DOCUMENT RESUME

ED 082 856 PS 006 948

AUTHOR Walker, Debbie Klein; And Others
TITLE The Quality of the Head Start Planned Variation Data. Volume I.
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 273p.; For Volume II, see PS 006 949

EDRS PRICE MF-$0.65 HC-$9.87
DESCRIPTORS Achievement Tests; 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, the first of two volumes, describes the cognitive, psychomotor, and socioemotional measures used in all years of the Head Start Planned Variation Evaluation. Part I discusses generally the issues involved in evaluating the quality of the data, and summarizes findings. Part II contains technical reports on 12 of the individual measures used in the evaluation, such as the Preschool Inventory and the Stanford-Binet Intelligence Test (which were considered especially useful), the Classroom Behavior Inventory, and the Brown IDS Self-Concept Referents Test. There is a description of each measure and the theory behind it, as well as a review of the available data on the measure's reliability, validity and other technical qualities. Judgments on the usefulness of the measures, and a summary of the information on which the judgments were based are presented. Data tables are included. (SET)

THE QUALITY OF THE HEAD START PLANNED VARIATION DATA

DEBBIE KLEIN WALKER
MARY JO BANE
ANTHONY BRYK

VOLUME I
(pages 1-269)

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

INTRODUCTION

This document describes the child outcome and background measures used in all years of the Head Start Planned Variation Evaluation. Implementation measures are described in other Huron Institute reports. The purpose is to enable readers to assess the quality of the data on which the analyses of the HSPV Study and similar studies and evaluations involving young children are based. We offer our own judgments on the usefulness of the measures, and have also summarized much of the information upon which our judgments were based.

The document has three parts. Part I is a general discussion of the issues involved in evaluating the quality of the data, and a summary of findings. Part II contains technical reports of the individual measures used in the evaluation. For each measure, there is a description of the measure and the theory behind it, as well as a review of the available data on reliability, validity and other technical qualities. Part III describes the procedures used during 1971-72 in the Huron Institute's examination of the quality of the data.

We wish to thank all those people at the Office of Child Development and at Stanford Research Institute who have helped in the development of this technical report on the quality of the data. We are especially appreciative of the guidance received from Thelma Zener, Lois-ellin Datta, and Esther Kresh.

PART I: GENERAL ISSUES

The general question raised in this section is how well the goals of the various sponsors participating in the Head Start Planned Variation experiment are assessed by the measures and procedures used in the evaluation. To answer this question, the measures used will be described and categorized according to the characteristics they purport to measure. This allows for an assessment of the scope of the battery on the assumption that the instruments measure what they say they do. The reliability of the instruments will then be looked at to see how much confidence we can have that stable unitary characteristics are being measured. Next, concurrent and face validity is examined in an attempt to arrive at more precise definitions of the characteristics the instruments really measure. Predictive validity is then discussed as a way of making some inferences about the importance of these characteristics. Next, background measures are briefly discussed. A summary including suggestions for future work in measurement of young children, especially in large-scale evaluations, is included at the end of this section.

Like most evaluators of programs for young children, we conclude that all that can be measured with any confidence are short-term cognitive gains. Like most evaluators, we lament this situation. We question the importance of short-term cognitive gains, since all the data seem to indicate that they have few long-term effects. The most important effects of Head Start may be in areas such as social and emotional growth where there are no adequate measures for use in a

large scale evaluation.

The HSPV evaluators devoted considerable effort to a search for adequate non-cognitive measures. That they were unable to find them results not from a lack of will or effort, but from the lack of adequate theory and technology in the field (see Walker, 1973).

The Measures

Selection of instruments. Several criteria were important in selecting child outcome measures for the HSPV battery. The first was that the battery as a whole reflect the various goals of the several sponsors. Since some sponsors concentrate on general intellectual development, some on specific academic skills and some on social and emotional development, all these areas were to be included in the battery. A second criterion was that the instruments selected be usable on a large scale with paraprofessional testers. The conditions under which Head Start testing was carried out demanded that tests be relatively easy to administer in a standard way. A third criterion, to be applied when possible, was that tests were to be selected which had gone through a fairly lengthy development process and had been used in other studies. Finally, wherever possible, tests were to be the same as those used in previous years and in the Follow Through evaluation. The process for test selection involved the Office of Child Development,

the Planned Variation sponsors, Stanford Research Institute and outside consultants in the field of child development. In the last year of the study the Huron Institute also participated in test selection.

This entire document speaks to the issue of how well these criteria were met. It should be noted at the outset, however, that they are very demanding. Although there is no dearth of instruments which have been used with young children,[1] there is a paucity of generally respected and widely used instruments in many areas. Consequently, although the test selection procedure was as thorough and careful as could be expected, the battery is very uneven in the extent to which various goals are measured.

Instrument descriptions. The following list provides brief descriptions of the instruments used in the HSPV evaluation from 1969-1972. They are described in detail in Part II of this document. Various categorizations of the measures -- by year of use, type of instrument, area measured -- appear later. All measures are individually administered:[2]

---

[1]For example, see the listings in Buros (1972), Educational Testing Service (1968), Walker (1973), and White et al., (1972).

[2]Copies of the tests and manuals in the forms used in the HSPV evaluation are available from the ERIC Clearinghouse for Tests, Measurement and Evaluation, Educational Testing Service, Princeton, N. J. 08540. References and full descriptions of the tests appear in Part II.

Brown IDS Self-Concept Referents Test (Brown): a measure of the child's self-concept.

California Preschool Social Competency Scale (CPSCS): a teacher rating scale of "preschool children's interpersonal behavior and the degree to which they assume social responsibility."

Classroom Behavior Inventory (CBI): a rating scale assessing social behavior in three areas: task orientation, extroversion and hostility.

Classroom Information Form (CIF): a summary form providing demographic information on the children in the sample.

Classroom Observation Instrument (COI): a measure of classroom interaction and activity patterns, by classroom and by individual.

ETS Enumeration Test: a test of three components of the enumeration process: counting, pointing and matching.

Eight-Block Sort Task: a measure of maternal teaching style and of mother-child interaction.

Ethnic Identity Questionnaire (EIQ); Children's Cultural Awareness Scale (CCAS): measures of the degree of awareness of racial identity.

Gumpgookies: a semi-projective test of motivation to achieve in school.

Hertzig-Birch Scoring: a measure of a child's style of responding to the cognitive demands of a testing situation.

Illinois Test of Psycholinguistic Abilities--Verbal Expression Subtest (ITPA): a measure of the child's ability to express himself verbally.

Motor Inhibition Test (MI): a measure of the child's ability to inhibit movement when requested.

NYU Booklet 3-D: an achievement test of relational, pre-math, pre-science and linguistic concepts.

NYU Booklet 4-A: an achievement test of knowledge of numbers, letters and shapes.

Parent Information Form (PIF): a measure of demographic information, parental attitudes and parental participation.

Peabody Picture Vocabulary Test (PPVT): a test of the receptive vocabulary component of verbal intelligence.

Preschool Inventory (PSI): a test of "achievement in areas regarded as necessary for success in school."

Relevant Redundant Cue Concept Acquisition Test (RRC): a measure of concept acquisition, learning ability and attention.

Stanford-Binet Intelligence Test: a measure of general intelligence.

Wide Range Achievement Test (WRAT): an achievement test of skills in reading, spelling and arithmetic.

Continuity. Table 1 lists the measures by the year in which they were given. Tables 2 and 3 indicate the sites and the proportion of children within each site who were tested on each measure. These tables summarize the extent to which continuity was achieved in the process of battery selection. Many changes were made in the battery from one year to the next. In any given year, however, there was considerable continuity from pretest to posttest. Moreover, many of the same areas were measured each year. Thus change scores can be calculated for each cohort, and general comparisons can be made from year to year. Although it might have been preferable to use the same measures each year in order to determine more accurately whether program effects were stable, such continuity would have involved sacrificing the opportunity to improve the battery. Since the same general areas are measured it is possible to make inferences about the stability of program effects on such characteristics

TABLE 1

Head Start Planned Variation Test Batteries

(Tests given to total sample unless otherwise specified)

1969-70	1970-71	Fall 1971	Spring 1972
N Y U Booklet 4A and 3D	California Preschool Social Competency Scale (CPSCS)	Classroom Behavior Inventory	CBI
Preschool Inventory (PSI)	N Y U Booklet 4A and 3D	Peabody Picture Vocabulary Test (PPVT)	PPVT
Motor Inhibition (MI)	PSI	Wide Range Achievement Test (WRAT)	WRAT
Stanford Binet with (1/2) Hertzig-Birch scoring	MI	PSI with H-B scoring	PSI with H-B scoring
8-Block sort (1/2)	Stanford Binet with H-B (1/2 Fall; 1/2 Spring)	Brown Self-Concept	Gumpgookies
	8-Block Sort (1/2 Spring only)	ETS Enumeration (1/3)	ETS Enumeration
	(If requested by site--)	ITPA verbal expression subtest (1/3)	ITPA verbal expression subtest (1/3)
	Ethnic Identity Quest (EIQ)	Motor Inhibition--toy truck (1/3)	Motor Inhibition--toy truck (1/3)
	Children's Cultural Awareness Scale (CCAS)	8-Block Sort (1/3)	Relevant Redundant Cue (RRC) (1/3)
			Brown (8 sites only)
			8-Block Sort (1/3 in 10 sites)
Parents' Questionnaire to 8-Block sample	Parents' Questionnaire to 8-Block sample	Parents' Questionnaire to 8-Block sample	
SRI classroom observations	SRI classroom observ.	SRI classroom observations (including observations on 6 individuals in 8 sites)	
Classroom Information Form	Classroom Inform. Form	Classroom Information Form	
Teacher Questionnaire	Teacher Questionnaire	Teacher Questionnaire	
Sponsor rating of teacher	Teacher-aide Questnre.	Teacher-aide Questionnaire	
	Sponsor rating of teacher	Sponsor rating of teacher	
	Director rating of teacher	Director rating of teacher	

TABLE 2

TEST BATTERIES FOR HSPV

(Levels represent testing levels at sites)

1969-70 All children in tested classes at each site were given the NYU 4A, NYU 3D, PSI and MI; 1/2 of the children in each site were given the SB with HB scoring and the other 1/2 of the children in each site were given the 8-Block Sort.

1970-71 Level I: All of the children in tested classes at each site were given the CPSCS.

Level II: All of the children in tested classes at each site were given the CPSCS, NYU 4A, NYU 3D, PSI and MI; EIQ and CCAS were given if the site requested them.

Level III: All of the children in tested classes at each site were given all of the tests of Level II; a random 1/2 of the children in the tested classes of each site were given the SB with HB scoring and the other 1/2 were given the 8-Block Sort.

1971-72 Level I: All of the children in tested classes at each site were given the Schaefer.

Level II:

Fall: All of the children in tested classes at each site were given the Schaefer, PPVT, WRAT, PSI with HB scoring, Brown Self-Concept; in addition, a random 1/3 of the children in tested classes at each site were given the ETS Enumeration, ITPA verbal expression subtest, Motor Inhibition (toy truck) and the 8-Block Sort.

Spring: All of the children in tested classes at each site were given the Schaefer, PPVT, WRAT, PSI with HB scoring, Gumpgookies, ETS Enumeration; in addition, a random 1/3 of the children in tested classes at each site were given the ITPA verbal expression subtest, Motor Inhibition (toy truck) and RRC.

Level III: All of the children in tested classes at each site were given the same tests of Level II; all of the children in tested classes at 8 sites were given the Brown in the spring; a random 1/3 of the children in tested classes at 10 sites were given the 8-Block Sort in the spring.

TABLE 3

HEAD START PLANNED VARIATIONS SITES 1969-1972

	1969-1970	1970-1971	1971-1972
NIMNICHT	Duluth	Duluth Fresno Tacoma Salt Lake* Buffalo*	Duluth Salt Lake Tacoma Buffalo*
TUCSON	LaFayette Lakewood	LaFayette Lakewood* Lincoln	LaFayette Lakewood Lincoln Des Moines**
BANK STREET	Tuskegee Wilmington	Tuskegee* Wilmington Boulder Elmira	Tuskegee Wilmington Boulder* Elmira
BECKER- ENGELMANN	Tupelo E. St. Louis	Tupelo E. St. Louis E. Las Vegas	Tupelo E. Las Vegas Providence**
BUSHELL	Oraibi Portageville	Oraibi Portageville Mounds	Oraibi* Portageville Mounds
WEIKART	Ft. Walton Central Ozarks	Ft. Walton Central Ozarks* Greeley Seattle	Ft. Walton Central Ozarks Greeley Seattle*
GORDON	Jacksonville Chattanooga	Jacksonville* Chattanooga Jonesboro Houston	Jacksonville* Chattanooga Jonesboro Houston
EDC	Washington Johnston Co.	Washington Paterson Johnston Co.	Washington* Paterson Johnston Co.
PITTSBURGH		Lock Haven	Loch Haven Montevideo
REC		Kansas City	Kansas City
NYU		St. Thomas	St. Thomas*
ENABLERS		Billings Colorado Sp. Bellows Falls Newburgh* Puerto Rico*	Billings Colorado Sp. Bellows Falls Newburgh Puerto Rico*

*Level I
**1971-72 sites with only comparison classes

as general cognitive skill, certain kinds of achievement, and motor inhibition. The use of different measures is, therefore, not as much of a handicap as might seem at first glance. It simply provides a stricter test for assessing the relative stability of program effects.

Coverage. Table 4 lists the instruments by areas of measurement. Three basic categories are used in classifying the measures: cognitive, psychomotor and socio-emotional or affective. No attempt has been made to provide an exhaustive list of all the possible sub-categories within the three major groups. Those sub-categories covered by the battery are merely listed.

Any category scheme is somewhat arbitrary. Our three general categories are widely used in classifying educational objectives[1] and encompass the major goals of most preschool programs. The sub-classifications are commonly-used descriptions of specific goals; it is possible, however, to imagine very different sets of categories. Some of the tests were very difficult to classify. For example, the ITPA is classified as a test of cognitive processes, although it might also be included among the measures of general ability or of achievement. The Hertzig-Birch is classified as a measure of cognitive style despite the fact that it is sometimes interpreted as a broader socio-emotional measure.

---

[1] See, for example, Bloom et al., 1956.

TABLE 4

AREAS OF MEASUREMENT

General Area	Specific Area	Measure
Cognitive	Achievement	WRAT PSI NYU 3D and 4A
	General intellectual development or "ability"	Stanford-Binet PPVT
	Cognitive style	Hertzig-Birch
	Cognitive processes	RRC Enumeration ITPA
Psychomotor		Motor Inhibition
Socio-emotional or affective	Social competencies	CPSCS CBI
	Self-concept	Brown
	Achievement motivation	Gumpgookies
	Ethnic attitudes	EIQ/CCAS
	Classroom Interaction	COI
Background	Demographic	CIF,PIF
	Mother-child interaction	8-Block Sort

Table 4 illustrates two major points. First, many important areas are covered by the battery, at least to the extent that the tests measure what they say they are measuring. Second, many possible categories are left out. The psychomotor area, for example, potentially includes a large number of categories: small muscle skill, large muscle development, and so on. The whole area of perceptual skills might have been included. Many attitudes, motivations and social skills which might have been included as sub-categories in the socio-emotional area are not covered.

Reliability Issues

This discussion focuses on the technical question of whether the test scores used in the analyses are interpretable estimates of classroom and individual "true scores." We have gone through several steps in attempting to arrive at an answer to this question, since any one of a number of technical characteristics can introduce unacceptable error components.

Before summarizing our information on the tests, it may be useful to distinguish two sources of error: randomness and bias. Internal reliability coefficients provide an estimate of the amount of variation in test scores. High internal reliability estimates indicate that the items on the test are generally measuring the same characteristic, and that a score on that test is close to the "true score"

which would be obtained on a perfect test for that characteristic. The lower the reliability of a test, the less confidence we have that the score a person receives is close to his true score. This means that if an individual took the same test several times, one could expect to find more difference in his scores on a low reliability test than on a high reliability test.

When test reliability is low, one can have little confidence in the accuracy of an individual's score. When dealing with group means, however, the situation is somewhat different. Two things happen to increase confidence in estimated group means and estimated differences between groups. First, when calculating the reliability of a test given to a group, one takes into account the number of people in each group. If one can assume that the error components for scores of individuals within a group are uncorrelated, then the ratio of the true variance to the error variance contained in a mean increases proportionately to the size of the group. This is analagous to increasing the length of the test,[1]

---

[1] Shaycoft (1962) gives the following formula for estimating the reliability of group means:

$$r_{\bar{a}\bar{a}} = 1-[\frac{1-r_{aa}}{n}]\ [\frac{s_a^2}{s_{\bar{a}}^2}]$$

where $r_{\bar{a}\bar{a}}$ is the reliability of the group mean, r_{aa} is the reliability of individual scores on the test, s_a^2 is the variance of the individual scores, $s_{\bar{a}}^2$ is the variance of the group means and n is the number of individuals in the group. Suppose we have a test whose reliability is .6. Suppose it is given to groups of 20, and suppose that the variance of the means is 10 per cent of the variance for individuals.

The formula shows that the reliability of the group mean will be

$$r_{\bar{a}\bar{a}} = 1-[\frac{1-0.6}{20}]\ [\frac{1.00}{.10}] = .80$$

This is considerably higher than the individual test reliability.

and gives a higher reliability coefficient. Second, the reliability of group means is also sensitive to the amount of "true" difference there is among the groups -- the greater the differences among the groups relative to the variation in the individual scores, the greater the reliability of group means. Thus, when group means are compared, or when classrooms are used as the unit of analysis, the reliability is often much higher than when individual scores are used. This suggests that low reliability tests are very dangerous to use when classifying individuals, but that they may well be both useful and appropriate when comparing groups. Even with classroom analyses, however, low reliability tests imply that the standard error will be larger than if the test were highly reliable. Precise comparisons become impossible.

Internal reliability estimates also provide a starting point for discussions of validity. Low reliability may be a warning signal in terms of face validity. If only a modest proportion of the variance in actual test scores is explained by true score estimates (which is what low reliability implies) it is possible that the unexplained variance is not simply random. Some other characteristic may be being measured quite accurately. These issues of face validity are taken up in the next section.

Biased scores are a more serious matter than scores which include even a large random error component. Scores

from tests that have ceiling effects, floor effects, or systematic tester effects for the HSPV sample do not yield an interpretable estimate of a true score, even at the classroom level. In evaluating the quality of the tests, then, we distinguish between those which appear to be biased for our sample and those which simply contain random error. Also distinguished, for purposes of assessing the precision of mean scores, are those tests with a high random error component from those with low error.

Since we are interested in test reliabilities for the HSPV sample and HSPV testing conditions, published information has not been relied on exclusively for the estimates. Instead, four procedures have been followed:

1. Estimating, from the literature, the internal consistencies of the tests and the stability of test scores from one administration to the next. Most weight has been placed on estimates from studies in which the age and composition of the sample was similar to that of the HSPV sample.

2. Estimating the internal consistency of the tests for the HSPV sample, by calculating KR-20 reliability coefficients for the sample and for various sub-groups in the sample. These sample coefficients have been computed for Fall 1969, Fall 1970 and Fall 1971 data.

3. Estimating the test-retest reliability coefficients for some of the tests for the HSPV sample, by conducting the reliability studies reported in Part III.

4. Estimating the amount of error variance or bias introduced by HSPV testing conditions. These estimates were obtained from the inter-tester reliability study reported in Part III. Further information was obtained from observations of testing.

The confidence we have in our reliability estimates varies considerably from test to test. Some of the tests have been widely used and extensively reported while others were developed specifically for the Head Start evaluation. Some tests were used for several years; others were used only once. For many of these latter tests our current estimates are based only on our own data. The 1971-72 testing was monitored not only by SRI but also by independent observers, and this gives us more confidence that those particular tests were given under standard conditions. The unevenness of all these checks on the data mean that we have much more confidence in some of the tests than in others. The PSI, for example, has been widely used. Thus, there are KR-20's for the HSPV sample, test-retest and inter-tester reliabilities, and observations of field administration. At the other extreme, there is very little data for the new Relevant Redundant Cues Test,

since it was given in only the spring of the third year.

The results of all our specific reliability investigations are reported on a test by test basis in Part III. Table 5 summarizes all the reliability findings from the HSPV Study and other sources.

Validity Issues

Concurrent and face validity: The first question in examining the validity of the HSPV instruments is whether they are measuring what they purport to measure. This is not always easy to determine. All tests measure both a general test taking ability and a general test taking motivation. Tests for young children almost always measure the ability to understand directions and the ability or motivation to pay attention to a task. Some tests measure more of these extraneous components than others, although it is impossible to obtain a precise measurement of them.

To some extent, what a test is measuring can be inferred from its correlations with other measures. For the more widely used tests, correlations found in other studies are reported in the individual test descriptions. Table 6 shows intercorrelations for the HSPV 1970 battery and Table 7 shows intercorrelations from the HSPV 1971 test batteries.

TABLE 5

SUMMARY OF HSPV TEST RELIABILITIES

Brown Self-Concept Test

Internal reliability moderate (KR-20's =.70's in HSPV Study, .60's in ETS Study); test-retest = .55 (2-3 weeks) in Follow Through Study; ceiling effects and response biases found in all studies.

California Preschool Social Competency Scale

Odd-even reliability = .96 in HPSV Study; no interjudge data on HSPV sample; high interjudge estimates (.80's) in manual; culturally-biased items.

Classroom Behavior Inventory

Test-retest (2 weeks) in .70's in HSPV Study; 3 independent traits shown in factor analyses; internal reliability high in published sources; inter-observer reliability estimates low for paraprofessionals in HSPV Study.

