....13

Both English, also both Spanish.....2

Both English, Spanish once.....2

One Spanish, one English.....1

APPENDIX G

Quality of the Testing Procedures

Testing procedures changed considerably during the three years of the HSPV evaluation, as a result of changes in the test battery and improvements in procedures. This section will describe training and monitoring procedures used in 1970-71 and 1971-72.[1] It will also report the findings of an independent monitoring procedure carried out in Spring 1972, which provides the best data we have on the quality of the testing procedures.

Organization of Testing

Testing in both years was done by local paraprofessionals, under the supervision of a local site coordinator. The site coordinator was selected by the Head Start director in each site, with the approval of SRI. The site coordinator, often in consultation with the Head Start director, hired a sufficient number of local testers to complete testing within the allotted three-week period. During 1970-71, at least one tester per site was a trained Binet tester. Testers in 1971-72 were not required to have special qualifications.

---

[1] For a more extensive discussion of testing procedures for all three years of the HSPV study, see Implementation of Head Start Planned Variation Testing and Data Collection Effort. Menlo Park, California: Stanford Research Institute, 1972.

The tests were arranged into "batteries," each of which constituted a single testing session for a child. Each tester was trained in one of the batteries. The organization of the batteries and the number of testing personnel is shown below:

1970-71, fall and spring	
Site coordinators	41
Binet testers	44
Auxiliary battery testers	68
NYU Booklets	
PSI	
Motor Inhibition	
EIQ, CCAS (19 sites)	
Eight - Block observers (spring only)	41 (approx.)
Eight - Block trainers (spring only)	41 (approx.)
1971, fall	
Site coordinators	33
Clerical assistants	40
Basic battery testers:	84
PPVT	
PSI	
WRAT	

Brown testers*	9
Supplementary battery testers:	35
Enumeration	
ITPA	
Motor Inhibition	
Eight - Block observers	39
Eight - Block trainers	40
1972, spring	
Site coordinators	35
Clerical assistants	38
Basic battery A testers:	85
Gumpgoohies	
WRAT	
Basic battery B testers	85
PPVT	
PSI	
Enumeration	
Supplementary battery testers	34
Relevant Redundant Cues	
Motor Inhibition	
ITPA	

*In most sites Brown testing was done by the supplementary battery tester or the basic battery tester.

1972, spring (con't)

Eight - Block observers	17
Eight - Block trainers	17

Training & Monitoring

1970 - 71: All site coordinators were trained at SRI during one 4-day session in August. Training was conducted by SRI Personnel. Auxiliary battery testers were trained on site by the site coordinators, immediately preceding the beginning of testing. Binet testers were trained in the use of the Hertzig-Birch scoring and in HSPV evaluation procedures by SRI personnel in full-day sessions at two separate locations. Eight - Block observers were trained by SRI personnel in three locations. Eight - Block observers trained Eight - Block trainers on site. Since the spring battery was the same as the fall, except for the Eight - Block sort, training of site coordinators was not repeated in the spring. Site coordinators apparently gave refresher training and training for new testers as needed.

Fall 1971: Training procedures for fall 1971 were basically the same as those for 1970-71. Site coordinators were trained during one 5-day session at SRI. Site coordinators trained basic battery testers,

supplementary battery testers and Brown testers on site. Eight - Block observers were trained in three groups by SRI Personnel during 3-day training sessions. Eight Block observers trained Eight - Block trainers on site.

Spring 1972: Additional training was given in spring 1972, since new tests were added to the spring battery, and since special efforts were made to insure that the last HSPV testing would be of high quality. Site coordinators were trained in two groups during five-day sessions at SRI. These longer sessions included practice testing of children and written tests on procedures. Site coordinators then trained basic battery, supplementary battery and Brown testers.

Monitoring: The procedures described above imply that training for most of the tests takes place on three levels: SRI personnel, site coordinators and local testers. The probability that procedural deviations will occur at one point or other along the line is higher, therefore, than it would be if all testers were trained together. There is also a possibility that site biases might be introduced, since all testers at a site are trained by one person. Monitoring of testing is necessary to ensure that these potential biases are not introduced. During 1970-71, SRI personnel visited the sites during the first two days of testing, to help

with final training and to check that standard testing procedures were being used. During 1971-72 independent observers as well as SRI personnel participated in monitoring testing. Their spring reports provide the basis for the following discussion of testing quality.