Eight-Block Sort Test

Very high inter-judge agreement on success scores; low to moderate inter-judge agreement on mother-child interaction codes; higher agreement using frequency counts than frequency per minute units.

ETS Enumeration Test

Test-retest coefficients (2 weeks) moderate in HSPV Study; KR-20's for total score = .70's in HSPV Study and ETS Study; KR-20's for subtests vary -- high for Counting and Touching, moderate for Same Number Matching and low for Same Order Matching.

Gumpgookies

High internal reliability for long form in published sources; questionable reliability for shorter form; possible ceiling effects and item biases.

Hertzig-Birch Scoring

No data available.

ITPA-Verbal Expression Subtest

Internal consistency estimates in published sources high for subtest (.70's - .80's); test-retest moderate (2 weeks) in HSPV Study; statistically significant (.04 level) tester effects in HSPV Study.

TABLE 5
(cont.)

Motor Inhibition Test

Questionable reliability; statistically significant (.001 level) tester effects in HSPV Study.

NYU Booklets 3D and 4A

HSPV KR-20's = .66(3D) and .69(4A); floor and ceiling effects.

Peabody Picture Vocabulary Test

Test-retest moderate in published sources; high (.90's) internal reliability in ETS Study; no HSPV data since calculating KR-20's is inappropriate for "tailored" test.

Preschool Inventory

High KR-20 estimates in HSPV Study =.90's (64-item) and .80's (32-item); high internal reliability estimates from other studies; high test-retest (2 week) reliability (.80's - .90's) in HSPV Study; no known biases.

Relevant Redundant Cue Test

Low reliability estimates -- presence of random fluctuation.

Stanford-Binet

Internal consistency estimates high (.80 - .90's) in manual; test-retest estimates high in published sources; no known biases.

Wide Range Achievement Test

HSPV KR-20's for Fall subtests -- .80 for Copying Marks and Recognizing Letters, .85 for Naming Letters, and .60 for Reading Numbers; no data on spring subtests; floor effects for some Fall subtests.

TABLE 6

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	.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).

TABLE 7 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)	.666 (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	.475 (1075)	.508 (1097)	.427 (1097)	.307 (1097)	.446 (1097)	.542 (1097)	.584 (1073)	.459 (1115)									
ETS ENUMERATION COUNTING	.492 (1075)	.504 (1097)	.322 (1097)	.359 (1097)	.500 (1097)	.620 (1097)	.625 (1073)	.384 (1115)	.761 (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	.259 (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)	.032 (637)	.136 (597)	.135 (597)	.047 (597)	.107 (597)	.118 (610)	.139 (610)			
EIGHT-BLOCK PLACEMENT	.304 (1119)	.222 (1148)	.271 (1148)	.145 (1148)	.207 (1148)	.304 (1148)	.305 (1090)	.303 (1096)	.322 (1032)	.415 (1032)	.200 (1032)	.180 (1032)	.212 (1113)	.183 (1113)	.005 (573)		
EIGHT-BLOCK REASON	.445 (1119)	.354 (1148)	.333 (1148)	.286 (1148)	.372 (1148)	.390 (1148)	.445 (1090)	.418 (1096)	.405 (1032)	.402 (1032)	.258 (1032)	.211 (1032)	.178 (1113)	.168 (1113)	.063 (573)	.520 (1211)	
EIGHT-BLOCK SUCCESS TOTA.	.439 (1119)	.346 (1148)	.351 (1148)	.267 (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.

The data which is available for making inferences about what the tests are measuring suggest that a general test-taking component, perhaps a general cognitive ability factor, is being tapped by many of the tests. It is difficult to estimate the importance of this component. Correlations of the cognitive tests in the Fall 1970 battery with the Stanford-Binet are considerably lower (.30 or .40) than the correlations one usually finds between achievement tests for older children and the Stanford-Binet. These may, however, result from analytical problems or low reliabilities. One of our analyses of the 64-item PSI, for example, shows a high correlation with mental age on the Stanford-Binet but a very moderate correlation with IQ. This suggests that if analyses of the PSI, and perhaps of other tests as well, were adequately controlled for age, a general ability component might emerge more strongly. The correlation in Table 6 of the PSI and the 3D illustrate the possible problems introduced by unreliability. If the correlation of .696 is corrected for unreliability, it becomes extremely high.[1] If we corrected other correlations for unreliability, we might find equally strong indications of a general cognitive component.[2]

---

[1] Using the formula $\frac{r_{1.2}}{\sqrt{t_1 t_2}}$ where t_1 and t_2 are estimates of the test reliability the estimated correlation between the true score components of the PSI and the 3D is $\frac{.7}{\sqrt{(.7)(.9)}} \cong .9$

[2] Other corrected correlations are reported in the technical reports of Part II.

The data we have, however, suggests that much of the non-error variance is unique to specific tests. This interpretation is supported by Shipman's (1971, 1972) factor analysis of the Year 1 and Year 2 data from the ETS longitudinal study. In both years two interpretable factors emerged from her analysis of about fifty tests: a general ability or test-taking factor which explained about 20 percent of the variance, and a response speed factor which explained an additional 5 percent. Additional factors tapped task-specific behaviors. Sub-clusters of tests did not emerge; i.e., the tests could not be grouped into such categories as vocabulary, classification ability, or analytic functioning. In Year 2 using an older sample, there was also a spontaneous verbalization factor and some evidence for generalization of specific personal and social behaviors across tasks. If these findings are supported by the HSPV analysis, interesting questions are raised about the proper interpretation of the data. Inferences about the effects of programs will have to be made on the basis of their effects on general cognitive ability and on a wide variety of specific tasks, whose individual significance will have to be determined.

Inferences about what the tests are measuring can also be made by examining the content of the tests, observing children's behavior in the testing situation, and analyzing response rates to items. Table 8 summarizes our general impressions of the face validity of the tests, based on both analytical and observational data.

TABLE 8

FACE VALIDITY

Test	Measures	Also Measures
1. Brown	self-concept (ability to verbalize acceptable responses)	verbal skills, rapport with examiner
2. CPSCS	social competency	acceptance of middle-class norms, teacher biases and response style
3. CBI	hostility, extroversion, task persistence	teacher biases and response styles
4. Enumeration	number concepts	attention
5. Gumpgookies	achievement motivation (ability to verbalize achievement-oriented responses)	cognitive understanding, attention, knowledge of middle-class norms, response set
6. Hertzig-Birch Scoring	cognitive style	internalization of expected behavior in testing situation
7. ITPA	verbal expression	test-specific motivation, quantity rather than quality
8. MI	ability to inhibit motion	understanding of directions, coordination, small muscle control
9. NYU 3D	relational, pre-math, pre-science, linguistic concepts	general ability
10. NYU 4A	numbers, letters, shapes	general ability
11. Peabody	receptive vocabulary	general ability, persistence
12. PSI	various school-related skills	general ability

TABLE 8

(Con't)

Test	Measures	Also Measures
13. Stanford-Binet	general intelligence	motivation, general learning
14. WRAT	specific academic skills	general intelligence, persistence
15. RRC	concept acquisition	persistence, cognitive style

If these impressions are correct, they point out one of the major limitations of currently available non-cognitive tests for young children -- that they are often tests of cognition and attention. Since one cannot tell what proportion of the variance is explained by extraneous rather than relevant characteristics, it is well to be cautious in interpreting the results. The relatively low internal reliabilities for non-cognitive tests also suggest the salience of extraneous characteristics.

The table also allows us to comment on another general problem, that of distinguishing measures of general cognitive ability from measures of achievement. Cognitive tests inevitably measure both ability and achievement: what people have learned, for example, affects their Stanford-Binet scores; their ability to pick up the directions of the WRAT affects their achievement scores. Researchers almost always find high correlations between achievement tests and IQ measures, and among achievement tests themselves.[1]

---

[1] The four "achievement" tests used in the Equality of Educational Opportunity Survey provide a good example of a common finding. The intercorrelations for sixth graders, and the loading on first principal components are as follows:

	2	3	4	r with 1st Principal Component
1. Non-verbal ability	.697	.702	.755	.85
2. Verbal ability		.717	.850	.90
3. Reading comprehension			.860	.90
4. Math computation				.96

(Table from Mosteller and Moynihan, p. 473, based on Mayeske et al., 1968. The table was computed from Table 2, corrected for unreliability.)

The correlations reported for the instruments in the HSPV battery are not as high as one might expect on the basis of research with older children, though as noted earlier, this may be an artifact of the analysis. This is somewhat reassuring, since it indicates that specific sorts of cognitive achievement may be measured.

Predictive validity: The second question to ask in looking at the validity of the HSPV battery is whether it is measuring anything important. Since performance on a test is obviously not important in itself, the question is whether performance on the test is related to other important outcomes considered or known to be important.

These "important outcomes" are not always long term effects. Many educators are interested in affective or cognitive skills not because they are important in achieving something else like school success or a higher income, but because they are important to a child in the present. These educators are not interested in whether scores on a test predict future scores on the test or, indeed, on anything else. They are, of course, interested in whether scores on the test are associated with some other indication of the characteristic which the test purports to measure. When they have no other measures (for example, in the case of self concept) they are forced to argue that what the test measures is the characteristic, and that scores on

the test are important in and of themselves. Whether this is an adequate argument depends, for these tests, on the quality of the theory and the face validity of the instrument.

The argument is more often made, however, that test scores are important because they are predictive of some longer term "important outcome." Test scores are of interest because they are a way of estimating long-term program effects without conducting longitudinal studies. Let us assume that what we are really interested in is whether a program improves school achievement and life chances. The way to find this out is to compare people who have been in the program with people who have not (but who are comparable in other ways) at various points in their lives. Since this is a long and costly process, evaluators look to test scores as a short-cut. Tests for this purpose are chosen on the basis of whether they predict those outcomes which would be measured if a lifetime evaluation were possible. Programs are then evaluated on the basis of their ability to raise scores on these tests.

Most of the discussion in this section will deal with the question of whether scores on the cognitive tests in the HSPV battery predict important long-term outcomes. Given our serious doubts about the face validity, and the lack of predictive and construct validity for non-cognitive instruments, they are not discussed here.

There are fairly reliable predictive data for tests of general cognitive ability. Even for these tests, however, it is important to remember that the data we have do not deal with the question of whether test scores which occur as the result of a planned intervention predict important outcomes. This question is most difficult with regard to tests of IQ or general cognitive development. IQ is generally considered to be important, in that it is a good predictor of a number of adult outcomes. It seems reasonable, therefore, to evaluate programs on the basis of their effects on IQ scores. But what we do not know is whether IQ measured at the end of intervention programs is a better or worse predictor of these outcomes than IQ measured at the beginning. We do not even know the extent to which induced IQ gains are stable. Evidence from any number of preschool programs indicates that the IQ advantage of experimental groups tends to disappear by third grade (see Stearns, 1971 for a summary of this literature.) There is almost no data available for answering the question of whether induced IQ gains which are stable have the same relationship to other outcomes as IQ in the normal population. We can, however, look at the relationships between IQ and other outcomes for the general population. Although the Stanford-Binet and other IQ tests have been widely used and extensively reported, there is no one longitudinal data set from which predictive validities can be inferred. The data in

this section come, therefore, from a variety of sources. The basic question is how well IQ measured at age 5 predicts various outcomes. The answer must be inferred from data on the correlation of IQ at age 5 with IQ at other ages, and data on the correlation of IQ around age 18 with adult outcomes.

Bloom (1949) summarizes the available data on the stability of IQ over time. Using the highest estimate (.71) found between intelligence at ages 4 and 17 (Bayley, 1949) implies that IQ at age 4 explains about 50 percent of the variance in adult IQ. Using this stability of IQ estimate and the data that exist on the relationship between adult IQ (at age 18) and two adult outcomes -- occupational status and income -- in path model analyses, Jencks et al. (1972) found that the inferred correlations between IQ at age 5 and adult income or status were low. The large proportion of the variance -- about 88 percent for status and 95 percent for income -- in these adult outcomes was unexplained by the early IQ estimates. This does not mean, however, that early IQ is unimportant. It has an important relationship to eventual years of schooling completed and a relationship to grades and to achievement tests. Furthermore, data on the relationship between IQ and other adult outcomes of major interest -- for example, job satisfaction, family stability, or general happiness--- do not exist.

The discussion thus far has dealt only with the relationship between general cognitive ability and adult outcomes. In addition to a general cognitive ability component, however, the tests in the HSPV battery measure a number of more specific skills. Presumably, the tests measure the skills which they say they measure: vocabulary on the PPVT; numbers, letters and spelling on the WRAT; a variety of information and concepts on the PSI. How important are these skills? More specifically, how much of an advantage is it to a child to learn these skills in preschool rather than in regular school? There is little relevant data which could help answer these questions. Although moderate correlations are reported between scores on many of the tests and school achievement (see Part II), it is hard to know how much both measures are influenced by general cognitive competence. There is apparently no data on the long-term predictive validity of the tests.

In the absence of data, we must fall back on theoretical arguments for the importance of early acquisition of skills. These are the same arguments which justify the importance of short-term IQ gains. One argument rests on structural considerations, the other on psychological. The structural argument says that possession of cognitive skills, whether general or specific, at the time of entrance into school, gives a child an advantage relative to his classmates.[1] This advantage means that he is more likely to be

[1] Wolff and Stern (1966) gave some indication that this might occur.

placed in a higher level group or class, and therefore to be taught more and expected to do well.[1] This may lead to higher achievement and a better chance of attending college.[2] More schooling is more important than IQ in producing adult advantages.[3]

Thus, an advantage may accrue to children who come into school ahead of their classmates, simply because of the grouping patterns in the schools and the advantages which accompany placement in a higher track.

A second sort of advantage is psychological and, again, it is a relative advantage. Children who begin school doing as well as or better than their classmates may feel better about themselves and about school than children who start off doing badly. This increased motivation may have an important effect on their later success. Unfortunately, there is no data with which to test this theory. If it is

---

[1]Suggestive evidence on intra-class grouping is given in Rist, 1970. The hypothesis gains some additional support from the Coleman findings on kindergarten. Within-school analyses revealed that children who had attended kindergarten achieved slightly higher than their classmates. But schools with higher proportions of children who had attended kindergarten did not have higher average achievement scores than schools with low proportions of kindergarten alumni. This indicates that the benefits of kindergarten may result more from being given a relative advantage than from learning something.

[2]See Jencks et al., 1972. When all other factors are controlled, students in the college track in 9th grade are 12 percent more likely to attend college than students in the non-college track.

[3]Blau and Duncan, 1967; Duncan, Featherman and Duncan, 1968.

true, however, it suggests that the short-term gains which preschool programs seem to be capable of producing may be more important than simple predictive validity estimates indicate.

Background Measures

Demographic information: The main source of demographic data on the HSPV sample is the Classroom Information Form filled out by teachers. This instrument was developed for the HSPV evaluation. Only one reliability study was done using the instrument. This study is reported in Appendix C. We have no reliability information for the first two years of the HSPV evaluation, but we have no reason to believe that the quality of the data differed much from that gathered in 1971. As far as we can tell, the data on preschool experience, family size and parental education is sufficiently reliable to be useful. The data on parental occupations is not particularly good, and that on language in the home is unreliable.

Mother-child interaction: One measure of mother-child interaction, the 8-Block Sorting Task, has been used in the HSPV evaluation. The strengths and weaknesses of this measure are summarized in the test description in Part II. At this stage in the development of such measures, their meaning is a matter for speculation and inquiry.

Summary and Future Directions

One benefit of the Head Start Planned Variation Study has been the developmental work in creating suitable measures in many child outcome domains for large scale evaluations. From an intensive review of the technical quality of all the measures used in the Head Start Planned Variation Study, we have most confidence in the achievement measures, especially the Preschool Inventory, and in the intelligence measure -- the Stanford-Binet. Our confidence in the quality of the non-cognitive instruments -- Brown, California Preschool Social Competency Scale, Gumpgookies and the Classroom Behavior Inventory -- is very low because of their poor psychometric properties and/or their lack of validity. In most cases, the more developmental and experimental measures, such as those that measure cognitive abilities (ETS Enumeration, ITPA-Verbal Expression Subtest, and Relevant Redundant Cues), those that assess cognitive-style (Hertzig-Birch codes), and those that measure mother-child interaction, need more refinement and study before we can conclude how meaningful their use in future large scale evaluations will be. Their use in the HSPV analyses in most cases will be minimal since they are still in the developmental stages and thus have many problems associated with their interpretation.

Based on our work in measurement with young children, the following directions for future work in this area are recommended:

1. Non-cognitive instruments presently available for young children are poor. Further development of paper and pencil tests in this area is unadvisable since we have serious doubts about their validity and psychometric properties. Since the basic problem underlying the development of non-cognitive measures for young children is the lack of adequate developmental theory, a major research effort needs to be launched in this area. After more theoretical work is done, a more adequate non-cognitive measurement technology for young children can be developed. At present, the one type of measure that looks most promising for future assessment of social and emotional development in young children is the observational instrument, especially those that are developed to test specific theories and hypotheses. Much refinement in the available observation instruments, however, needs to be done before such measures can be meaningfully used in large-scale evaluations.

2. Further analyses with the achievement measures, such as the WRAT and the NYU Booklets, needs to be done. Consideration of these achievement measures as criterion-reference measures instead of norm-reference measures may yield more information in future planned variation evaluations. These measures are custom-tailored to assessing the goals of the more structured programs in the HSPV Study. In addition, other criterion-referenced measures ought to be developed.

3. Since the child outcome analyses show that many of the results are item or test specific, a detailed reanalysis of many of these tests and other tests from similar evaluations might result in the generation of a test battery that would be more sensitive to possible treatment differences.

4. Further study of the more experimental measures in the battery needs to be done before they are used in other large-scale evaluations. Many of these more experimental measures (i.e., Hertzig-Birch codes, ETS Enumeration, ITPA Verbal Expression Subtest, etc) tap important skills or abilities underlining cognitive competence. Future work in these developmental areas is encouraged.

5. The one psychomotor measure used in the HSPV Study -- the Motor Inhibition Test -- is inadequate. Further development of tests to assess psychomotor development in young children for use in large-scale evaluations is needed.

6. Developing ways of assessing the quality and quantity of mother-child interaction will be beneficial to future evaluations of early intervention programs. Intensive refinement of the Eight-Block Sort observational procedure needs to be done before it will generate such information. In addition, the development of other mother-child measures is encouraged.

7. Further inquiry and discussion on how test scores should be calculated and reported (i.e., grade equivalent, standard score, raw score, etc.) is necessary in improving the interpretability and the comparability of evaluation findings.

8. Greater thought and examination should be given to what elements should be considered in the conceptualization of an evaluation of a planned variation experiment. Such inquiry should address the issues involved in comparing sponsors (or programs) which have a wide variety.of goals and objectives.

9. It is recommended that technical and norming information from future large-scale evaluations be provided in a similar report form so that a cumulative knowledge based about instruments suitable for use with young children can be developed.

BIBLIOGRAPHY

Bayley, N. Consistency and variability in the growth of intelligency from birth to 18 years. The Pedagogical Seminary and Journal of Genetic Psychology, 1949, 75, 165-196.

Blau, P., & Duncan, O. D. The American occupational structure. New York: John Wiley, 1967.

Bloom, B. S., Stability and change in human characteristics. New York: John Wiley, 1964.

Bloom, B.S., (Ed.) Taxomomy of educational objectives: the classification of educational goals. Handbook I - Cognitive domain. New York: Longmans, Green, 1956.

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

Duncan, O. D., Featherman, D. L., & Duncan, B. Socioeconomic background and occupational achievement: Extensions of a basic model. Final Report, Project No. 5-0074 (EO-191) Contract No. OE-5-85-072. Washington, D.C.: U.S. Office of Education, Bureau of Research, 1968.

Educational Testing Service. Disadvantaged children and their first school experiences: Theoretical considerations and measurement strategies. Prepared for OEO Contract #4206 and OEO Grant #CG-8256. Princeton, N. J.: ETS, 1968.

Jencks, C., et al. Inequality: A reassessment of the effects of family and schooling on America. New York: Basic Books, 1972.

Mayeske, W., et al. Item response analysis of the educational opportunities survey student questionnaires. Technical Note 64, U.S. Office of Education, April 1968 (mimeo).

Mosteller, F., & Moynihan, D. P. Equality of educational opportunity. New York: Random House, 1972.

Reiss, A. J., et al. Occupations and social status. New York: Free Press, 1961.

Rist, R. C. Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. Harvard Educational Review, 1970, 40, 3, 411-45.

Shaycroft, M. The statistical characteristics of school means. In Flanagan, et al. Studies of the American High School. University of Pittsburgh, 1962.

Shipman, V., et al. Disadvantaged children and their first school experiences: Structure and development of cognitive competencies and styles prior to school entry. Prepared for HEW Grant #H-8256. Princeton, N.J.: Educational Testing Service, 1971.

Shipman, V.C. Disadvantaged children and their first school experiences: Structural ability and change in the test performance of urban preschool children. Prepared for HEW Grant #H-8256. Princeton, N.J.: Educational Testing Service, 1972.

Stearns, M. S. Report on preschool programs: the effects of preschool programs on disadvantaged children and their families. Washington, D.C.: Office of Child Development, HEW, 1971.

Walker, D.K. Socio-emotional measures for preschool and kindergarten children: A handbook. San Francisco, Calif.: Jossey-Bass, 1973.

Wolff, M., & Stein, A. Six months later: A comparison of children who had Head Start, Summer, 1965, with their classmates in kindergarten. Ferhauf Graduate School of Education, Yeshiva University, 1966 (mimeo).

White, S. H., et al. Federal programs for young children: review and recommendations. Prepared for contract #HEW-OS-71-170 with the Office of the Assistant Secretary for Planning and Evaluation, HEW. Cambridge, Mass.: Huron Institute, 1972.

PART II: TECHNICAL REPORTS ON THE MEASUREMENT INSTRUMENTS

NOTE

Part II includes a technical report on the nineteen measures used to assess child outcomes and background variables in any or all of the three years of the Head Start Planned Variation Study. Each measure's report is a separate entity which includes the following information, if available: the purpose of the test; an abbreviated description of the test with the scoring procedures; a history of the development of the instrument; relevant technical information on standardization procedures, reliability and validity; remarks; and references. Each section includes technical information generated from the Head Start Planned Variation sample of either the second or third year. For all HSPV analyses, only "valid" tests, as recorded by the tester, were used. If the instrument has been used in either the ETS Longitudinal Head Start Study or the Home Start Study, the detailed findings of these studies involving similar preschool populations are reported in the appropriate technical sections. In addition, any study found in the literature using the instrument with a preschool population is reported; there are few such studies mentioned, however, since many measures used in the HSPV Study are new and in the developmental phases.

The nineteen measures described in Part II are listed below with the year they were used in the HSPV Study (note: Year 1 = 1969-70, Year 2 = 1970-71, and Year 3 = 1971-72):

Brown IDS Self-Concept Referents Test - Year 3
California Preschool Social Competency Scale - Year 2
Classroom Behavior Inventory - Year 3
Classroom Information Form - All three years (different versions)
Classroom Observation Procedure - All three years (different versions)
Eight-Block Sort Task - Year 1, Year 2 (Spring only), Year 3 (different versions)
Ethnic Identity Questionnaire, Children's Cultural Awareness Scale - Year 2
ETS Enumeration Test - Year 3
Gumpgookies - Year 3 (Spring only)
Hertzig-Birch Scoring - Year 1 and Year 2 with the Stanford-Binet, Year 3 with the 32-item Preschool Inventory
Illinois Test of Psycholinguistic Abilities -- Verbal Expression Subtest - Year 3
Motor Inhibition Test - Three subtests in Year 1 and Year 2, one subtest in Year 3
NYU Booklets 3D, 4A - Year 1, Year 2
Parent Information Form - All three years (different versions)
Peabody Picture Vocabulary Test - Year 3
Preschool Inventory - 64-item version in Year 1 and Year 2, 32-item version in Year 3
Relevant Redundant Cue Concept Acquisition Task - Year 3 (Spring only)
Stanford-Binet Intelligence Test - Year 1, Year 2
Wide Range Achievement Test - Year 3 (some subtests only in Spring)

Tables 1 - 3 in Part I describe in detail the HSPV samples used in each year for each measure. Instruments used to describe the implementation process are described in other Huron Institute reports: Implementation of Head Start Planned Variation: 1970-71 by C. V. Lukas and C. Wohlleb, Implementation of Head Start Planned Variation: 1971-72 by C. V. Lukas, and An Exploratory Study of the Match Between Classroom Practice and Educational Theory by A. C. Monaghan. Further discussions of the child outcome instruments used in the first two years of the study are available in two Huron Institute reports: Some Short Term Effects of Project Head Start: A Preliminary Report on the Second Year of Planned

Variation -- 1970-71 by Marshall S. Smith and Cognitive Effects of Preschool Programs on Different Types of Children by Helen Featherstone.

Brown IDS Self-Concept Referents Test

Purpose

The Brown IDS Self-Concept Referents Test is designed to examine a child's self-concept as well as his perceptions of what others think of him. Since developmental theory and empirical studies support the fact that young children probably cannot differentiate between their perceptions of themselves and others' perceptions of them, only that part of the test measuring the child's perceptions of himself was used in the Head Start Planned Variation Study. Many educators and psychologists in early childhood education have pointed out the importance of a child's self-concept in learning and development. Since a primary goal of most preschool programs is to increase a child's self-concept, there has been a great need for assessment techniques in this area. Unfortunately, in contrast to numerous instruments available for measuring language development and cognitive functioning there are very few good instruments available for assessing the development of self-concept in young children (Buros, 1970; Coller, 1971; Walker, 1973). For a review of the scarce literature on the emergence and development of self-concept in young children, see Lacrosse et al. (1970) and Wylie (1961).

Description

At the beginning of the test a full-length colored Polaroid photograph is taken of each child in a standard

setting. The tester then asks the child to make a response about the picture to make sure the child recognizes himself in the picture. While looking at the picture, the child is asked 16 bipolar questions in an "either-or" format. Eight of the questions present the pairs of opposite adjectives (e.g., "Is (child's name) happy or is he (she) sad?"), while the remaining eight questions ask the child whether he does or does not possess a certain characteristic (e.g., "Does (child's name) like to play with other kids?"). These latter items (such as #s 4, 5 and 9) involve more complicated syntax than the items based on pairs of adjectives (such as items 1 - 3). The items presented to the child are listed below in abbreviated form with the correct answer underlined.

1. happy-sad
2. clean-dirty
3. ugly-good looking
4. likes to play with other kids--doesn't like to play with other kids
5. likes to talk a lot--doesn't like to talk a lot
6. likes to have other kids' things--likes to have own things
7. bad-good
8. smart-stupid
9. scared of a lot of things--not scared of a lot of things
10. likes the way his clothes look--doesn't like the way his clothes look
11. scared of a lot of people--not scared of a lot of people
12. strong-weak
13. sick-healthy
14. likes the way his face looks--doesn't like the way his face looks
15. has a lot of friends--doesn't have a lot of friends
16. when gets up in the morning and thinks about going to school, feels really good--doesn't feel really good

If the child doesn't answer the question the first time, it is repeated in its entirety. If the child still doesn't answer, the question is read a final time in two separate parts (i.e., "Is (child's name) happy? Is (child's name) sad?" instead of "Is (child's name) happy or sad?"). The items are scored "1", "0", or indeterminate; "1" stands for the more socially desirable attribute. Two scores -- called unadjusted and adjusted -- were used in the HSPV analyses. The unadjusted score equals the total number of correct answers (maximum = 16). The adjusted score equals the percentage of correct responses for those items clearly answered in a positive or negative way (maximum = 100).

Results of a coding reliability study on this test done in Fall 1971 at Stanford Research Institute are very favorable (see Appendix D).

Development of Instrument

The Brown IDS Self-Concept Test was first used by Brown (1966) in 1966 with 38 lower class black preschoolers and 36 middle class white preschoolers. Using 14 bipolar questions he first asked the children how they perceived themselves and then how their mothers, their teachers, and their peers perceived them. Brown reports that the black children's self-perceptions were significantly less favorable than those of the white children. The black children also perceived their teachers as seeing them in

a less favorable position. There was no difference between the two groups in their perceptions of either their mothers' or peers' evaluation of them. Test-retest reliability for the self-referent responses was .71 for blacks and .76 for whites. These findings were later replicated by Brown in 1967 (Shipman et al., 1971).

A study, conducted by Clark et al. (1967), raises questions about the validity of the Brown study. Clark and his associates used a fifty item completely non-verbal instrument, the U-scale, to challenge the Brown finding that black children have less favorable self-concepts than white children. The U-scale items depict a U-figure in both a positive and negative situation; the child is asked to show by pointing which drawing "is the real U?" In a sample of 95 black children and 52 white children the self-concepts of the black children, as measured by the U test, were not significantly different from those of the white children, who exhibited superior vocabulary skills.

Versions of the Brown test were also used in the first two years of the ETS Longitudinal Study (Shipman et al., 1971; Shipman, 1972) and in a Follow Through pilot study (Emrick, 1972). The version used in the ETS Study included 15 items, 14 of which were scored as self-referents. In the second year the teacher referent was also used with those children in a preschool program. The 21-item version used in the Follow Through Study included the 16 items of the HSPV version plus 5 teacher-referent items.

Norms

Norms for the adjusted and unadjusted Brown IDS Self-Concept scores for the Fall 1971 HSPV sample are available in Tables 1 - 16. 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: (note: the first table listed is for unadjusted scores; the second table listed is for adjusted scores) total sample (Tables 1, 9); males (Tables 2, 10), females (Tables 3, 11), children with previous preschool experience (Tables 4, 12), children with no previous preschool experience (Tables 5, 13), white children (Tables 6, 14), black children (Tables 7, 15), and Mexican-American children (Tables 8, 16). The mean adjusted score for the total sample is 11.585 (S.D. = 3.271, N = 2866); the mean adjusted score for the total sample is 83.389 (S.D. = 14.271; N = 2866). A developmental age trend in both scores can be seen in the norm tables. Norms for the same two scores and the number of items omitted are available for three month age intervals (42-44 months to 57-59 months for Year I and 51-53 months to 66-69 months for Year 2) for the children in the ETS Head Start Longitudinal sample (Shipman, 1972). The mean unadjusted and adjusted scores for Year 1 (42-59 months) were 10.7 (S.D. = 2.45, N = 1371) and 82.0 (S.D. = 14.6, N = 1371); the respective scores for Year 2 (51 - 69) months were 11.8 (S.D. = 2.00, N = 1285) and 86.2 (S.D. = 12.8, N = 1285).

TABLE 1

DISTRIBUTION OF BROWN SELF-CONCEPT SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	11.333	2.494
39-41	4	8.250	4.146
42-44	20	9.050	2.889
45-47	66	9.924	3.457
48-50	252	10.468	3.688
51-53	459	10.813	3.604
54-56	458	11.439	3.376
57-59	450	11.473	3.154
60-62	365	12.395	2.961
63-65	244	12.148	2.819
66-68	254	12.354	2.662
69-71	200	12.530	2.689
72-74	83	12.807	2.148
75-77	5	13.200	1.600
78-80	3	12.000	2.160
TOTAL	2866	11.585	3.271

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

[2]Maximum score = 16.

TABLE 2

DISTRIBUTION OF BROWN SELF-CONCEPT SCORES FOR MALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	1	3.000	-----
42-44	7	9.429	2.871
45-47	42	10.119	3.223
48-50	131	10.794	3.663
51-53	249	10.418	3.514
54-56	233	11.275	3.610
57-59	228	11.474	3 223
60-62	200	12.370	2.960
63-65	114	11.921	2.980
66-68	139	12.381	2.651
69-71	91	12.747	2.625
72-74	44	12.977	1.815
75-77	3	14.333	0.943
78-80	2	13.500	0.500
TOTAL	1484	11.523	3.319

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

[2]Maximum score = 16.

TABLE 3

DISTRIBUTION OF BROWN SELF-CONCEPT SCORES FOR FEMALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	11.333	2.494
39-41	3	10.000	3.266
42-44	13	8.846	2.878
45-47	24	9.583	3.807
48-50	121	10.116	3.682
51-53	210	11.281	3.653
54-56	225	11.609	3.105
57-59	222	11.473	3.081
60-62	165	12.424	2.963
63-65	130	12.346	2.654
66-68	115	12.322	2.675
69-71	109	12.349	2.727
72-74	39	12.615	2.456
75-77	2	11.500	0.500
78-80	1	3.000	-----
TOTAL	1382	11.652	3.217

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

[2]Maximum score = 16.

TABLE 4

DISTRIBUTION OF BROWN SELF-CONCEPT 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	11.500	1.500
45-47	6	9.333	3.543
48-50	34	11.618	2.797
51-53	62	10.935	3.126
54-56	57	11.193	3.103
57-59	76	11.447	3.139
60-62	121	12.959	2.563
63-65	94	12.638	2.374
66-68	96	12.844	2.391
69-71	94	12.809	2.485
72-74	35	13.143	2.307
75-77	3	14.333	0.943
78-80	2	11.000	2.000
TOTAL	682	12.284	2.805

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

[2]Maximum score = 16.

TABLE 5

DISTRIBUTION OF BROWN SELF-CONCEPT 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	3	11.333	2.494
39-41	4	8.250	4.146
42-44	17	8.706	2.946
45-47	59	10.051	3.432
48-50	212	10.321	3.799
51-53	382	10.848	3.642
54-56	388	11.536	3.376
57-59	361	11.604	3.059
60-62	232	12.190	3.023
63-65	141	11.972	2.876
66-68	152	12.092	2.727
69-71	103	12.262	2.859
72-74	48	12.563	1.989
75-77	2	11.500	0.500
78-80	1	14.000	-----
TOTAL	2105	11.431	3.332

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

[2]Maximum score = 16.

TABLE 6

DISTRIBUTION OF BROWN SELF-CONCEPT SCORES FOR WHITE CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	3	6.333	2.867
42-44	8	8.000	2.236
45-47	28	9.964	3.510
48-50	91	11.055	3.146
51-53	180	10.961	3.463
54-56	205	11.693	3.252
57-59	187	11.824	2.780
60-62	160	12.894	2.600
63-65	91	12.132	2.746
66-68	97	12.546	2.420
69-71	73	12.041	2.577
72-74	49	12.959	1.873
75-77	3	14.333	0.943
78-80	2	11.000	2.000
TOTAL	1177	11.820	3.045

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

[2]Maximum score = 16.

TABLE 7

DISTRIBUTION OF BROWN SELF-CONCEPT SCORES FOR BLACK CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score [2]	S.D.
36-38	3	11.333	2.494
39-41	1	14.000	-----
42-44	12	9.750	3.058
45-47	38	9.895	3.417
48-50	131	9.771	3.847
51-53	200	10.810	3.679
54-56	191	11.267	3.455
57-59	182	11.319	3.296
60-62	129	12.357	3.037
63-65	101	12.980	2.251
66-68	104	12.865	2.414
69-71	103	13.165	2.406
72-74	32	12.563	2.536
75-77	2	11.500	0.500
78-80	1	14.000	-----
TOTAL	1230	11.572	3.381

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

[2] Maximum score = 16.

TABLE 8

DISTRIBUTION OF BROWN SELF-CONCEPT 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	12.087	4.138
51-53	70	10.514	3.714
54-56	52	11.288	3.586
57-59	66	10.742	3.678
60-62	67	11.194	3.342
63-65	50	10.560	3.281
66-68	50	10.960	3.124
69-71	20	10.900	3.434
72-74	---	-----	-----
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	398	10.940	3.542

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

[2] Maximum score = 16.

TABLE 9

DISTRIBUTION OF BROWN-SELF CONCEPT ADJUSTED SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	83.333	5.907
39-41	4	3.500	21.546
42-44	20	73.700	13.199
45-47	66	75.803	14.137
48-50	252	77.972	19.280
51-53	459	81.227	15.488
54-56	458	82.963	14.637
57-59	450	83.313	13.743
60-62	365	86.967	11.321
63-65	244	85.332	11.363
66-68	254	85.941	11.592
69-71	200	86.450	12.037
72-74	83	87.482	9.754
75-77	5	82.000	9.879
78-80	3	79.000	18.019
TOTAL	2866	83.389	14.271

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

[2] Maximum score = 100.

TABLE 10

DISTRIBUTION OF BROWN SELF-CONCEPT ADJUSTED SCORES FOR MALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score [2]	S.D.
36-38	---	------	------
39-41	1	42.000	------
42-44	7	77.714	9.300
45-47	42	77.619	12.307
48-50	131	79.824	17.469
51-53	249	90.418	14.471
54-56	233	82.820	15.740
57-59	228	83.702	13.828
60-62	200	85.735	11.878
63-65	114	84.667	12.049
66-68	139	85.669	11.276
69-71	91	85.637	12.397
72-74	44	86.682	8.849
75-77	3	89.000	5.657
78-80	2	90.500	9.500
TOTAL	1484	83.201	14.002

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

[2] Maximum score = 100.

TABLE 11

DISTRIBUTION OF BROWN SELF-CONCEPT ADJUSTED SCORES FOR FEMALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	83.333	5.907
39-41	3	70.667	20.336
42-44	13	71.538	14.425
45-47	24	72.625	16.391
48-50	121	75.967	20.880
51-53	210	82.186	16.562
54-56	225	83.111	13.398
57-59	222	82.914	13.644
60-62	165	88.461	10.692
63-65	130	85.915	10.692
66-68	115	86.270	11.954
69-71	109	87.128	11.684
72-74	39	88.385	10.611
75-77	2	71.500	3.500
78-80	1	56.000	------
TOTAL	1382	83.590	14.553

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

[2]Maximum score = 100.

TABLE 12

DISTRIBUTION OF BROWN SELF-CONCEPT ADJUSTED SCORES FOR ALL CHILDREN WITH NO PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE

Age (Months)	N [1]	Mean Score [2]	S.D.
36-38	---	------	------
39-41	---	------	------
42-44	2	88.000	2.000
45-47	6	74.500	7.762
48-50	34	83.706	11.382
51-53	62	80.419	16.195
54-56	57	82.667	12.822
57-59	76	83.316	13.336
60-62	121	87.471	10.742
63-65	94	85.064	10.352
66-68	96	85.458	12.341
69-71	94	87.351	10.920
72-74	35	88.029	11.300
75-77	3	89.000	5.657
78-80	2	68.500	12.500
TOTAL	682	85.013	12.314

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

[2] Maximum score = 100.

TABLE 13

DISTRIBUTION OF BROWN SELF-CONCEPT ADJUSTED 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	3	83.333	5.907
39-41	4	63.500	21.546
42-44	17	72.706	13.091
45-47	59	75.780	14.691
48-50	212	77.217	20.321
51-53	382	81.641	15.029
54-56	388	82.848	14.944
57-59	361	83.568	13.515
60-62	232	86.858	11.660
63-65	141	85.631	11.836
66-68	152	86.099	11.079
69-71	103	85.573	13.041
72-74	48	87.083	8.428
75-77	2	71.500	3.500
78-80	1	100.000	-----
TOTAL	2105	82.960	14.737

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

[2] Maximum score = 100.

TABLE 14

DISTRIBUTION OF BROWN SELF-CONCEPT ADJUSTED SCORES FOR WHITE CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	------	------
39-41	3	55.667	19.328
42-44	8	67.875	10.729
45-47	28	74.143	14.577
48-50	91	78.846	16.456
51-53	180	82.267	14.241
54-56	205	84.580	13.424
57-59	187	84.257	12.322
60-62	160	87.150	11.689
63-65	91	84.253	13.159
66-68	97	85.835	11.880
69-71	73	84.466	11.588
72-74	49	88.898	8.274
75-77	3	89.000	5.657
78-80	2	68.500	12.500
TOTAL	1177	83.880	13.460

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

[2]Maximum score = 100.

TABLE 15

DISTRIBUTION OF BROWN SELF-CONCEPT ADJUSTED SOCRES FOR BLACK CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	83.333	5.907
39-41	1	87.000	------
42-44	12	77.583	13.263
45-47	38	77.026	13.676
48-50	131	77.183	20.848
51-53	200	81.430	16.101
54-56	191	81.099	15.803
57-59	182	82.148	13.776
60-62	129	86.349	10.631
63-65	101	85.970	10.305
66-68	104	85.548	11.268
69-71	103	87.854	11.529
72-74	32	85.219	11.516
75-77	2	71.500	3.500
78-80	1	100.000	------
TOTAL	1230	82.741	14.826

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

[2] Maximum score = 100.

TABLE 16

DISTRIBUTION OF BROWN SELF-CONCEPT ADJUSTED 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	81.391	20.818
51-53	70	77.957	16.816
54-56	52	84.423	12.546
57-59	66	84.148	13.776
60-62	67	87.597	11.659
63-65	50	86.300	9.667
66-68	50	87.160	11.895
69-71	20	86.650	14.434
72-74	---	------	------
75-77	---	------	------
78-80	---	------	------
TOTAL	398	84.327	14.674

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

[2]Maximum score = 100.

Score and Item Characteristics

A frequency distribution of the Brown unadjusted scores for Fall 1971 HSPV sample is found in Table 17. This distribution is negatively skewed and shows a ceiling effect, since 18.5% of the children (N= 3067) scored at the top two scores. In Spring 1972, the scores were also negatively skewed; 23.6% of the children scored at the top two scores and 37.0% scored at the top three scores (N = 853). Distributions for the Brown adjusted scores for Fall 1971 and Spring 1972 were also negatively skewed. In Fall 1971, 17.8% of the children (total N = 3067) scored 100, 0% scored 96-99, and 23.3% scored 90-95. Shipman (1971, 1972) also found that the self concept scores were high and negatively skewed, indicating a strong tendency to select positive attributes.

The mean number of items omitted decreased with age in both the ETS Longitudinal Study and the HSPV Study. The mean number of items omitted for the total Fall 1971 HSPV sample was 2.262 (S.D. = 3.099) (See Table 18). This was higher than both the Year 1 (mean = 1.5, S.D. = 2.97) and Year 2 (mean =.5, S.D. = 1.48) means found in the ETS Study (Shipman, 1972). Results from both studies show that the test is more difficult for younger children to complete.

The frequency distribution of the number of items omitted for the Fall 1971 HSPV sample (Table 19) is a positively skewed distribution with 40.6% of the children omitting no items and 16.8% omitting only one item. In Spring 1972, 50.1% of the children (N = 853) omitted no items. Because of this

TABLE 17

FREQUENCY DISTRIBUTION OF BROWN SELF-CONCEPT SCORES FOR FALL 1971 HSPV SAMPLE[1]

Score[2]	# of Children	%	x = nearest 10 children
0	13	.4	x
1	14	.5	x
2	22	.7	xx
3	28	.9	xxx
4	46	1.5	xxxxx
5	73	2.4	xxxxxxx
6	90	2.9	xxxxxxxxx
7	142	4.6	xxxxxxxxxxxxx
8	140	4.6	xxxxxxxxxxxxx
9	200	6.5	xxxxxxxxxxxxxxxxxxx
10	248	8.1	xxxxxxxxxxxxxxxxxxxxxxxx
11	298	9.7	xxxxxxxxxxxxxxxxxxxxxxxxxxxx
12	347	11.3	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
13	406	13.2	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
14	431	14.1	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
15	369	12.0	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
16	200	6.5	xxxxxxxxxxxxxxxxxx
Total	3067		

[1] Includes PV and non-PV children.