Independent Assessment of Testing Quality

Five independent observers visited fifteen sites during spring 1972 testing. The five observers were hired by OCD, independent of SRI. All were specialists in both child development and testing. They were instructed to observe the actual testing situations, placing themselves so that they could hear the tester and child and observe the tester's codings. They recorded all instances of coaching, procedual and coding errors which they observed.

The numbers of testers and test administrations which were observed are:

Test battery	Number of testers	Number of children
Basic battery A	28	52
Basic battery B	29	39
Supplementary battery	11	20
	68	111

The observers were asked to record their general impressions of site and tester biases, in addition to

recording specific errors. Their records give the impression that testing was generally of extremely high quality.

Of the fifteen sites visited, only three elicited negative general comments from the observers:

Houston: "site supervisor seemed uncommitted to quality training and supervision" (but reports on testers were all good).

Greeley: "a few of the testers acted overtly hostile to minority children."

Des Moines: "more little errors in most of the testers than in (other sites which the observer visited) . . . the lack of what really constitutes an acceptable probe is certainly a site bias here."

All other site comments were positive. For example:

Fort Walton: "testers are well trained and have good rapport techniques. Generally . . . data will be all valid."

Loch Haven: "the testers seemed excellently prepared."

Jonesboro: "all the testers demonstrated the same high attention to detail that is so crucial to getting good data."

Bellows Falls: "the testers were very capable . . . all testers seemed to have a good grasp of what they were doing."

The observers recorded general negative comments about only eight of the sixty-eight testers whom they observed.[1] These eight were the only testers who the observers felt might be eliciting a biased or invalid test score. All other testers were considered to be well within acceptable limits. Many received rave reviews.

The observers were instructed to record each error they observed in test administration. This is a demanding instruction, and some of the observers were more perfectionist than others. Nonetheless, their reports indicate that they did record every error which seemed important. (Unimportant errors were less consistently reported. One observer, for example, recorded a procedural error when the PSI checkers were not precisely spaced. Others did not seem to be quite this precise).

Table 1 shows the frequencies of recorded errors, by test and type of error (coaching, procedural, coding, timing).

---

[1] Their comments were
--"very directive and brusque...depressed children's scores."
--"messed up on scoring the PSI...a lot of probing."
--"upon first observation is an inadequate tester...quite unclear about when to repeat items and when or how to prompt..."
--"seems on the hostile side and makes almost no attempt to be pleasant."
--"horrible testing conditions...like a goddamned zoo."

TABLE 1

FREQUENCY OF RECORDED ERRORS, BY TEST AND TYPE OF ERROR, MADE BY FIVE INDEPENDENT OBSERVERS ABOUT SPRING 1972 DATA COLLECTION

	Coaching	Procedural	Coding	Timing	Start	Stop
PPVT (39)[1]	1	0	3	----	5	3
PSI (39)	4	18	16	---		
Enumeration (39)	9	5	6	---		
Gumpgoohies (52)	0	1	0	---		
WRAT (52)	1	9	8	2		
RRC (20)	0	2	1	---		
MI (20)	0	0	0	2		
ITPA (20)	1	1	3	---		

[1]Number of observed test administrations is in parentheses.

A coaching error would be a verbal or non-verbal cue given by the tester to the child to indicate his response was satisfactory or not (i.e., allowing a child a second chance on an item, consistently smiling at a child's correct responses, etc.) A procedural error would be a deviation from instructions for the test (i.e., pointing to a correct gumpgookie when reading a response, starting incorrectly on the PPVT, proceeding to the test without giving the adequate pretest tasks on the MI, etc.) A coding error would be a mistake made in recording the child's response (i.e., not circling "V" for a verbal response on the PSI, etc.) A timing error would be a mistake in recording the start or finish of a test or an error in time given for some specifically timed items, such as those on the WRAT. The number of errors recorded is probably larger than they would be if the observers had been instructed to consistently record "important errors" They are smaller than if each observer had recorded tiny procedural deviations.)

Errors were recorded by item. Considering the large number of items on the various tests, the number of errors is quite low. On only two tests, (Enumeration and the PSI) does the total number of errors average more than .5 per test. Many of the errors would not be expected to influence a student's score at all. No systematic baises were in evidence. All in all, we can be quite confident that the general level of testing competence in spring 1972 was high.