[2] Score = # correct; i.e., unadjusted score.

TABLE 18

MEAN AND STANDARD DEVIATION FOR NUMBER OF ITEMS OMITTED ON BROWN SELF-CONCEPT TEST FOR FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	2.333	3.300
39-41	4	3.750	3.345
42-44	20	3.850	3.482
45-47	60	3.167	3.370
48-50	252	2.794	3.675
51-53	459	2.895	3.263
54-56	450	2.342	3.248
57-59	450	2.360	3.045
60-62	305	1.952	2.715
63-65	244	1.885	2.603
66-68	254	1.709	2.351
69-71	200	1.595	2.293
72-74	83	1.398	1.983
75-77	5	.....	.....
TOTAL	2863	2.262	3.099
Males	1432	2.290	3.145
Females	1331	2.231	3.048

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

[2]Maximum # = 16.

TABLE 19

FREQUENCY DISTRIBUTION FOR NUMBER OF OMITTED ITEMS ON BROWN SELF-CONCEPT TEST FOR FALL 1971 HSPV SAMPLE[1]

# of items omitted	# of children	%	x = nearest 50 children
0	1245	40.6	xxxxxxxxxxxxxxxxxxxxxxxx
1	516	16.8	xxxxxxxxxx
2	307	10.0	xxxxxx
3	195	6.4	xxxx
4	145	4.7	xxx
5	140	4.6	xxx
6	135	4.4	xxx
7	87	2.8	xx
8	84	2.8	xx
9	83	2.7	xx
10	36	1.2	x
11	34	1.1	x
12	18	.6	
13	10	.3	
14	16	.5	
15	9	.3	
16	7	.2	
N =	3067		

[1] Includes PV and non-PV children.

skewed distribution and floor effect, the number of items omitted was not selected as a Brown Self-Concept score for further analyses.

Item analyses also reveal that most items were fairly easy for most children. The percent of children passing each item, based on the total number of items in the test and based on the scorable answers only, are presented in Tables 20 and 21 for the HSPV Fall 1971 sample. For ages 4, 4½, 5, and 5½, 50% of the children or more passed every item (see Table 20). Using only scorable answers, the percent of children passing each item is much higher (see Table 21).

R-biserials for the unadjusted total score were generally high in the ETS Study (Shipman, 1972). In Year 1, they ranged from .42 - .64 with nine over .60; in Year 2, they ranged from .40 - .79 with eleven over .60. The lowest in each case was for item #6 (likes to have other kid's things vs. own things).

In the Fall 1971 Follow Through pilot study (Emrick, 1972), a 21-item version of the Brown was given to kindergarten and entering first grade children in 17 projects. In general the unadjusted scores were quite high, indicating potential ceiling effects of some projects. Detailed item analyses for each project indicated that responding was uniformly positive and high.

TABLE 20

BROWN IDS SELF-CONCEPT TEST: PERCENT PASSING EACH ITEM BASED ON TOTAL NUMBER OF ITEMS IN TEST[1]

	AGES[2]				
Items	3½	4	4½	5	5½
1	44	64	69	71	72
2	74	86	88	81	93
3	74	85	88	90	90
4	51	60	67	72	74
5	44	55	60	60	60
6	62	71	75	81	78
7	81	85	88	90	91
8	51	55	66	76	81
9	25	55	62	68	70
10	74	68	71	75	78
11	44	59	65	71	75
12	44	64	69	74	74
13	37	56	60	69	72
14	40	62	67	72	72
15	48	61	63	70	68
16	51	59	58	66	73
N =	27	483	880	772	477

[1]Unscorable or omitted items are included in the base of items from which the percent passed is computed.

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

TABLE 21

BROWN IDS SELF-CONCEPT TEST: PERCENT PASSING EACH ITEM

BASED ON SCORABLE ANSWERS[1]

	AGES[2]				
Items	3½ (n)	4 (n)	4½ (n)	5 (n)	5½ (n)
1	54 (22)	75 (416)	77 (793)	77 (706)	78 (443)
2	76 (26)	88 (468)	90 (861)	93 (760)	94 (470)
3	83 (24)	93 (442)	87 (833)	94 (741)	95 (450)
4	77 (18)	82 (354)	88 (669)	92 (604)	91 (388)
5	80 (15)	75 (352)	80 (663)	77 (608)	76 (379)
6	77 (22)	83 (412)	86 (774)	87 (721)	85 (441)
7	84 (26)	90 (460)	92 (846)	94 (746)	93 (466)
8	54 (22)	67 (394)	77 (753)	85 (693)	89 (432)
9	46 (15)	70 (380)	77 (715)	80 (658)	81 (410)
10	100 (20)	91 (362)	92 (677)	95 (609)	94 (393)
11	70 (17)	77 (368)	81 (708)	85 (647)	87 (411)
12	52 (23)	71 (431)	75 (816)	78 (737)	75 (468)
13	47 (21)	64 (424)	67 (794)	75 (717)	76 (450)
14	100 (11)	86 (350)	88 (672)	90 (614)	90 (383)
15	92 (14)	80 (371)	82 (675)	87 (625)	83 (391)
16	77 (18)	81 (350)	80 (645)	85 (596)	86 (404)
Mean N	19.6	395.9	743.4	673.4	423.7

[1] Percent based only on items answered; no unscorable or omitted items are included in the base from which the percent is computed.

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

Table 22 presents the intercorrelations of some possible Brown scores for all the children in the Fall 1971 HSPV sample. The correlation between the adjusted and unadjusted scores is .62. Whether or not the child smiled when his picture was taken correlated lowly (around .10) with all other scores. The number of items omitted correlated -.17 with total adjusted score and -.85 with the total unadusted score.

In the ETS Study the correlation between the adjusted and unadjusted score was .83 in Year 1 and .93 in Year 2. Smiling or not correlated lowly with all other scores. The number of items omitted correlated -.15 in Year 1 and -.22 in Year 2 with the adjusted score, and -.67 in Year 1 and -.56 in Year 2 with the unadjusted score (Shipman, 1972).

Reliability

The KR-20's for the Brown unadjusted score using the Fall 1971 HSPV sample are listed in Table 23. The KR-20 for the total sample (n = 3068) was .723. These estimates for 92 subsamples with a size greater than 20 ranged from .568 for old white females with no previous preschool experience (n = 208) to .820 for young male Mexican-American children with no previous preschool experience (n = 76). 93.5% of the KR-20's were between .60 and .79. Since the adjusted total score is a percentage score based on a different number of items for each child, KR-20 calculations are not appropriate.

TABLE 22

INTERCORRELATIONS OF BROWN SELF-CONCEPT SCORES FOR FALL 1971 HSPV SAMPLE[1]

	Total Score Adjusted	Total Score Unadjusted	Number Items Omitted	Number Items Answered
Total Score Unadjusted	.62			
Number Items Omitted	-.17	-.85		
Number Items Answered	.17	.85	-1.00	
Child Smiled	.10	.14	- .10	.10

N = 3067 for all correlations, except for those with child smiled where N = 2946.

[1]Sample includes PV and non-PV children.

TABLE 23

KR-20 RELIABILITIES FOR FALL 1971 BROWN UNADJUSTED SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	3068	11.450	3.348	.723
Black	1309	11.409	3.480	.742
White	1255	11.716	3.086	.685
Mexican-American	436	10.812	3.644	.759
Male	1541	11.445	3.364	.724
Female	1450	11.555	3.280	.716
Young[3]	1334	10.774	3.641	.748
Old	1637	12.097	2.911	.667
Previous Preschool	738	12.114	2.900	.667
No Previous Preschool	2247	11.294	3.421	.729

[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.

The alpha coefficient for the total unadjusted score in the ETS Study was .64 in Year 1 ($\hat{n}$ = 1372) and .59 in Year 2 (n - 1299) (Shipman, 1972).

The KR-20 for the total number of omitted items in the Fall 1971 HSPV sample was .813 (n = 3067). The alpha coefficient for the same score in the ETS Study was .91 in Year 1 (n = 1441) and .84 in Year 2 (n = 1314) (Shipman, 1972).

Internal reliability and test-retest reliability estimates were also obtained in the Fall 1971 Follow Through pilot study (Emrick, 1972). The overall initial test reliability coefficient (KR-20) was .816 (range = .478 to .961); the overall coefficient for the retests several weeks later was .787 (range = .143 to .945). The test-retest coefficient for the total sample (n = 632) after a 2-3 week interval was .545.

Correlations with Other Tests

The correlations of the Brown with the other tests in the Fall 1971 HSPV battery are presented in Table 24. All of the correlations were low. The highest correlations with the Brown unadjusted score were .323 with the 32-item PSI and .322 with the PPVT. The highest correlations with the Brown adjusted score was .259 with the 32-item PSI.

In the ETS Longitudinal Study (Shipman, 1972) correlations with the adjusted total score were quite low (.01 -.19) in both years. In Year 1, the highest correlations were .19 with the PPVT and .17 with the Preschool Embedded Figures Test.

TABLE 24 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	.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	.283 (1075)	.358 (1097)	.293 (1097)	.196 (1097)	.271 (1097)	.363 (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 (2679)				
MI-TRUCK[3]	.174 (607)	.061 (625)	.085 (625)	.083 (625)	.121 (625)	.056 (625)	.164 (608)	.032 (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)	.303 (1096)	.322 (1032)	.415 (1032)	.200 (1032)	.180 (1032)	.212 (1113)	.183 (1113)	.005 (573)		
EIGHT-BLOCK REASON	.445 (1119)	.364 (1148)	.353 (1145)	.286 (1148)	.372 (1148)	.390 (1148)	.443 (1090)	.418 (1096)	.405 (1032)	.402 (1032)	.258 (1032)	.211 (1032)	.178 (1113)	.166 (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.

In Year 2 the highest correlation was .15 with Form B of the PPVT, a productive language measure.

In the Fall 1971 Follow Through pilot study (Emrick, 1972), the Brown unadjusted score was correlated with the other measures for both the test and retest sessions. Correlations with a 29-item experimental version of the PSI were .293 (test) and .378 (retest); correlations with the ITPA Verbal Expression Subtest were .248 (test) and .314 (retest); correlations with Faces, an inventory of pupil's attitudes to self, home and school, were .229 (test) and .257 (retest). Results of all these studies indicating low associations with other tests may mostly be due to the dearth of other affective measures in the batteries.

Remarks

The technical findings of the HSPV Study are generally consistent with the Follow Through 1971 pilot study (Emrick, 1972) and the ETS Longitudinal Study (Shipman et al., 1971; Shipman, 1972). Reliability estimates for the Brown unadjusted scores are acceptable but item analyses reveal there are ceiling effects and tendencies for children to respond in a socially desirable way. Due to these contradictions and uninterpretable findings, SRI (Emrick, 1972) concluded that the Brown in its present form not be used in future large scale evaluations.

In the theoretical realm there are questions concerning the Brown's validity, since one cannot be sure if it is the child's self-concept that is actually being measured. Many of the items require a high degree of verbal comprehension and syntax understanding, indicating that the Brown may be more of a cognitive test -- and more specifically, a vocabulary test -- for younger children. Children in the HSPV Study consistently omitted about 12% of the responses. They had slightly more difficulty with items involving a quantity (such as item #9 -- scared of a lot of things, and item #11 -- scared of a lot of people) and negatives (such as item #4 -- deesn't like to play with other kids, and item #14 -- doesn't like the way his face looks). From inspecting test protocols Shipman (1972) found that there were uneven differences in wording of items and that double negatives were particularly difficult for young children. In additions, some items are ambiguous and culturally biased; for example, item #5 -- likes to talk a lot or doesn't -- consistently had a relatively low percentage of children passing it correctly in the HSPV data. The correct response -- likes to talk a lot -- is not the more socially desirable response in all situations and/or in all subcultures of the American population. The item's connection to a child's self-concept is ambiguous.

As a child becomes older Shipman (1972) found that the self-concept factor obtained in factor analyses no longer correlated with the general intellectual competency factor in Year 2 as it had in Year 1. Thus, the Brown may be a

cognitive test for young children and a more affective measure for older children. Even if this were true, there is still the problem of response biases towards the perceived socially acceptable response when older children are given the test; this is substantiated in the Follow Through data with older children.

Future reports from the ETS Longitudinal Study may show if the three stage developmental pattern Shipman (1972) hypothesizes exists: (1) self-concept is not clearly differentiated and so cannot be assessed by a verbal report measure; (2) self-concept is differentiated along a global "good-bad" continuum; and (3) self-concept is differentiated along a variety of dimensions. If this pattern is correct, scores would increase in stage 2 and then decrease with maturation and experience in stage 3.

It will also be interesting to see if the supplementary scoring used in Year 2 of the ETS Study will further explicate the meaning of the Brown Self-Concept Test. This supplementary scoring was developed to capture the child's initial response pattern to each Brown item: verbalize one specific alternative, verbalize both alternatives, qualified answer, said "yes" or "no" only, non-verbal response only, verbal and non-verbal response, no response, substituted another task-related response, irrelevant response (Shipman, 1972).

Future research into the development and meaning of self-concept of young children must be done before valid and meaningful self-concept instruments can be developed or existing instruments can be adapted into a meaningful and

valid framework. Because of these theoretical problems and the conflicting technical findings reported for the Brown Self-Concept Test, it is recommended that the Brown not be used in this form in future large scale evaluations.

References

Brown, B. The assessment of self-concept among four-year-old Negro and white children: A comparative study using the Brown IDS Self-Concept Referents Test. Paper presented at the Eastern Psychological Association Meeting, New York City, April, 1966.

Buros, O.,K. (Ed.) Personality tests and reviews. Highland Park, N. J.; Gryphon, 1970.

Clark, E. T., Ozenhosky, R. J., Barz, A. K., & O'Leary, J. V. Self-concept and vocabulary development in Negro and white preschool children. Paper presented at the Eastern Psychological Association Meeting, April, 1967.

Coller, A. R. Self-concept measures; An annotated bibliography. Princeton, N. J.; Head Start Test Collection Report, Educational Testing Service, 1971.

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, HEW under contract #OEC-0-8-522480-4633, Menlo Park, Calif. SRI, 1972.

LaCrosse, E. R., Lee, P. C., Litman, F., Ogilvie, D., Stodolsky, S., & White, B. L. The first six years of life: A report on current research and educational practice. Genetic Psychology Monographs, 1970, 82(2), 161-266.

Shipman, V., 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. J.: Educational Testing Service, 1971.

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

Walker, D. K. Socio-emotional measures for preschool and kindergarten children: A Handbook. San Francisco: Jossey-Bass, 1973.

Wylie, R. The self-concept: A critical survey of pertinent research literature. Lincoln, Nebraska: University of Nebraska Press, 1961.

California Preschool Social Competency Scale

Purpose

The California Preschool Social Competency Scale (CPSCS) is a rating scale designed to "measure the adequacy of preschool children's interpersonal behavior and the degree to which they assume social responsibility" (Levine et al., 1969, p. 3). Many psychologists and educators state that the development of a child's social skills is related to a child's personality and mental development. Many preschool programs focus on the importance of a child's social competence and related social skills. Studying the development of a child's social skills is key to understanding the socialization processes operating within one's environment.

Description

The scale consists of 30 items which cover a wide range of behaviors such as response to routine, response to the unfamiliar, following instructions, making explanations, sharing, helping others, initiating activities, giving direction to activities, reaction to frustration, and accepting limits (Table 1). Each item contains four descriptive statements which represent varying levels of competence for the particular behavior. A teacher or any other adult who has had an opportunity to observe the child

TABLE 1

NAMES OF ITEMS FROM THE CPSCS

1. Identification
2. Using Names of Others
3. Greeting New Child
4. Safe Use of Equipment
5. Reporting Accidents
6. Continuing in Activities
7. Performing Tasks
8. Following Verbal Instructions
9. Following New Instructions
10. Remembering Instructions
11. Making Explanation to Other Children
12. Communicating Wants
13. Borrowing
14. Returning Property
15. Sharing
16. Helping Others
17. Playing With Others
18. Initiating Involvement
19. Initiating Group Activities
20. Giving Direction to Play
21. Taking Turns
22. Reaction to Frustration
23. Dependence Upon Adults
24. Accepting Limits
25. Effecting Transitions
26. Changes in Routine
27. Reassurance in Public Places
28. Response to Unfamiliar Adults
29. Unfamiliar Situations
30. Seeking Help

in a variety of situations places the child's habitual or typical behavior into one of the four levels. On each item, level one represents the lowest level of competence and level four the highest. Each item is rated independently and no special test situation is required. Each test item is a Guttman scale, meaning that the levels are cumulative -- i.e., a child who performs at level 4 is assumed to be able to perform at all the preceeding levels. The total social competency raw score (maximum score = 120) is the sum of all the level ratings for the 30 items. A child's raw score can be translated into percentile scores according to age, sex and occupational level of parents (see manual, 1969). For the HSPV analysis, a child's score was used only if 27 or more items were completed by the teacher.

Development of Instrument

The CPSCS was developed in 1969 by Levine, Elzey, and Lewis at San Francisco State College (1969); portions of the scale are adopted from the Cain-Levine Social Competency Scale of 1963. The standardization of items and the determination of norms were based on teacher ratings of children who were attending preschool. The behaviors selected were those that one would expect to be developed in the process of socialization of all preschoolers of all socioeconomic groupings.

In developing the scale and selecting behaviors, the degree of the child's independence was the primary focus. To be selected as an item for the CPSCS, the item must have

been observable within the context of a preschool environment, age differentiable, applicable to both sexes, unidimensional in content, easily scaled into four levels, and judged important to a child's social competency development. Based on the judgments and criticisms of mainly "teacher groups enrolled in graduate programs in early childhood education," an initial form of 34 items was devised and given to 1,165 two-to-five-year-old children in California. From statistical analyses of the items on this initial form, the 30 item scale used in HSPV was developed.

Standardization

The original norming sample for the CPSCS is based on 800 children equally divided according to chronological age, sex and parent's occupational level (Levine, Elzey & Lewis, 1969). The norming sample approximates the proportion of preschool children in major urban centers for each geographic region of the United States. Prior to establishing norms for the final form of the scale, a three way analysis of variance for variables of sex, age, and occupational level showed that all main effects were significant beyond the .01 level while none of the interactions were significant.[1] Therefore, separate norms were established for each of the four age groups (2-6 to 2-11, 3-0 to 3-11, 4-0 to 4-11, 5-0 to 5-6) by sex and occupational level. The mean and standard deviation of the raw scores at each age level for each group were

[1] Size of effects was not stated.

used for the computation of the percentile norms. Correlations between age and social competence are higher for high occupational levels (male = .51; female = .49) than for low occupational levels (male = .38; female = .39). In general, as children get older their scores on the CPSCS increase so that there is a ceiling effect for the last two age norms.

Finally, correlations between the individual items and total competency score were computed for each sex at high and low occupational levels. In general, these correlations were comparable across the four different groupings. Over eighty percent of the items showed correlations of .50 or above with the total score.

CPSCS norms for raw scores are also available for the Fall 1970 HSPV sample (Tables 2-8). 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 (Table 2), females (Table 3), males (Table 4), children with no previous preschool experience (Table 5), children with previous preschool experience (Table 6), black children (Table 7), and white children (Table 8).

TABLE 2

DISTRIBUTION OF CPSCS SCORES FOR ALL CHILDREN IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	9	73.111	9.085
45-47	66	76.364	15.959
48-50	214	75.813	13.829
51-53	323	76.808	18.566
54-56	356	80.003	16.970
57-59	351	80.396	18.490
60-62	263	80.932	18.236
63-65	231	81.095	16.890
66-68	175	84.486	15.379
69-71	175	84.600	17.466
TOTAL	2163	80.000	17.423

[1] Includes all children; not in Level I sites, Oraibi, or Fresno; who had adequate information on sex, age, race, preschool experience, and 27 or more of the items.

[2] Maximum score = 120

TABLE 3

DISTRIBUTION OF CPSCS SCORES FOR FEMALES IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	4	74.250	9.203
45-47	33	81.000	11.893
48-50	103	76.476	16.559
51-53	154	77.630	18.465
54-56	185	82.114	17.110
57-59	160	82.519	18.098
60-62	134	80.522	20.129
63-65	119	81.966	17.038
66-68	86	87.337	14.439
69-71	84	85.905	13.816
TOTAL	1062	81.419	17.864

[1]Includes children not in Level I sites, Omaha or Fresno, who had adequate information on sex, age, race, preschool experience, and 27 or more of the test items.

[2]Maximum score = 120.

TABLE 4

DISTRIBUTION OF CPSCS SCORES FOR MALES IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	5	72.200	8.886
45-47	33	71.727	18.026
48-50	111	75.198	10.652
51-53	169	76.059	18.626
54-56	171	77.719	16.516
57-59	191	78.618	18.627
60-62	129	81.857	16.024
63-65	112	80.170	16.681
66-68	[illegible]	81.730	15.753
69-71	[illegible]	83.396	16.025
TOT[illegible]	1101	78.630	16.874

[1]Includes children; not in Level I sites, or Oraibi, or Fresno; who had adequate information on sex, age, race, preschool experience, and 27 or more of the test items.

[2]Maximum score = 120.

TABLE 5

DISTRIBUTION OF CPSCS 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	9	73.111	9.085
45-47	60	75.417	15.596
48-50	191	75.607	14.035
51-53	296	76.466	18.943
54-56	295	78.851	16.880
57-59	287	79.561	18.812
60-62	196	79.929	17.543
63-65	162	80.414	15.563
66-68	128	84.188	16.020
69-71	129	84.109	18.435
TOTAL	1753	79.106	17.447

[1] Includes children; not in Level I sites, Oraibi, or Fresno; who had adequate information on sex, age, race, preschool experience, and 27 or more of the test items.

[2] Maximum score = 120.

TABLE 6

DISTRIBUTION OF CPSCS 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	85.833	16.446
48-50	23	77.522	11.843
51-53	27	80.556	13.211
54-56	61	85.574	16.283
57-59	64	84.141	16.456
60-62	67	83.866	19.837
63-65	69	82.696	19.562
66-68	47	85.298	13.445
69-71	46	85.978	14.316
TOTAL	410	83.822	16.790

[1] Includes children not in Level I sites, Oraibi, or Fresno; who had adequate information on sex, age, race, preschool experience, and 27 or more of the test items.

[2] Maximum score = 120.

TABLE 7

DISTRIBUTION OF CPSCS SCORES FOR BLACK CHILDREN IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	3	78.667	6.549
45-47	41	74.049	16.251
48-50	130	76.238	14.242
51-53	172	77.384	13.550
54-56	194	80.598	15.632
57-59	193	79.368	17.885
60-62	106	79.179	15.462
63-65	111	78.559	17.264
66-68	84	82.357	15.419
69-71	102	85.118	16.499
TOTAL	1136	79.366	16.027

[1] Includes children; not in Level I sites, Oraibi, or Fresno; who had adequate information on age, sex, race, preschool experience, and 27 or more of the test items.

[2] Maximum score = 120.

TABLE 8

DISTRIBUTION OF CPSCS SCORES FOR WHITE CHILDREN IN THE FALL 1970 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
42-44	6	70.333	8.901
45-47	25	80.160	14.699
48-50	75	74.973	13.427
51-53	121	78.455	19.530
54-56	126	79.833	18.845
57-59	115	81.017	19.124
60-62	104	83.529	17.325
63-65	86	83.616	17.245
66-68	59	87.898	13.861
69-71	51	82.725	20.453
TOTAL	768	80.991	18.057

[1] Includes children; not in Level I sites, Oraibi, or Fresno; who had adequate information on sex, race, age, preschool experience, and 27 or more of the test items.

[2] Maximum score = 120.

Reliability

Reliability data for the CPSCS is reported in the manual for three studies (Levine et al., 1969). In Texas, independent ratings by classroom teachers, by the director of the program and by the consultant to the program were obtained on 24 children. In Minnesota independent teacher-director ratings were obtained on 15 children, and in California independent ratings of teachers and assistant teachers were obtained on 71 children in six summer Head Start programs. The Pearson "r" reliability coefficients for these studies ranged from .75 to .86 (average = .79). These are probably conservative estimates since interjudge differences in the use of the scale and knowledge of the children being rated were not considered. Odd-even reliability coefficients computed by age, sex and parent's occupational level for the norming sample were from .90 to .98, after correcting by the Spearman-Brown formula.

In the Fall 1970 HSPV sample the odd-even reliability coefficient, after correcting by the Spearman-Brown phrophecy formula, was .962 for 3857 children. The coefficient for two random half correlations for the same sample was .951.

Validity

Table 9 presents the correlations of the CPSCS with the other tests of the Fall 1970 HSPV battery. The highest correlations were in the .30's: .390 with the 64-item PSI,

TABLE 9

INTERCORRELATIONS OF FALL 1970 SCORES FROM THE CPSCS, NYU BOOKLETS 3D AND 4A, 64-ITEM PSI, MI SUBTESTS, LIGHT-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	.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)	.068 (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)	.160 (552)	.356 (556)	.110 (276)	.212 (278)	.097 (273)	.159 (276)	.739 (576)	.924 (576)		
IQ	.321 (752)	.427 (753)	.365 (749)	.510 (752)	.152 (352)	.229 (352)	.128 (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).

.373 with MA (Mental Age), and .321 with the Stanford-Binet IQ. All of these correlations have questionable usefulness since the CPSCS scores were pooled across teachers who seem to be using different rating criteria.

Remarks

Even though the norming procedures of the CPSCS seem quite adequate, the procedure by which the test was constructed and the content of items are questionable. First, the item selection procedure needs to be further elaborated by the authors. Were the "teacher groups enrolled in graduate programs in early childhood education" in-service teachers or graduate students with no teaching experience? In addition, it is difficult to see any consistent criteria for the selection of items and there is no clearly defined theoretical structure underlying the specification of behavior to be rated on the scale.

Second, the content of some of the items is questionably culture-bound. One such item is "Greeting new child" (item #3). This item may be culturally biased since level 4 ("He nearly always makes verbal contact with child without physical contact") is assumed more competent than level 3 ("He makes a limited and brief physical contact with child and some verbal contact"). The defined "more competent" behavior reflects the white, middle class norm in the United States. Many subcultures of the American population, such as South

Americans and Italians, highly value level 2 and level 3 behaviors; perhaps these behaviors should be defined as "most competent" for children from these subculture groups.

There is very little validity information and no inter-judge reliability data for the Head Start Planned Variation sample. What little data exists on the latter in the manual is based on small samples. Since it appears that teachers have initially different average biases, it is recommended that the CPCSS not be used to compare across classrooms. It is impossible to meaningully pool ratings across teachers who seem to be using different "frames of references" (i.e., different criteria) in their ratings. Undoubtedly, such analyses would include unknown biases that could not be explained and might be misinterpreted. Until some of the problems outlined here are solved, this scale should not be used for summative evaluation purposes but may be effectively used in formative evaluation efforts.

Reference

Levine, S., Elzey, F. F., & Lewis, M. California Preschool Competency Scale Manual. Palo Alto, Calif.: Consulting Psychologists Press, Inc., 1969.

Classroom Behavior Inventory[1]

Purpose

The Classroom Behavior Inventory (CBI) is designed to assess children's behaviors in the socio-emotional domain. The three specific traits of social and personality development that this rating instrument measures are task orientation, extraversion, and hostility. Growth and development of social skills is an important aspect of most all preschool programs.

Description

The Classroom Behavior Inventory is a 15 item Likert-type rating scale of children's behaviors. For each of the 15 items in the inventory, a rater (usually the teacher) indicates the degree to which the behavior described in the item is characteristic of the child being rated. The child's behavior is placed on a seven point scale: never, almost never, occasionally, half the time, frequently, almost always, always. The following is a list of the items in the order of presentation to the raters:

1. Pays attention to what he's doing when other things are going around him.
2. Tries to be with another person or group of people.
3. Gets impatient or unpleasant if he can't get what he wants when he wants it.
4. Stays with a job until he finishes it.
5. Likes to take part in activities with others.
6. Slow to forgive when offended.
7. Becomes very absorbed in what he is doing.

---

[1]Also called the "Shaefer Behavior Inventory" in other sources.

8. Enjoys being with others.
9. Stays angry for a long time after a quarrel.
10. Works earnestly at his classwork. Doesn't take it lightly.
11. Seeks contact with others.
12. Complains or whines if he can't get his own way.
13. Watches carefully when a teacher or classmate is showing how to do something.
14. Does not wait for others to approach him, but makes the first friendly move.
15. Angry when he has to wait his turn or share with others.

Even though the items are presented in a random order to the rater, they actually fall into three basic categories: task orientation (1,4,7,10,13); extraversion (2,5,8,11,14); and hostility (3,6,9,12,15). Each item is scored on a seven point scale from "1" for "never" to "7" for "always". Thus, a low score represents an infrequent manifestation of the trait measured.

Development of Instrument

The Classroom Behavior Inventory was developed by E. Schaefer, M. Aaronson and V. Small from Schaefer and Droppleman's Classroom Behavior Checklist, which has been used in the United States and Europe (Schaefer, 1971). The original checklist was created from Schaefer's circumplex model of child behavior (Schaefer, 1961). The original preschool/primer version of the inventory consisted of 60 items representing twelve different behavior traits from verbal expressiveness, hyperactivity, and kindness to resentfulness. In a factor analysis of scores of 1579 Head Start children

on the 60 item inventory, three bipolar factors emerged: lovable-hostile, extraversion-introversion, and task orientation-distractibility. From the 60 item inventory data a 25 item inventory with a five point rating scale was developed and administered by teachers in two schools to approximately 1500 children in grades K-3. A factor analysis of this study revealed the same three basic bipolar traits. From these data on the 25 item inventory, a 15 item inventory with a five point rating scale, in which all of the items were stated in a positive manner, was developed and administered to approximately 3600 children in six schools in grades K-5. From the results of this analysis three items were replaced on the inventory to form the final version of the inventory that was used in Project FOCUS (Small, 1971).

Head Start Planned Variation Reliability Studies

In the fall of 1971 a CBI test-retest reliability study using the HSPV sample was conducted by Huron Institute and SRI. The details of this study are reported in Appendix C. In general the test-retest reliability coefficients for a two week period were in the .70's for the sample of 464 children from four sites. The correlation coefficients were .760 (task orientation), .740 (extraversion), and .726 (hostility). The reliability estimates calculated from a one way analysis of variance repeated measures (two times)

design were very similar: .754 (task orientation), .737 (extraversion), and .704 (hostility).

As part of this reliability study a principal components analysis followed by a varimax rotation on the pooled seventh and ninth week scores for the total sample was done. This factor analysis revealed that the 15 items do cluster around the three bipolar traits which Schaefer reported and used in developing the rating scale (see Table 1 in Appendix C). Altogether the three factors explained 80.4% of the total variance.

In Fall 1971 Huron Institute conducted a Classroom Behavior Inventory inter-rater reliability study in thirteen Head Start Planned Variation classrooms. The raters were either classroom aides or other paraprofessionals. Product-moment correlations and Spearman rank-order correlations between the two raters are reported in Table 1 for each CBI subtest and classroom. In general, the two correlation coefficients were similar for each classroom. The inter-rater reliability was highest for the Task Orientation subtest (median = .60, .62) and lowest for the Extraversion subtest (median = .46, .49). These moderate to low inter-rater estimates may have been higher if teachers had been the raters!

Table 2 lists the mean rater level for each subtest for each classroom in the Head Start Planned Variation inter-rater reliability study. The large discrepancies between the mean scores for the two raters, especially for the extraversion subtest, indicate that the scores should probably not be aggregated across classrooms for analysis purposes since non-

TABLE 1

CORRELATIONS BETWEEN TWO RATERS ON CLASSROOM BEHAVIOR INVENTORY SUBTESTS[1]

	Task Orientation			Extraversion			Hostility		
Classroom	n	Product-Moment r	Spearman Rank-order r	n	Product-Moment r	Spearman Rank-order r	n	Product-Moment r	Spearman Rank-order r
1	19	.45	.40	17	.61**	.50	17	.10	.08
2	19	.84***	.87	19	.51*	.49	18	.78***	.68
3	14	.39	.41	14	.56*	.53	14	.49	.30
4	11	.62*	.74	11	-.02	.12	11	.16	-.07
5	18	.67**	.60	17	.39	.24	17	.47	.46
6	16	.16	.16	18	-.08	-.15	16	.25	.26
7	16	-.19	-.19	15	.37	.44	16	.14	.32
8	18	.75***	.71	18	.58*	.66	18	.73***	.66
9	17	.60*	.63	17	.20	-.02	17	.78***	.82
10	16	.59*	.60	16	-.15	-.32	16	.60*	.53
11	19	.91***	.92	19	.59**	.75	19	.51*	.64
12	20	.76***	.73	19	.66**	.53	17	.69**	.66
13	17	.66**	.52	18	-.06	.11	16	.35	.31
Median		.62	.60		.39	.44		.49	.46

[1]The pair of raters at each classroom were classroom aides or other paraprofessionals.

Significance: * .05
** .01
*** .001

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR EACH OBSERVER ON THE CLASSROOM BEHAVIOR INVENTORY SUBTESTS

A. Task Orientation
(Max. Score = 35)

Class	Observer 1			Observer 2		
	n	Mean	S.D.	n	Mean	S.D.
1	19	18.3	1.9	19	25.7	7.1
2	19	21.1	10.8	19	26.9	6.0
3	15	17.6	3.3	15	28.0	5.7
4	11	29.0	4.8	11	19.9	8.1
5	18	23.4	5.4	18	22.8	7.1
6	16	25.1	2.8	19	26.2	2.7
7	17	21.9	4.7	16	22.2	6.5
8	18	29.3	6.9	18	27.3	3.7
9	17	24.1	7.9	17	33.8	4.8
10	16	22.4	3.2	16	24.7	4.6
11	19	24.4	6.4	19	24.4	8.7
12	20	23.9	7.2	20	26.5	6.7
13	19	25.1	5.7	18	26.2	7.1

B. Extraversion
(Max. Score = 35)

Class	Observer 1			Observer 2		
	n	Mean	S.D.	n	Mean	S.D.
1	19	19.7	2.8	17	29.1	5.4
2	19	22.9	2.9	19	27.4	3.7
3	15	24.5	5.2	15	30.0	6.0
4	11	32.8	2.0	11	23.5	7.6
5	18	27.1	4.2	17	23.0	4.0
6	18	26.2	3.6	19	24.4	3.2
7	17	23.4	5.3	15	25.2	3.9
8	18	30.1	2.7	18	28.3	4.0
9	17	24.1	4.7	17	33.8	3.1
10	16	26.3	2.8	16	30.8	1.2
11	19	31.2	2.2	19	30.9	5.3
12	20	24.3	3.8	19	27.6	6.6
13	18	27.7	5.1	20	29.1	3.4

(Table 2 cont.)

C. Hostility
(Max. Score = 35)

Class	Observer 1			Observer 2		
	n	Mean	S.D.	n	Mean	S.D.
1	19	11.5	2.7	17	10.0	.8
2	19	11.8	10.1	18	10.1	7.3
3	15	13.5	4.6	15	15.7	6.8
4	11	9.0	4.8	11	15.4	8.1
5	17	16.6	3.7	18	13.7	3.5
6	16	20.1	2.8	19	17.9	7.8
7	17	15.0	4.5	16	15.2	4.1
8	18	9.2	4.9	18	15.1	5.2
9	17	12.2	8.4	17	8.1	4.8
10	16	12.9	4.8	16	12.4	1.9
11	19	14.5	5.9	19	12.3	7.9
12	19	14.9	5.7	17	18.3	9.3
13	16	12.1	5.6	20	13.8	6.9

biased, meaningful statements could not be made.

Home Start Study

The Classroom Behavior Inventory used in the HSPV Study was also used in the pilot year (1972-73) of the Home Start evaluation (Hi-Scope, 1973). A factor analysis of these data produced the same three factors found by Schaefer and in the HSPV reliability study. Altogether the three factors explained 49.7% of the total variance.

Internal consistency reliability (coefficient alpha) estimates calculated for the three subtests from the Home Start pilot data (n = approximately 180) were .72 for Task Orientation, .72 for Extraversion, and .67 for Hostility. These are quite adequate, considering that the subtests have only five items. No inter-rater reliability estimates are reported.

Correlations of items with total test scores were .40 - .51 for Task Orientation, .24 - .67 for Extraversion, and .31 - .57 for Hostility. In every case an item correlated higher with its assigned scale than with either of the other two scales. A close look at the percentage of children rated at each level for each item in the Home Start data reveals that there is a definite ceiling effect for Extraversion scores and a possible one for Task Orientation scores.

Score Characteristics

When the 15 item version with a five point response scale was given to approximately 2200 children in grades

K-4, different patterns for the subtest scores were found (Small, 1971). Average hostility and extraversion scores remained fairly stable for both boys and girls over grades K-4, while task oriented behaviors tended to decrease from grades K-4 and were consistently lower for boys than for girls at all grade levels. Average hostility scores were located near the lowest possible scores while average extraversion scores were located near the highest possible score. Task oriented behaviors were somewhat less extremely distributed.

Similarly, characteristic patterns for each subtest of the 15 item version with a seven point response scale were found across the four sites used in the HSPV test-retest study (Appendix C). At every site the hostility subtest scores were located at the lowest possible scores. Scores of the extraversion and task orientation subtests were somewhat evenly distributed across sites from the middle to the top of the scoring range. There was a definite ceiling effect for each of these two subtests in only one out of four sites.

The means and standard deviations for each of the three CBI subtests for the total Fall 1971 sample and selected subsamples are listed in Table 3. For the total sample (n = 4943) the mean for Task Orientation was 22.446 (S.D. = 7.109); for Extraversion, 24.196 (S.D. = 6.370); and for Hostility, 13.637 (S.D. = 6.661). There are no large differences in mean scores on any subtest for any of the subsamples.

TABLE 3

DISTRIBUTION OF CLASSROOM BEHAVIOR INVENTORY SUBTEST SCORES FOR HSPV FALL 1971 SAMPLE[1]

	n	Task Orientation[2]		Extraversion[2]		Hostility[2]	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
Total	4943	22.446	7.109	24.196	6.370	13.637	6.661
Male	2523	21.242	7.147	24.763	6.428	14.131	6.750
Females	2420	23.701	6.848	25.076	6.305	13.123	6.527
Previous Preschool	1161	22.451	7.409	25.699	6.328	14.230	6.960
No Previous Preschool	3498	22.500	7.029	24.703	6.380	13.471	6.526
White	1706	22.302	7.153	24.574	6.448	13.916	6.626
Black	2355	22.145	7.027	25.223	6.221	13.856	6.732
Mexican-American	513	23.587	7.040	24.889	6.444	12.692	6.383

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

[2] Maximum score = 35.

Validity

In addition to the CBI's face validity, the factor analyses described in previous sections give evidence of the construct validity of the instrument. Correlations of the Classroom Behavior Inventory subtest scores (aggregated across all children) with the other tests of the Fall 1971 HSPV battery (Table 4) are examples of the CBI's concurrent validity. All of the correlations with both cognitive and non-cognitive tests were low. The largest correlation was .29 between the Task Orientation subtest and the 32-item PSI.

For the Fall 1971 HSPV sample (n = 4962), the correlations between the CBI's subtests were .38 (Task Orientation with Extraversion), -.39 (Task Orientation with Hostility), and -.13 (Hostility with Extraversion).

TABLE 4

CORRELATIONS OF CLASSROOM BEHAVIOR INVENTORY SUBTEST SCORES WITH OTHER TESTS OF THE FALL 1971 HSPV TEST BATTERY FOR THE TOTAL FALL 1971 HSPV SAMPLE[1]

	Task Orien. Subtest	Extra-version Subtest	Hostility Subtest	n
PPVT	.24	.16	-.03	2947
WRAT Copy Marks	.24	.11	-.07	2937
" Recog. Letters	.23	.12	-.06	2937
" Naming Letters	.15	.08	-.04	2937
" Reading Nos.	.20	.08	-.05	2937
32-item PSI	.29	.18	-.08	2927
Brown Unadjusted	.13	.10	-.05	2818
Brown Adjusted	.12	.10	-.07	2818
MI-Truck[2]	-.05	-.05	-.01	627
ITPA-Verbal Expres.	.23	.15	-.07	1187
ETS-Total Score[3]	.25	.15	-.10	1113
8-Block Success Total	.22	.20	-.10	1192

[1]Children in sample are those with adequate information not in Level I sites.

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

[3]ETS Enumeration score = sum of counting, touching and same number matching subtest scores.

Remarks

The various factor analyses done on the Classroom Behavior Inventory items convincingly show that the various subtests hold together well and measure independent traits. In addition, the test-retest reliability estimates form the HSPV Study and the internal consistency estimates from the Home Start Study are adequate for subtests of a rating scale. The inter-rater estimates for paraprofessionals are moderate; there is no inter-judge information for teachers or other professionals.

Despite these favorable technical points, the score characteristics of the subtests, including a ceiling effect for Task Orientation and Extraversion scores and a floor effect for Hostility scores, must be seriously considered in interpreting results. The difference in mean rater levels for the subtests in the classrooms of one HSPV reliability study are most distressing, and indicate that aggregation of CBI scores across classrooms and comparisons among classrooms would be uninterpretable at this time.

More information about the validity of the CBI is needed. In addition, more work needs to be done on interpreting various score profiles. Is the child who scores high on Task Orientation and Extraversion and low on Hostility the most developed in these socio-emotional domains? What do different profiles of scores indicate about particular children and particular classrooms?

Since there appears to be a large amount of within classroom variation in scores, it is recommended that the Classroom Behavior

Inventory not be used to compare across classrooms. It is impossible to meaningfully pool ratings across teachers who seem to be using different criteria in rating. Until some of the issues discussed above are resolved, the CBI should not be used as a summative evaluation tool but may be used effectively as a formative tool.

References

Hi-Scope Foundation/Abt Associates. Home Start Evaluation Study. Draft interim report to the Office of Child Development, Department of Health Education and Welfare - #HEW-OS-72-127 Ypsilanti, Michigan; Hi-Scope/Abt, 1973.

Schaefer, E. S. Converging conceptual models for maternal behavior and child behavior. In J. C. Glidewell (Ed.), Parental attitudes and child behavior. Springfield, Illinois: Thomas, 1961.

Schaefer, E. S. Development of hierarchical configurational models for parent behavior and child behavior. In J. P. Hill (Ed.), Minnesota Symposia on Child Psychology. Vol. 5. Minneapolis: University of Minnesota Press, 1971.

Small, V. H. Supervisor, Department of Curriculum and Instruction, Montgomery County Public Schools, Maryland. Personal communication, 1971.

Classroom Information Form

Purpose

The purpose of the Classroom Information Form (CIF) is to obtain demographic data on the children in the Head Start Planned Variation sample.

Description

The following are the variables measured on the CIF:

1. previous preschool experience
2. mother's education
3. mother's occupation and employment status
4. father's education
5. father's occupation and employment status
6. child's age and sex
7. ethnic group
8. language spoken at home
9. days in attendance
10. number of persons in the household
11. income

The forms are filled out by Head Start teachers. The information on the child's family is usually obtained from the application form which the parents fill out. Attendance information comes from the teacher's roster.

Reliability

The reliability of the demographic data was checked against parent reports using a procedure which is reported in Appendix E. The data on previous Head Start experience seem extremely reliable. Across all sites in the reliability study there was 79.9% perfect agreement. The data on education are

better than the data on occupation, and the data on mothers are slightly better than on fathers. The data on employment status are very unreliable. The data on income were not checked. Number of people in the home is an extremely reliable variable while language spoken in the home is extremely unreliable. See Appendix E for more detailed results about the quality of this demographic data.

Remarks

One advantage of the CIF over a parent interview or questionnaire is its low cost. In addition, a teacher has access to the Head Start center's records and does not have to rely solely on first-hand knowledge. If the importance of filling out such a form is stressed and the process monitored closely, the CIF response rate will be high as it was in the last two years of the HSPV study. Thus, if adequate controls are used, important demographic data can be obtained from teachers for very little cost. Even if the response rate is high, however, the reliability of the data will vary depending on the subject area (i.e., income, education, etc.).

Classroom Observation Procedure

Purpose

The Classroom Observation Procedure (CO) is designed to assess the degree of successful implementation of classroom processes and child outcomes from the varying programs.

Description

The Classroom Observation Procedure has several parts:

1. Five Minute Observation. The Five Minute Interaction measure was designed to record the interaction patterns of the classroom. The instrument has categories for recording four basic aspects of an interaction: Who, To Whom, What and How. The Who and To Whom categories include teacher, assistant/aide, volunteer, child, different child, two children, small group, large group, everyone, materials, and confusion. The What categories include direct request, choice request, respond, teach/inform, comment/play, praise, acknowledge, help, cooperate, corrective feedback, no response/ignore/I don't know, refuse/reject, observe, and confusion. The How categories include happy, sad, negative, angry, guide to alternative, reason, control by praising, question, firm, demean, threaten, punish, touch, object, and symbol.

For each Five Minute Observation, the observer focuses on one activity which can consist of children working independently or of the teacher working with either an

individual child or a group of children. All interactions in the selected activity are recorded for five minutes. The standard number of interactions recorded during this interval is 60. The observer first observes each of thirteen specified activities (listed below under the Classroom Checklist) if they occur in the classroom. Next, activities which occur most frequently are selected. Thus, the activities observed reflect the dominant activity of each class as well as representing a wide range of activities.

2. Classroom Checklist. This provides a record, for one point in time, of all the activities and groupings which are going on in the classroom. A checklist is filled out immediately before each Five Minute Observation. The observer records for each child the size of the group he is working in (alone, one child, small group, large group), whether he is working with an adult (teacher, volunteer), and what he is doing (activity). There are thirteen activity groupings: snack lunch, group time/sharing/rest/story/singing/dancing, numbers, reading/alphabet/language development, social studies/geography, science/natural world, guessing games/table games/puzzles, arts/crafts, sewing/cooking/pounding/sawing, blocks/trucks, dolls/dress up, and active play. Each adult in the classroom is also accounted for. Thus the Classroom Checklist, which is filled out approximately 12 times for each observed classroom, gives an overall description of the activities of the classroom.

3. Physical Environment. The observer makes a series of judgments relating to lighting, noise level and seating arrangement for each classroom. This section was only used in Fall 1970. For Spring 1971 data on Observatio. Summary Form, on which the observer records information about enrollment, attendance, class duration and number of adults by role, was included.

Table 1 lists the variables used in analyzing the Spring 1971 CO data (Lukas and Wohlleb, 1973). The variables used in the HSPV analysis are limited to a selected group of those being used by SRI in analyzing the Follow Through data. (SRI, 1972).

Development of Instrument

The SRI Classroom Observation Instrument, developed at SRI under the direction of J. Stallings, is based mainly on the interaction process measure of N. Flanders (1969). It has been modified in consultation with Planned Variation sponsors in order to pick up the behaviors which they consider important.

Early versions of the CO were somewhat different in their procedures. For example, the 1970 version included a series of summary ratings, and asked for physical layout information in a somewhat different way than did the 1971 and 1972 versions. Changes were made in order to obtain more reliable and more useful information. The CO procedure is quite complicated. Details can be best understood by examining the manual prepared by SRI (n.d.).

At present extensive and comprehensive analysis of this instrument is being done by J. Stallings and her associates

TABLE 1

LIST OF CLASSROOM OBSERVATION VARIABLES

1. Activity A: snack, lunch, any eating activity
2. Activity B: group time: story-reading, singing, TV, record-playing, dancing, usually entire class in one group
3. Activity C: academic activities: numbers, alphabet, reading, language development (with or without curriculum materials)
4. Activity D: inquiry activities: finding out about people and how they live; finding out about the natural world (magnets, shapes, sound)
5. Activity E: table games, guessing games, working puzzles
6. Activity F: arts and crafts and domestic activities: cooking, sewing, pounding or sawing
7. Activity G: blocks, trucks, dolls, dress-up, water play
8. Adults with children in academic activities
9. Academic activities (frequency of occurrence)
10. Independent child activity (child observed as alone in any activity)
11. Wide variety of activities
12. Adult interactions with one or two children
13. Aide's participation in academic activities
14. Adult informing children symbolically (adult teaching with pictures, letters, numerals, etc.)
15. Adult direct questioning of child (questions to which there is a definite expected response either verbal or non-verbal, e.g., "Will you bring the water pitcher here?"; "What do 3 and 1 make?"
16. Child response to adult direct question (verbal or non-verbal; right or wrong)
17. Adult praise and corrective feedback (guide to alternative, reason, control by praising, question--includes any accompanying expressions of emotion)

(Table 1 cont.)

18. Adult feedback to child response (variable 16 followed immediately by variable 17)

19. Adult informing children (teaching, explaining, instructing)

20. Adult asking "thought" questions (questions to which there is no particular expected response, no right or wrong answer)

21. Adult informing child with concrete objects (concrete objects being any tangible, real object such as blocks, Cuisenaire rods. scales, clay, etc.)

22. Adult acknowledgement to child (includes any accompanying emotions)

23. Child self-learning with concrete objects (e.g., child alone working out math problem with scales or Cuisenaire rods; includes play as well as "work")

24. Child self-learning (child teaching or informing himself either with or without "machine" such as language master or typewriter; does not include code for comment, play)

25. Child teaching another child (child informing or explaining to another child)

26. Child self-learning with symbols (child alone "learning" with paper and pencil, numerals, letters, workbooks, etc.)

27. Child asking questions (includes all kinds of questions, requests in the form of questions)

28. Child self-expression (comment, play, show-and-tell)

29. Adult communication focus: one child

30. Adult communication focus: small group

31. Adult communication focus: large group

32. Adult praise/acknowledgement of children (adult complimenting or commenting more or less favorably on child's behavior)

33. Adult "positive" corrective feedback (adult attempting to alter child's (or group's) behavior by guiding to alternative activity, giving a reason why behavior is unacceptable, controlling by praise of other children, or questioning child as to his behavior)

34. Adult "negative" corrective feedback (adult attempting to alter child's (or group's) behavior by firmness, demeaning, threatening or punishing in a sad, negative, or angry manner)

35. Adult "negative" behavior (adult doing anything in a sad, negative, angry, firm, demeaning, threatening or punishing manner)

36. Child "negative" behavior (same as variable 35)

37. Negative behavior (variable 35 + variable 36)

38. Adult positive affect toward children (adult communicating to child in happy manner)

39. Child positive affect toward adults

40. All positive affect (all evidence of "happiness")

41. Child positive affect

(Variables 42 through 51 are derivations of variables 1 through 41)

42. Independent children in academic activities (variable 9 minus variable 8)

43. Teachers and volunteers with children in academic activities (variable 8 minus variable 13)

44. Independent children in non-academic activities (variable 8 minus variable 13)

45. Adult informing children other than symbolically or with concrete objects (variable 19 minus variables 14 and 21)

46. Adult praise of children (variable 32 minus variable 22)

47. Adult corrective feedback (either variable 17 minus variable 46 or variable 33 plus variable 34)

48. Adult negative behavior other than corrective feedback (variable 35 minus variable 34)

49. Child positive affect to other children (variable 41 minus variable 39)

50. Child informing self other than symbolically (variable 24 minus variable 26)

51. Adult positive affect to other adults (variable 40 minus variables 41 and 38)

(in press) at SRI for the Follow Through evaluation.

Reliability

Since the CO was designed specifically for the Planned Variation and Follow Through evaluations, no outside estimates of reliability and validity are available. Observers are trained to 80 percent agreement on all codes. In the Follow Through data (SRI, 1970) percent agreement between paired observers is used as a reliability estimate. The average percent agreement achieved on a set of 63 variables is 76.9% (S.D. = 15.6). On more than one-half of the variables, the agreement is greater than 81%. There is higher agreement on variables that are least often recorded (mean agreement = 86.8%, S.D. = 12.1) than those most frequently recorded (mean agreement = 69.0%, S.D. = 13.4).

Remarks

Even though the reliability and validity of the Classroom Observation Procedure have not been adequately explicated, this procedure has great potential in illuminating the processes that characterize various early grade and preschool classrooms. It can be used as both an independent and dependent variable in analyses. The information obtained by this procedure can be viewed as the "treatment", and thus, as independent variables in the analysis. Changes in the findings can also be interpreted as possible effects of the HSPV models, and thus, as dependent variables. Further refinement of the observation process both as a child outcome/non-cognitive measure and as a classroom process/implementation measure is encouraged.

References

Flanders, N.A. Analyzing classroom interaction. Reading, Mass.: Addison-Wesley, 1969.

Lukas, C.V., & Wohlleb, C. Implementation of Head Start Planned Variation: 1970-71. Prepared for Office of Child Development, Grant #H1926. Cambridge, Mass.: Huron Institute, 1973.

Stallings, J., Wilcox, M., and others. Follow Through Classroom Observation Evaluation. Menlo Park: SRI, in press.

Stanford Research Institute. Training manual for classroom observation, Menlo Park, California: Stanford Research Institute, n.d.

Stanford Research Institute. Evaluation of the National Follow Through Program 1969-71: Appendix B - Follow Through Classroom Observations. Report for Bureau of Elementary and Secondary Education, Department of Health, Education and Welfare, Menlo Park, California: SRI, 1972.

Eight-Block Sort Task

Purpose

The Eight-Block Sort Task is a measure of maternal teaching style and interaction styles between mother and child. A major assumption underlying the use of this task is that the mother is the major socializing agent for the preschool child. Therefore, it is the mother who is primarily responsible for the way in which a child's world is structured and transmitted to him. The mother's interaction with the child mediates how the child interprets and processes his environment around him.

It is hoped that this opportunity for a mother to interact freely with her child in a standardized situation will give the tester and/or observer information about the mother's style of interaction with her child. More specifically, the Eight-Block sort allows one to look at the modes of communication between mother and child, the mother's structuring of the learning situation for the child, the child's responses to mother's teaching demands, and the motivational controls that the mother employs to guide the behavior and performance of the child.

Description

The Eight-Block Sort Task uses blocks differing according to four attributes--height(tall or short), mark (X or 0 on

top), color (red, yellow, green or blue), and shape (rectangular or circular cross section). The task involves sorting eight of the blocks according to two of the four attributes--height and mark. The four groups remaining after a correct sort are (1) tall blocks marked X, (2) short blocks marked X, (3) tall blocks marked 0, and (4) short blocks marked 0. After the mother is taught the eight-block sort task by the tester (called the trainer), she is told to teach the task to the child in any manner she wishes. The mother is instructed to teach the child not only how to sort the blocks correctly into groups of the same height and same mark, but also how to correctly explain reasons for the groupings.

The task situation is divided into three phases: (1) training of the mother by a trained tester; (2) training of the child by the mother; and (3) testing of the child on task comprehension by the tester. During the first phase--the mother's training session--the trainer teaches the mother to sort the blocks correctly in a standardized procedure. The success of the mother in learning the task is recorded by either the tester or a trained observer. During the second phase--the child's training session--the mother is free to instruct the child in any way she wants. The training session lasts as long as the mother thinks it is necessary for the child to learn the task or a maximum time of twenty minutes, whichever comes first. The tester or a trained

observer records behaviors of the mother and the child during this training session. During the third phase--the child's testing session--the tester asks the child to place two previously unseen blocks (a tall X and a short 0) on the board with the original eight blocks which are sorted into the four groups defined by height and mark. After the child places a block, the tester asks the child to give reasons for his placement. The mother is instructed to remain neutral during the testing. The child's responses and the mother's behavior during the testing phase are recorded by the tester or a trained observer. During the first year of the HSPV study, the observation of the mother's and child's behaviors during the three phases were made by the tester; during the second and third year of study, the observations were made by a trained observer.

Scoring and Analysis

For analysis of the Eight-Block Sort data obtained in the first year of the HSPV Study the following five summary variables were defined and studied by SRI (1971): (1) verbal communication--an indication of mother's verbal task-related communication to the child; (2) task description--an indication of how the mother orients the child toward the task dimensions and how the mother trains the child in discrimination of these dimensions; (3) regulation--an indication of whether the mother uses positive or negative, verbal or non-verbal means of regulating the child's behavior; (4) child's verbal

responsiveness--an indication of the amount of verbal labeling elicited from or volunteered by the child; and (5) child success--an indication of the child's success in placing the blocks and in giving reasons for the placements. For all of the summary variables, a higher score indicates a more positive behavior on the part of the mother or the child in the task situation. In the SRI analysis, scores were standardized to a mean of 50 and a standard deviation of 10. Because of the difference in formats of the fall and spring rating forms (observational forms),only the definitions of variable 5 are exactly the same. Close analysis of the definitions of the other four variables for the fall and the spring shows they are not comparable (Walker, 1972).

The analysis of the Eight-Block Sort data for the second and third years of the HSPV Study includes an analysis of the child's success in the post-task session. Each child's total success score ranges from 0 to 8 with a maximum of four points for correct placement of the two blocks and a maximum of four points for correct verbalizations about the placements. The total success score is broken into the following parts:

Criterion	Score
1. Placement of short 0 test block in correct group	0-2
2. Verbalization of "short" or same height in explaining placement	0-1

Criterion	Score
3. Verbalization of same mark, 0, or other descriptive label used by mother when teaching (e.g., "cheerios") in explaining placement	0-1
4. Placement of tall X test block in correct group	0-2
5. Verbalization of "tall" or same height in explaining placement	0-1
6. Verbalization of same mark, X, or other descriptive label used by mother when teaching (e.g., "airplane") in explaining placement	0-1

Points for verbalization are given only if the child has placed the block according to the relevant dimension (height or mark). Variables derived from the 1971-72 forms to describe maternal teaching style and mother-child interaction are explicated in great detail in Appendix B.

Development of Task

The Eight-Block Sort Task was developed and first used by R. Hess and V. Shipman. Using this task they have shown that there are differences in mother-child interaction patterns which are associated with different social class and background indices (Hess & Shipman, 1965; Hess et al., 1968). For example, in a study with 163 black mother-child pairs, the middle class mothers used more elaborated codes and more person-oriented explanations than did the lower class mothers who used predominantly restricted codes and status-oriented explanations. Furthermore, in a follow-up study (Hess et. al.,

1969) differences in maternal teaching styles, as illuminated by this task, were not only associated with social class indices, but also with a child's academic performance in the first two years of school. Variables that differentiated the more effective maternal teacher were greater orientation to the task, reinforcement of more correct responses than errors, use of specific language, greater reliance on verbal feedback from the child, and more use of praise than blame in controlling the child's behaviors.

ETS Longitudinal Study

The Eight-Block Sort Task is also being used in the ETS Head Start Longitudinal Study (Shipman et al., 1971). Each child's total performance on the post-task test was scored 0 to 8 points with a maximum of four points for correct verbalizations and a maximum of four points for correct placement. Correlations of the placement and verbalization subscores were .81 and .86 with the total scores; their correlation with each other was .32. The placement score was inflated since all children were credited with a higher score in questionable protocols where the examiner had not probed enough.

Results from the ETS study show that the majority of the children (n = 1495) could place the blocks correctly (72.2% - short "0"; 64.3% - tall "X") while only a small minority could verbalize the reasons for correct placements (approximately 20% verbalized one dimension correctly and

11% verbalized both dimensions). The mean placement score for the entire sample was 3.18 (S.D. = 1.09) while the mean verbalization score was .86 (S.D. = 1.29). Percentile distributions of total score by age and sex for the total ETS sample were developed. In general, even though differences in total scores between sexes were negligible, the girls obtained higher verbal scores than the boys.

In a future ETS report, the Eight-Block Sort data will be analyzed in relation to several maternal variables such as teaching style, use of feedback, orienting, and use of control strategies.

Home Start Study

The Eight-Block Sort Task with a similar observation form to that used in the last year of the HSPV Study was used in the pilot year of the Home Start Study (Hi-Scope, 1973). Child success scores and two maternal variables were coded during the child's training and testing. All other maternal and child variables were recorded from audio tapes of the session.

Over half of the Home Start children (ages 3 to 5½) in this pilot study placed each of the blocks correctly: short O -62% (n = 164), tall X - 51% (n = 160). Only a small percentage could verbalize the total correct reason for placement (short O - 16%, n = 74; tall X - 20%, n = 75); about one-half could verbalize one dimension correctly

(short 0 - 60%, n = 74; tall X - 50%, n = 75). Correlations of the placement and verbalization subscores were .79 and .84 with the total scores; their correlation with each other was .34. The authors point out that a number of tester errors in administering the test might explain some of the performance levels. Even though the verbal success scores were much higher in the Home Start pilot study, the intercorrelation of scores and the placement scores were very similar to those found in the ETS Longitudinal Study (Shipman, et al., 1971).

Head Start Planned Variation Study

Distributions of the total success score and two subscores (placement and reason/verbalization) for the Fall 1971 HSPV sample can be found in Tables 1 - 3. The mean placement score for the entire sample (N = 1,203) was 3.086 (S.D. = 1.211), while the mean verbalization score was 1.379 (S.D. = 1.515). Both subscores and the total score increased with age. From the distribution table for placement scores, it can be seen that there is a ceiling effect for older preschool children (greater than 66 months). A ceiling effect for this subscore was also found in the small sample (n = 20) in the inter-judge reliability study in Fall 1971. (See Appendix B for details.)

Correlations of the three success scores with the other tests in the Fall 1970 battery and the Fall 1971 battery are shown in Tables 4 and 5. In both years there were no high correlations with other tests in the battery. In Fall 1970 the highest correlation was between the total

TABLE 1

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

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	3.333	3.399
39-41	1	4.000	-----
42-44	11	2.545	2.311
45-47	33	3.394	2.117
48-50	116	3.793	2.272
51-53	201	3.527	2.307
54-56	195	3.897	2.212
57-59	184	4.723	2.052
60-62	137	4.898	2.245
63-65	98	4.612	2.481
66-68	111	5.829	2.238
69-71	72	5.944	1.978
72-74	41	5.902	2.105
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	1203	4.466	2.380

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

[2]Maximum score = 8.

TABLE 2

DISTRIBUTION OF EIGHT-BLOCK CORRECT PLACEMENT SCORES FOR ALL CHILDREN IN THE FALL 1971 SAMPLE[1]

Age (Months)	N	Mean Score [2]	S.D.
36-38	3	2.000	1.633
39-41	1	3.000	-----
42-44	11	1.909	1.311
45-47	33	2.515	1.209
48-50	116	2.784	1.285
51-53	201	2.657	1.413
54-56	195	2.903	1.230
57-59	184	3.342	0.998
60-62	137	3.336	1.027
63-65	98	3.061	1.260
66-68	111	3.541	0.888
69-71	72	3.681	0.723
72-74	41	3.585	0.826
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	1203	3.086	1.211

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

[2] Maximum score = 4.

TABLE 3

DISTRIBUTION OF EIGHT-BLOCK CORRECT REASON SCORES FOR ALL CHILDREN IN THE FALL 1971 SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	1.333	1.886
39-41	1	1.000	-----
42-44	11	0.636	1.226
45-47	33	0.879	1.225
48-50	116	1.009	1.329
51-53	201	0.871	1.267
54-56	195	0.995	1.330
57-59	184	1.380	1.443
60-62	137	1.562	1.556
63-65	98	1.551	1.533
66-68	111	2.288	1.602
69-71	72	2.264	1.590
72-74	41	2.317	1.537
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	1203	1.379	1.515

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

[2]Maximum score = 4.

TABLE 4

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	.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).

TABLE 5 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)	.506 (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	.252 (1075)	.356 (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)	.295 (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)	.032 (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)	.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)	.266 (1148)	.372 (1148)	.390 (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.

success score and the 64-item Preschool Inventory score (r = .356, n = 556). In Fall 1971, the highest correlation was between the verbalization subscore and the Peabody Picture Vocabulary score (r = .445, n = 1,119).

In Fall 1970 (n = 576), correlations of the placement and the verbalization subscores were .74 and .93 with the total score; their correlation with each other was .43. In Fall 1971 (n = 1,211), correlations of the placement and verbalization subscores were .84 and .90 with the total score; their correlation with each other was .52

Similar correlations among the three scores were also found in the reliability study in Fall, 1971 (n = 20). Correlations of the placement and verbalization subscores were .73 and .94 with the total score; their correlation with each other was .44. The inter-correlations of the variables used in the reliability study (using two different units of analysis) for each of two paraprofessional observers can be found in Appendix B.

Reliability

From the ETS Longitudinal data, the estimated reliability (coefficient alpha) was .55 for total placement and .86 for total verbalization (Shipman et al., 1971).

In the Home Start pilot study (Hi-Scope, 1973), 40 categories were coded from 10 separate tapes, making 400 inter-judge comparisons. In 83 of these cases (21%) the frequency counts coded differed by five or more. The authors concluded some of the categories used needed to be refined before

further observations were made.

In the fall of 1971 Huron Institute and SRI conducted an inter-Observer reliability study using the Eight-Block Sort observer form used in Fall 1971 and Spring 1972. Definitions and basic statistics for 35 maternal teaching style and mother-child interaction variables, three time variables and seven success variables are located in Appendix B. In this study two paraprofessional observers simultaneously watched 20 children and three observers (two paraprofessionals and one expert) simultaneously observed eight children. There was perfect agreement among both sets of observers on the total score, the placement score and the verbalization score during the post-task test. The correlation coefficients between two observers for recording time in minutes were adequate: orientation time = .78, training time = .94 and total time = .94.

Since there was a lack of time limit (under a ceiling of 20 minutes) for the orientation and training periods, it is possible that two mothers could have the same absolute counts of a particular behavior in quite different time periods. From a child development viewpoint it can be argued that the percentage of total time a mother or child spends on a particular behavior is a more accurate picture of what is happening than a straight frequency count. Since it was not certain which unit of analysis is best for analyzing the data, both units were used in the reliability study for analyzing maternal style and mother-child interaction variables.

Reliability estimates between observers for the maternal style and mother-child interaction variables were calculated in two ways: correlation coefficients (Pearson product-moment) and r's calculated from a one-way analysis of variance repeated measures design. In general, the reliability estimates for two observers, using the frequency counts as a unit of analysis (n = 20), were quite adequate. Sixty percent of 30 product-moment correlations were greater than .90. The reliability coefficients estimated from the one-way ANOVA design were very similar for most of the 30 variables. When frequency per minute was used as the unit of analysis, the results for both types of reliability estimate were much lower than those using frequency counts as the unit of analysis and there was a wider discrepancy between the two estimates. For instance, only one out of 11 (9%) variables had a product-moment correlation greater than .90 when frequency per minute was used.

Finally, very few F tests for observer effects were significant, using either unit of analysis, pointing out that there were generally no significant portions of variance attributable to differences between observers' ratings.

Remarks

Potentially there is a wealth of information that can be learned from the Eight-Block Sort Task, as it allows one to watch a mother teach her own child a particular task. It is questionable, however, how much can be obtained from the HSPV data, since four of the five observation forms

used during the three years of study are different. The only variable that is recorded exactly the same on all the forms is the child's success on the post-task after the training. Even though the variables using frequency counts as a unit of analysis appear to be more reliable, more studies need to be done to clarify the merits of the two units: frequency counts and frequency per minute.

Some of the interesting questions which could be answered using the Eight-Block Sort Task procedure are listed below:

1. How does the mother's orientation behavior relate to the child's success score?
2. How does the mother's behavior during the training session relate to the final score? In other words, what are the components of the mother's teaching style?
3. How does the child's verbal and non-verbal behavior during the training session relate to the child's success?
4. What kind of a control system does the mother use in handling and teaching her child?
5. How does the mother's performance in her training session relate to her behavior in the orientation and training session and to the child's success?
6. How does the mother's style of teaching compare with the teacher's style on the same task?
7. What is the quality and content of the mother's and of the child's verbalizations?

8. How does the mother's teaching style relate to background variables?
9. How does the child's success relate to his performance on other outcome measures?

Because of the limitations of the observation forms used in the HSPV Study several of these questions cannot be answered at all and many cannot be answered in much detail. In future evaluations using the Eight-Block Sort it is recommended that the observation form be designed to answer the specific questions the researcher is trying to answer. Further refinement of this observation procedure is encouraged so that maximum use of the data can be made in future large scale evaluation efforts.

References

Hess, R. D., & Shipman, V. C. Early experience and the socialization of cognitive modes in children. Child Development, 1965, 36, 869-888.

Hess, R. D., Shipman, V. C., Brophy, J. E., & Bear, R. M. The Cognitive environments of urban preschool children. Chicago: University of Chicago Press, 1968.

Hess, R. D., Shipman, V. C., Brophy, J. E., & Bear, R. M. The cognitive environments of urban preschool children: follow-up phase. Chicago: University of Chicago Press, 1969.

Hi-Scope Foundation/Abt Associates. Home Start Evaluation Study. Draft Interim Report to Office of Child Development Department of Health, Education and Welfare - #HEW-OS-72-127. Ypsilanti, Michigan: Hi-Scope/Abt, 1973.

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. J.: Educational Testing Service, 1971.

Stanford Research Institute. Implementation of planned variation in Head Start: Preliminary evaluations of planned variation in Head Start according to Follow Through approaches (1969-1970). Washington, D. C.: Office of Child Development, Department of Health, Education and Welfare, 1971.

Walker, D. K. Memorandum: Critique of the 1969-70 Head Start Planned Variation Eight-Block Sort Analysis. Cambridge, Mass.: Huron Institute, 1972.

Ethnic Identity Questionnaire
Children's Cultural Awareness Scale

Purpose

The Ethnic Identity Questionnaire (EIQ) was developed to investigate the ethnic identity of Mexican-American children and the Children's Cultural Awareness Scale (CCAS) was developed to explore the cultural awareness of Black children in the Head Start Planned Variation Study. They respond to the fact that strengthening cultural awareness and pride are objectives of some sponsors as well as a general concern of Head Start. The measures were developed and used on an experimental basis as part of the Planned Variation effort to expand measurement to significant areas not typically assessed in evaluations.

Description

The EIQ (available in both English and Spanish versions) was developed by Manuel Ramirez III at the University of California, Riverside. The instrument consists of seven questions which ask about things that are indigenous to the Mexican-American culture. The CCAS was developed by Edward J. Barnes at the University of Pittsburgh. The scale consists of thirteen items which ask about specific aspects of the Black culture in America. A tape recorder and pictures are used in connection with several of the items.

Remarks

Both of these questionnaires were pretested on small samples by their developers and used in the second year of the Head Start Planned Variation Study, upon the approval of the local Head Start Director and the Policy Council. Because these tests were in the development stage, they are being analyzed on a pilot basis rather than as part of this study. The findings of these experimental studies will be reported elsewhere.

ETS Enumeration Test

Purpose

The ETS Enumeration Test is designed to measure components of the cognitive process required in learning the concept of number. The test consists of four sections, each of which explores a different aspect of the enumeration process. The test is an attempt to investigate several aspects of the developmental process involved in the conception of number as hypothesized by Piaget (1952). One subtest--touching items--is an extension and revision of the Potter and Levy method of studying how one attends systematically to items in an array (Potter & Levy, 1968).

Description

The ETS Enumeration Test used in the HSPV Study can be divided into the following four subtests:

1. Counting. Three items require the child to count the dots (or colored circles) on a page and then say how many there are. The arrangement and number of dots in the items are six random, nine in a row, and nine random.

2. Touching. Six items require the child to touch each of the dots on a page "once and only once". The dots, all of the same color and varying in number from six to nine, are arranged in three types of arrays: single line, rows, random.

3. Same Number Matching. Eight items require the child to find the picture out of three presented that has the same number of objects (although not necessarily in the same array) as the stimulus picture. The stimulus pictures include 3 birds, 4 pennies, 3 cylinders, 5 walnuts, 5 fish, 7 apples, 9 balloons, and 7 lollipops.

4. Same Order Matching. Six items require the child to find the picture out of three presented that has the same ordering of objects as the stimulus picture. The stimulus pictures include 3 flowers, clothes on a clothesline, 3 fish going through a tunnel, a 4-car train, 6 beads on a string and 2 turtles.

Each item of the counting subtest is scored in two parts: 0-1 points for counting and 0-1 points for telling how many circles there are. The "counting" part is scored either "1" (correct) or "0" (wrong). A "1" (correct) is given for the "telling how many circles" part if (1) the child gives the correct number of circles on the page, whether or not he counted correctly to that number, or (2) the child gives the number of circles that corresponds to the number he counted to, even though it was technically incorrect; all other responses are scored "0" (wrong). The items in the touching and matching subtests are scored either "1" (correct) or "0" (wrong). The subtest scores as well as a total score, consisting of the sum of scores for all four subtests (range 0-26) and a partial score, consisting of the sum of scores for the first three subtests (range 0-20), were used in the analyses of the

the ETS Enumeration Test. Coding reliabilities for this test are very high (see Appendix D).

Development of Instrument

For use in the first year of their longitudinal study, ETS developed an enumeration test, called ETS Enumeration I, based on Potter and Levy's method (ETS, 1970; Potter & Levy, 1968; Shipman et al., 1971). ETS Enumeration I had 12 touching items, involving two-colored or three-colored circles arranged randomly or one, two, or three rows, and one counting aloud item. Both Shipman and Potter have found that (1) success on touching items was correlated with age, (2) arrays containing the smallest number of dots were easiest, and (3) random arrangements of dots were most difficult. From the results of the ETS Enumeration I, ETS developed a longer and more comprehensive test, called ETS Enumeration II, for use in the second year of their longitudinal study. Enumeration II which is similar to the HSPV version, has four counting items, four touching items, eight same number matching items and five same order matching items. The HSPV version of the ETS Enumeration test was also used in the pilot testing phase of the Home Start Study (Hi-Scope, 1973).

Norms

Norms for the ETS Enumeration total scores (sum of scores of all four subtests) and for the ETS Enumeration partial

scores (sum of scores from the Counting, Touching and Same Number Matching subtests) for the Fall 1971 HSPV sample are available in Tables 1 - 16. Based on 16 three month age intervals from 36-38 months to 78-80 months, these tables give the number of children, mean score, and the standard deviation at each age level for the following subgroups of the HSPV sample: (Note: the first table listed is for the total score and the second table listed is for the partial score) total sample (Tables 1, 9), males (Tables 2, 10), females (Tables 3, 11), children with previous preschool experience (Tables 4, 12), children with no previous preschool experience (Tables 5, 13), white children (Tables 6, 14), black children (Tables 7, 15), and Mexican-American children (Tables 8, 16). The mean of the total sample was 11.647 (S.D. - 4.842) for the total score and 9.140 (S.D. 4.155) for the partial score. The mean scores for females (total score = 12.018, S.D. = 4.960; partial score = 9.524, S.D. = 4.206) were higher than the mean scores for males (total score = 11.301, S.D. = 4.702; partial score = 8.783, S.D. = 4.076). The mean scores for children with previous preschool experience (total score = 12.695; S.D. = 4.641; partial score = 10.136, S.D. = 4.020) were higher than the mean scores for children without previous preschool experience (total score = 11.339, S.D. = 4.865; partial score = 8.840, S.D. = 4.161). Scores in all tables increased with age.

Norms for the four individual subtest scores for the Fall 1971 HSPV sample are presented in Tables 17 - 20. The

mean scores for the total sample were 3.996 (S.D. = 2.061) for the Counting subtest, 4.442 (S.D. = 1.668) for the Touching subtest, 4.159 (S.D. = 1.794) for the Same Number Matching subtest, and 3.174 (S.D. = 1.113) for the Same Order Matching subtest. Scores on the four subtests increased as a function of age.

Norms (means, standard deviations, percentiles) for the ETS Enumeration I scores used in Year 1 of the ETS Longitudinal Study are available for children, aged 42-59 months; norms for the ETS Enumeration II scores used in Year 2 of the study are available for ~~children, aged~~ 51-69 months (Shipman, 1972). The mean score for the ETS Enumeration I (range 0-12) was 5.9 (S.D. = 3.52; N = 1395). The mean score for the ETS Enumeration II (range 0-25) was 12.8 (S.D. = 5.06, N = 1193). Sex differences favoring girls for both sets of scores were significant at the .001 level. When subjects were divided into two age groups at the median age, a significant age difference favoring the "older" chilren was found in both years of the ETS Study.

TABLE 1

DISTRIBUTION OF ETS ENUMERATION TOTAL SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	2.333	1.247
39-41	1	3.000	-----
42-44	8	8.375	3.199
45-47	28	8.893	3.811
48-50	96	9.083	4.094
51-53	173	9.561	4.276
54-56	188	10.394	4.286
57-59	175	10.634	3.873
60-62	135	13.067	4.253
63-65	93	13.656	4.580
66-68	113	14.469	4.173
69-71	74	15.473	5.131
72-74	41	15.756	4.658
75-77	1	22.000	-----
78-80	---	-----	-----
TOTAL	1129	11.647	4.842

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

[2]Maximum score = 26; total score is the sum of scores of the four subtests.

TABLE 2

DISTRIBUTION OF ETS ENUMERATION TOTAL SCORES FOR MALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	1	3.000	-----
42-44	3	9.333	1.247
45-47	15	8.800	4.370
48-50	52	9.038	4.201
51-53	98	9.010	3.797
54-56	100	10.200	4.301
57-59	86	10.291	3.644
60-62	58	12.621	4.298
63-65	47	13.957	4.467
66-68	61	13.836	4.046
69-71	37	14.946	4.838
72-74	26	14.731	4.579
75-77	1	22.000	-----
78-80	---	-----	-----
TOTAL	585	11.301	4.702

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

[2] Maximum score = 26; total score is the sum of scores of the four subtests.

TABLE 3

DISTRIBUTION OF ETS ENUMERATION TOTAL SCORES FOR FEMALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	2.333	1.247
39-41	---	-----	----
42-44	5	7.800	3.816
45-47	13	9.000	3.038
48-50	44	9.136	3.963
51-53	75	10.280	4.734
54-56	88	10.614	4.257
57-59	89	10.966	4.054
60-62	77	13.403	4.188
63-65	46	13.348	4.673
66-68	52	15.212	4.199
69-71	37	16.000	4.357
72-74	15	17.533	4.241
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	544	12.018	4.960

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

[2]Maximum score = 26; total score is the sum of scores of the four subtests.

TABLE 4

DISTRIBUTION OF ETS ENUMERATION TOTAL 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	---	-----	-----
45-47	7	9.286	4.300
48-50	16	11.188	3.468
51-53	35	10.257	4.265
54-56	21	10.571	3.762
57-59	25	10.400	4.534
60-62	47	13.021	3.895
63-65	39	13.744	4.265
66-68	38	13.974	4.049
69-71	29	15.655	4.971
72-74	14	15.214	4.109
75-77	1	22.000	-----
78-80	---	-----	-----
TOTAL	272	12.695	4.641

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

[2]Maximum score = 26; total score is the sum of scores of the four subtests.

TABLE 5

DISTRIBUTION OF ETS ENUMERATION TOTAL 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	3	2.333	1.247
39-41	1	3.000	-----
42-44	8	8.375	3.199
45-47	21	8.762	3.624
48-50	77	8.688	4.097
51-53	132	9.500	4.206
54-56	162	10.358	4.343
57-59	145	10.586	3.748
60-62	86	13.081	4.478
63-65	54	13.593	4.794
66-68	73	14.726	4.269
69-71	43	15.581	5.231
72-74	27	16.037	4.895
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	832	11.339	4.865

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

[2]Maximum score = 26; total score is the sum of scores of the four subtests.

TABLE 6

DISTRIBUTION OF ETS ENUMERATION TOTAL SCORES FOR WHITE CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	1	3.000	-----
42-44	2	11.500	3.500
45-47	13	9.231	2.860
48-50	40	8.925	3.856
51-53	72	9.819	4.426
54-56	84	10.810	4.489
57-59	70	10.643	4.053
60-62	64	13.688	4.304
63-65	32	15.156	4.162
66-68	44	15.636	3.637
69-71	33	17.455	4.466
72-74	26	17.154	4.312
75-77	1	22.000	-----
78-80	---	-----	-----
TOTAL	482	12.357	5.069

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

[2]Maximum score = 26; total score is the sum of scores of the four subtests.

TABLE 7

DISTRIBUTION OF ETS ENUMERATION TOTAL SCORES FOR BLACK CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	2.333	1.247
39-41	---	-----	-----
42-44	6	7.333	2.285
45-47	15	8.600	4.454
48-50	42	8.500	4.371
51-53	70	9.514	4.146
54-56	78	9.654	4.218
57-59	76	10.382	3.930
60-62	46	12.152	3.741
63-65	45	11.689	4.060
66-68	42	12.833	3.970
69-71	28	12.786	5.122
72-74	14	13.786	4.003
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	465	10.581	4.455

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

[2] Maximum score = 26; total score is the sum of scores of the four subtests.

TABLE 8

DISTRIBUTION OF ETS ENUMERATION TOTAL 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	11	11.455	2.675
51-53	30	9.133	4.209
54-56	24	11.458	2.784
57-59	27	11.444	3.131
60-62	20	13.000	4.062
63-65	15	15.733	4.139
66-68	25	14.720	4.341
69-71	13	16.231	4.079
72-74	---	-----	-----
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	165	12.479	4.410

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

[2]Maximum score = 26; total score is the sum of scores of the four subtests.

TABLE 9

DISTRIBUTION OF ETS ENUMERATION PARTIAL SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	2.000	1.414
39-41	1	2.000	-----
42-44	8	5.750	3.192
45-47	28	6.607	3.384
48-50	96	7.063	3.335
51-53	173	7.243	3.600
54-56	188	8.181	3.579
57-59	175	8.286	3.427
60-62	135	10.200	3.858
63-65	93	10.882	3.951
66-68	113	11.593	3.696
69-71	74	12.554	4.182
72-74	41	12.585	4.060
75-77	1	17.000	-----
78-80	---	------	-----
TOTAL	1129	9.140	4.155

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

[2] Maximum score = 20; partial score is the sum of scores of the counting, touching and same number subtests.

TABLE 10

DISTRIBUTION OF ETS ENUMERATION PARTIAL SCORES FOR MALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	------	-----
39-41	1	2.000	-----
42-44	3	7.000	1.414
45-47	15	6.333	3.806
48-50	52	7.135	3.397
51-53	98	6.694	3.246
54-56	100	7.960	3.597
57-59	86	7.791	3.464
60-62	58	9.828	3.940
63-65	47	11.170	3.954
66-68	61	10.918	3.554
69-71	37	12.000	3.952
72-74	26	11.731	3.918
75-77	1	17.000	-----
78-80	---	------	-----
TOTAL	585	8.783	4.076

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

[2]Maximum score = 20; partial score is the sum of scores of the counting, touching and same number subtests.

TABLE 11

DISTRIBUTION OF ETS ENUMERATION PARTIAL SCORES FOR FEMALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score [2]	S.D.
36-38	3	2.000	1.414
39-41	---	-------	-----
42-44	5	5.000	3.688
45-47	13	6.923	2.786
48-50	44	6.977	3.258
51-53	75	7.960	3.900
54-56	88	8.432	3.541
57-59	89	8.764	3.322
60-62	77	10.481	3.771
63-65	46	10.587	3.926
66-68	52	12.385	3.701
69-71	37	13.108	4.330
72-74	15	14.067	3.872
75-77	---	-------	-----
78-80	---	-------	-----
TOTAL	544	9.524	4.206

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

[2] Maximum score = 20; partial score is the sum of scores of the counting, touching and same number subtests.

TABLE 12

DISTRIBUTION OF ETS ENUMERATION PARTIAL 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	---	-------	------
45-47	7	6.857	3.132
48-50	16	8.750	3.307
51-53	35	7.714	3.021
54-56	21	7.905	2.991
57-59	25	8.440	3.869
60-62	47	10.064	3.502
63-65	39	11.205	3.495
66-68	38	11.632	3.414
69-71	29	12.793	4.012
72-74	14	13.000	3.485
75-77	1	17.000	------
78-80	---	-------	------
TOTAL	272	10.136	4.020

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

[2]Maximum score = 20; partial score is the sum of scores of the counting, touching and same number subtests.

TABLE 13

DISTRIBUTION OF ETS ENUMERATION PARTIAL SCORES FOR ALL CHILDREN WITH NO PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 NSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	2.000	1.414
39-41	1	2.000	------
42-44	8	5.750	3.192
45-47	21	6.524	3.445
48-50	77	6.753	3.244
51-53	132	7.197	3.526
54-56	162	8.204	3.650
57-59	145	8.166	3.343
60-62	86	10.279	4.074
63-65	54	10.648	4.235
66-68	73	11.603	3.877
69-71	43	12.581	4.293
72-74	27	12.370	4.313
75-77	---	------	------
78-80	---	------	------
TOTAL	832	8.840	4.161

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

[2]Maximum score = 20; partial score is the sum of scores of the counting, touching and same number subtests.

TABLE 14

DISTRIBUTION OF ETS ENUMERATION PARTIAL SCORES FOR WHITE CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	------	-----
39-41	1	2.000	-----
42-44	2	8.500	3.500
45-47	13	6.846	3.348
48-50	40	6.950	3.154
51-53	72	7.556	3.738
54-56	84	8.667	3.755
57-59	70	8.500	3.528
60-62	64	10.656	3.938
63-65	32	11.656	3.989
66-68	44	12.500	3.401
69-71	33	14.061	3.446
72-74	26	13.538	3.734
75-77	1	17.000	-----
78-80	---	------	-----
TOTAL	482	9.732	4.315

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

[2] Maximum score = 20; partial score is the sum of scores of the counting, touching and same number subtests.

TABLE 15

DISTRIBUTION OF ETS ENUMERATION PARTIAL SCORES FOR BLACK CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	2.000	1.414
39-41	----	------	-----
42-44	6	4.833	2.478
45-47	15	6.400	3.402
48-50	42	6.643	3.624
51-53	70	7.143	3.502
54-56	78	7.462	3.511
57-59	76	8.026	3.572
60-62	46	9.457	3.437
63-65	45	9.667	3.483
66-68	42	10.619	3.664
69-71	28	10.536	4.412
72-74	14	11.429	3.755
75-77	---	------	-----
78-80	---	------	-----
TOTAL	465	8.329	3.912

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

[2] Maximum score = 20; partial score is the sum of scores of the counting, touching and same number subtests.

TABLE 16

DISTRIBUTION OF ETS ENUMERATION PARTIAL 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	11	8.818	2.249
51-53	30	6.733	3.473
54-56	24	8.875	2.315
57-59	27	8.556	2.726
60-62	20	10.300	3.579
63-65	15	12.333	3.771
66-68	25	11.320	3.739
69-71	13	13.077	3.562
72-74	---	------	-----
75-77	---	------	-----
78-80	---	------	-----
TOTAL	165	9.618	3.797

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

[2]Maximum score = 20; partial score is the sum of scores of the counting, touching and same number subtests.

TABLE 17

DISTRIBUTION OF ETS ENUMERATION - COUNTING SUBTEST SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	2.000	2.828
39-41	4	2.250	0.829
42-44	17	2.000	1.815
45-47	57	2.105	1.870
48-50	217	2.613	2.133
51-53	394	3.152	2.043
54-56	404	3.601	2.121
57-59	401	3.703	2.032
60-62	337	4.638	1.677
63-65	245	5.029	1.407
66-68	232	5.138	1.358
69-71	197	5.193	1.419
72-74	80	5.175	1.421
75-77	5	3.800	2.135
78-80	2	4.000	2.000
TOTAL	2595	3.996	2.061

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

[2]Maximum score = 6.

TABLE 18

DISTRIBUTION OF ETS ENUMERATION - TOUCHING SUBTEST SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	2.333	1.247
39-41	4	3.000	2.121
42-44	17	3.529	1.819
45-47	57	3.368	1.650
48-50	217	3.922	1.796
51-53	394	4.076	1.677
54-56	404	4.193	1.754
57-59	401	4.299	1.696
60-62	337	4.697	1.526
63-65	245	4.890	1.406
66-68	232	4.953	1.415
69-71	197	5.147	1.338
72-74	80	5.225	1.224
75-77	5	2.800	1.939
78-80	2	3.000	3.000
TOTAL	2595	4.442	1.668

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

[2] Maximum score = 6.

TABLE 19

DISTRIBUTION OF ETS ENUMERATION - SAME NUMBER MATCHING SUBTEST SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	1.667	0.471
39-41	4	4.250	0.433
42-44	17	3.824	1.689
45-47	57	3.333	1.790
48-50	217	3.378	1.667
51-53	394	3.525	1.683
54-56	404	3.676	1.676
57-59	401	4.032	1.623
60-62	337	4.430	1.690
63-65	245	4.845	1.619
66-68	232	4.901	1.677
69-71	197	5.066	1.808
72-74	80	5.675	1.649
75-77	5	3.800	0.980
78-80	2	2.000	2.000
TOTAL	2595	4.159	1.794

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

[2]Maximum score = 8.

TABLE 20

DISTRIBUTION OF ETS ENUMERATION - SAME ORDER MATCHING SUBTEST SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	3.333	1.247
39-41	4	3.000	0.707
42-44	17	2.824	1.248
45-47	57	2.842	1.136
48-50	217	2.917	1.083
51-53	394	3.036	1.073
54-56	404	3.022	1.104
57-59	401	3.075	1.003
60-62	337	3.291	1.097
63-65	245	3.343	1.131
66-68	232	3.496	1.126
69-71	197	3.437	1.141
72-74	80	3.563	1.243
75-77	5	2.800	1.166
78-80	2	3.500	0.500
TOTAL	2595	3.174	1.113

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

[2] Maximum score = 6.

Score and Item Characteristics

The scoring system for the three Counting subtest "tell me how many there are" items was more lenient in the HSPV Study than in the ETS Longiuudinal Study and in the Home Start Study. In the HSPV Study one point was given for these items (2A, 3A, 4A) if the child obtained a code 1 (same number as the correct sequence ending number), code 2 (single number correct but not same as what number counted to), or code 3 (single number incorrect but identical to last number of sequence counted). A child was given credit for code 3 since it recognized that a child, even though he couldn't count correctly, could have the concept that a cardinal number given to a group must be the same as the last ordinal number in a series used to count the objects in the group. The number of children receiving credit for each code for each of the three items in the HSPV Study in Fall 1971 is listed below:

	Item 2A	Item 3A	Item 4A
Code 1	308	255	202
Code 2	32	18	9
Code 3	136	164	231
Total n	476	437	442

A large proportion of children responding to each item had code 3 responses. Thus, the scores from the HSPV Study are probably higher than corresponding scores from the ETS

Study and the Home Start Study, since these studies only gave credit for code 1 and code 2 responses.

A principal components analyses followed by a varimax rotation of the Fall 1971 ETS Enumeration items resulted in six factors, corresponding to six eigenvalues greater than one: 4.747, 2.478, 1.833, 1.688, 1.382, 1.005. Together these eigenvalues accounted for 51% of the total variance; the first and second values accounted for 18% and 10% of the total variance. The six rotated factors with the 26 item loadings on them are displayed in Table 21. Factor I and Factor VI clearly represented the two parts of the counting subtest; Factor I had the "tell me how many there are" items load highest on it, while Factor VI had the "counting" items load highest on it. Factor III replicated the Touching subtest items. The Same Number Matching and the Same Order Matching subtest items loaded on three factors (II, IV, and V). The only common theme associated with the items that loaded on these three factors was the position of the correct response on the page. Factor II included those items whose correct response was "b" or selecting the picture directly under the stimulus picture on a page. Factor IV included all the items with the "a" response and Factor V included all the items with the "c" response. One slight exception to this pattern was that item 27-c, an "a" response item, loaded highest (.57) on Factor II, the "b" response factor, and second highest (.41) on Factor V, the "a" response factor. These results are very similar to the factor analysis done on the same ETS Enumeration

TABLE 21

FACTOR ANALYSIS OF ETS ENUMERATION TEST
(GIVEN FALL 1971 TO TOTAL HSPV SAMPLE)

Items	Rotated Factors (% total variance) I (18%)	II (10%)	III (7%)	IV (6%)	V (6%)	VI (4%)
Counting:						
2-A count	.20	.07	.20	.10	.01	.71
tell	.79	.11	.15	.08	.05	.19
3-A count	.27	.05	.20	.08	.03	.67
tell	.80	.07	.10	.09	.01	.26
4-A count	.31	.00	.17	.06	-.10	.60
tell	.80	.09	.16	.03	.04	.24
Touching:						
6-B	.08	.01	.64	.03	.00	.04
7-B	.05	.04	.71	.05	-.02	.02
8-B	.02	.06	.68	.07	.01	.05
9-B	.09	.06	.68	-.03	.01	.10
10-B	.05	.08	.62	.03	.07	.23
11-B	.13	.06	.55	.09	.02	.10
Same # Match:						
13-C	.14	.25	.15	.55	.09	.02
14-C	-.04	.60	-.05	-.10	.06	.19
15-C	.05	.67	.11	.10	-.12	.16
16-C	.10	-.02	.06	.68	-.02	-.07
17-C	.09	.13	-.01	.11	-.72	-.09
18-C	.05	.68	.06	.19	-.08	.11
19-C	.19	.12	.12	.63	-.27	-.04
20-C	-.09	-.08	-.03	-.07	-.67	-.02
Same Order Match:						
22-C	.02	-.02	-.04	.65	.03	.19
23-C	.11	.69	.10	.04	.02	-.09
24-C	-.06	.09	-.04	.00	-.60	.12
25-C	-.24	.01	-.02	.52	.08	.20
26-C	.14	.53	.04	-.05	.02	-.22
27-C	-.04	.57	.12	.41	-.26	-.03

Test used in the Home Start pilot study (Hi-Scope, 1973). Out of ten factors generated by the Home Start data, one was entirely the Counting subtest items; one was entirely the Touching subtest items; and the remaining eight were only identifiable in terms of the correct responses on the items of the Same Number Matching and Same Order Matching subtests.

The percent of children passing each item are listed in Table 22 for five age groups of the Fall 1971 HSPV sample. For all ages the Touching items as a group were easiest while several of the Same Order Matching items were among the hardest; this pattern was also found in the ETS Study (Shipman, 1972) and Home Start Study (Hi-Scope, 1973). In all three studies, the last item (7 lollipops) in the Same Number Matching subtest was the hardest and the last item (2 turtles) in the Same Order Matching subtest was the easiest.

In the HSPV Study, the difficulty of counting and saying how many there are were about the same except for the older (5½ year old) children where counting was harder. In the Home Start Study (Hi-Scope, 1973), counting in every instance was easier than saying how many things were counted; this discrepancy can probably be explained by the easier scoring system used for these items in the HSPV Study.

Item intercorrelations and item-total correlations (not corrected for overlap) are reported in Table 23 for the Fall 1971 HSPV sample. In general, the item intercorrelations were

TABLE 22

ETS ENUMERATION: PER CENT PASSING EACH ITEM

Items		Ages[1] 3½	4	4½	5	5½
Counting	1a	10	21	34	47	61
	1b	10	23	28	46	63
	2a	00	21	25	33	49
	2b	00	17	24	43	61
	3a	10	13	16	21	39
	3b	10	21	26	42	58
Touching	1	50	74	82	84	89
	2	60	67	69	80	79
	3	50	66	65	75	76
	4	40	55	56	64	76
	5	50	40	44	55	69
	6	30	31	32	42	51
Same Number Matching	1	20	31	39	41	53
	2	50	32	32	35	43
	3	30	49	53	59	76
	4	10	20	22	25	31
	5	50	35	34	34	33
	6	50	46	47	57	76
	7	20	22	35	38	50
	8	20	16	16	09	10
Same Order Matching	1	20	13	22	26	29
	2	60	42	45	58	68
	3	40	33	32	28	34
	4	10	16	18	25	27
	5	40	30	32	36	41
	6	90	69	79	87	88
N =		10	142	267	197	141

[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).

TABLE 23

ETS ENUMERATION TOTAL/ITEM INTERCORRELATIONS

		Counting Items (A)						Touching Items (B)						Same Number Matching (C)								Same Order Matching (D)					
		1a	1b	2a	2b	3a	3b	1	2	3	4	5	6	1	2	3	4	5	6	7	8	1	2	3	4	5	6
	1b	.33																									
Counting	2a	.49	.35																								
Items	2b	.36	.63	.35																							
(A)	3a	.39	.31	.40	.37																						
	3b	.32	.64	.37	.66	.36																					
	1	.19	.17	.20	.17	.14	.16																				
	2	.18	.16	.18	.14	.15	.18	.39																			
Touching	3	.20	.15	.18	.11	.16	.18	.32	.41																		
Items	4	.22	.20	.22	.17	.17	.21	.40	.35	.34																	
(B)	5	.24	.20	.22	.19	.26	.20	.25	.36	.33	.37																
	6	.18	.19	.23	.18	.19	.22	.22	.27	.32	.29	.36															
	1	.13	.18	.15	.18	.13	.17	.11	.13	.13	.13	.15	.11														
	2	.05	.10	.08	.05	.02	.10	.02	.04	.05	.04	.04	.01	.07													
	3	.15	.16	.14	.14	.11	.19	.10	.11	.12	.12	.14	.13	.23	.28												
Same	4	.10	.09	.11	08	.08	.07	.05	.08	.09	.03	.06	.06	.27	.01	.04											
Number	5	.02	.01	.01	.04	.03	.00	-.03	.02	.01	-.02	-.02	-.01	-.00	.01	.19	.06										
Matching	6	.14	.14	.11	.16	.13	.14	.09	.09	.11	.10	.10	.11	.25	.27	.47	.14	.15									
(C)	7	.14	.17	.14	.15	.16	.14	.11	.08	.12	.08	.09	.18	.27	.02	.20	.36	.28	.23								
	8	-.05	-.09	-.06	-.07	-.04	-.10	-.02	-.03	-.08	-.03	-.06	-.04	-.04	-.01	-.03	-.01	.24	-.05	-.01							
	1	.15	.12	.13	.12	.12	.08	.03	.05	.05	.01	.03	.08	.27	.03	.08	.26	.05	.08	.28	-.02						
Same	2	.09	.14	.06	.13	.03	.11	.06	.08	.09	.07	.12	.13	.19	.26	.34	.04	.08	.34	.14	-.02	-.00					
Order	3	-.01	-.01	-.02	-.02	.07	.01	-.02	-.01	.01	-.02	-.04	-.01	.03	.06	.10	.02	.19	.10	.10	.18	.01	.02				
Matching	4	.07	-.00	.03	.01	-.01	-.01	.02	.04	.01	-.01	.05	.04	.14	-.01	.05	.16	.01	.07	.16	-.01	.22	.03	.01			
(D)	5	-.00	.08	.02	.03	.02	.05	.02	.05	.02	.05	.03	.06	.07	.17	.20	.03	.09	.24	.07	-.03	.01	.32	.01	-.05		
	6	.08	.08	.06	.10	.03	.05	.09	.13	.13	.10	.09	.09	.31	.21	.38	.22	.23	.40	.31	.11	.21	.39	.15	.19	.20	
Counting Total		.67	.76	.68	.79	.64	.78	.24	.23	.23	.28	.30	.28	.22	.10	.20	.12	.03	.19	.21	-.10	.17	.13	.00	.02	.05	.10
Touching Total		.30	.27	.31	.24	.28	.29	.61	.69	.68	.69	.69	.64	.19	.05	.18	.09	-.01	.15	.17	-.07	.06	.14	-.02	.04	.06	.16
Same Number Total		.18	.21	.17	.20	.16	.19	.12	.14	.15	.12	.13	.14	.53	.43	.62	.46	.47	.64	.60	.20	.26	.36	.19	.14	.22	.55
Same Order Total		.12	.14	.09	.12	.09	.10	.06	.11	.10	.06	.09	.13	.32	.24	.37	.23	.21	.40	.33	.06	.45	.60	.40	.42	.52	.67
Partial Total[1]		.54	.59	.55	.58	.51	.60	.44	.48	.48	.49	.51	.48	.43	.26	.46	.31	.22	.45	.44	.01	.23	.29	.08	.09	.15	.36
Total Score		.50	.54	.50	.53	.46	.54	.39	.44	.44	.44	.46	.45	.46	.29	.50	.41	.25	.50	.48	.03	.32	.42	.18	.20	.27	.50

n = 1135 (total HSPV sample)

[1]Partial total score = sum of counting, touching and same number matching scores.

very low. The highest correlations, generally in the .30-.40 range, were between items within the Counting subtest and between items within the Touching subtest. Most of the items in the Same Number Matching subtest and in the Same Order Matching subtest correlated lowly with other items both in the same subtest and in other subtests. The same pattern of item intercorrelations can be found in the Home Start analyses of the ETS Enumeration Test (Hi-Scope, 1973).

The item-subtotal correlations from both the HSPV data and the Home Start data show that each item correlated highest with its particular subtest score. In all cases an item correlated higher with its subtest score than with the total score.

In the HSPV data, the item-total correlations were highest for the Counting subtest items (mean = .51, range = .46 - .54) and lowest for the Same Order Matching items (mean = .32, range = .18 - .50). The same findings were found in the ETS data (Shipman, 1972) and in the Home Start data (Hi-Scope, 1973). The item-total correlations for the matching subtests found in the Home Start Study were much lower than those found in the HSPV data and ETS data. Almost half of these correlations in the Home Start data were negative while all were positive in the other two studies.

The correlations among the four subtest partial and total scores for the Fall 1973 HSPV sample are listed in Table 24. The Same Order Matching subtest correlated least with the Counting subtest (.151) and the Touching subtest (.140), and

TABLE 24

INTERCORRELATIONS OF ETS ENUMERATION SUBTEST SCORES, PARTIAL SCORE AND TOTAL SCORE FOR FALL 1971 HSPV SAMPLE[1]

	ETS Total Score[2]	ETS Partial Score	Counting Subtest	Touching Subtest	Same # Matching Subtest	Same # Matching Subtest
ETS Enumeration Partial Score[3]	.966					
Counting Subtest	.713	.781				
Touching Subtest	.658	.721	.390			
Same # Matching Subtest	.725	.664	.257	.202		
Same Order Matching Subtest	.611	.385	.151	.140	.554	
Same Matching Score[4]	.765	.618	.240	.199	.918	.839

[1] N = 1135. Sample includes all children with adequate information not in Level I sites.

[2] Total score = sum of all four subtest scores.

[3] Partial score = sum of counting, touching, and same number matching subtest scores.

[4] Matching score = sum of same number matching and same order matching subtest scores.

highest with the Same Number Matching subtest (.554). The Counting subtest correlated highest with the partial score (.781), while the Same Number Matching score correlated highest with the total score (.725). The subtests correlated with the total score as follows: .725 - Same Number Matching, .713 - Counting, .658 - Touching, and .611 - Same Order Matching. The correlation between the total score and partial score was .966. Similar relationships among the four subtests can be found in the pretest and posttest Enumeration scores of the Fall 1971 HSPV reliability study (see Appendix A) reported in Table 25.

In the ETS Longitudinal Study (Shipman, 1972), the subtests correlated with the total score in the following order: .88 - Counting, .70 - Same Number Matching, .63 - Touching, and .41 - Same Order Matching. Intercorrelations between the four subtests were also relatively low, ranging from .32 to .42.

Reliability

In the fall of 1971, the ETS Enumeration Test was included in a test-retest/inter-tester reliability study conducted by Huron Institute and SRI. The details are reported in Appendix A. In general, the test-retest reliability coefficients for the various subtests after two weeks (for approximately 20 children) were moderate. For the counting subtest they ranged from .496 (paraprofessional A - paraprofessional B) to .946 (paraprofessional A - paraprofessional A); for the touching

TABLE 25

INTERCORRELATIONS AMONG SUBTESTS OF THE ETS ENUMERATION TEST FROM THE TEST-RETEST/INTER-TESTER RELIABILITY STUDY

Pretest Scores (n = 129)

	Counting	Touching	Matching-same #	Matching-same order	Matching-total
Counting	1.000				
Touching	0.387	1.000			
Matching-same #	0.308	0.163	1.000		
Matching-same order	0.187	0.197	0.289	1.000	
Matching-total	0.321	0.215	0.895	0.686	1.000

Posttest Scores (n = 129)

	Counting	Touching	Matching-same #	Matching-same/order	Matching-total
Counting	1.000				
Touching	0.522	1.000			
Matching-same #	0.359	0.453	1.000		
Matching-same order	0.325	0.279	0.444	1.000	
Matching-total	0.403	0.448	0.907	0.780	1.000

subtest, .028 (expert - paraprofessional B) to .906 (paraprofessional B - paraprofessional B); for the matching numbers of objects subtest, .036 (paraprofessional B - paraprofessional A) to .847 (expert - expert); for the matching orderings of objects, .132 (paraprofessional A - expert) to .608 (paraprofessional A - paraprofessional A); for both matching subtest, .108 (paraprofesional B - paraprofessional A) to .807 (paraprofessional B - paraprofessional B). Although no significant tester effects for any of the subtests were found, close analysis of the data reveals that many of the effects were close to the .05 level of significance and may have been if the sample size was larger.

Internal reliability coefficients (KR-20's) for ETS Enumeration subtest scores, total scores and partial scores are listed in Tables 26 - 31 for the total Fall 1971 HSPV sample and selected subsamples (black, white, Mexican-American male, female, young, old, previous preschool, and no previous preschool). The KR-20's for the total sample (n = 1135) were .681 for the Counting subtest, (Table 26), .622 for the Touching subtest (Table 27), .508 for the Same Number Matching subtest (Table 28), .354 for the Same Order Matching subtest (Table 29), .751 for the partial test (Table 30), and .766 for the total test (Table 31). For 85 subsamples with a size greater than 20, the KR-20's for the total test scores ranged from .492 for young Mexican-American males with no previous preschool experience (n = 32) to .811 for white females with no previous preschool experience (n = 179). One-third of the estimates were greater than .75. KR-20's for the partial test scores

calculated for the same 85 subsamples ranged from .378 for young Mexican-American males with no previous preschool experience (n = 32) to .800 for old white females with no previous preschool experience (n = 97). About one-fourth of the estimates were greater than .75.

The internal consistency (coefficient alpha) reliability of ETS Enumeration I scores from Year 1 of the ETS Study was .85 (n = 1459). The alpha estimates for ETS Enumeration II scores from Year 2 were .77 for the total scores, .88 for Counting scores, .57 for Touching/Pointing scores, .41 for Same Number Matching scores, and .11 for Same Order Matching scores (n = 1194 - 1292) (Shipman, 1972).

Coefficient alphas computed for the pilot Home Start sample (size not given) scores were .80 for the Counting scores, .78 for the Touching scores, .16 for the Same Number Matching scores, and -.07 for the Same Order Matching scores (Hi-Scope, 1973).

In all three studies the most reliable subtests were the first two: Counting and Touching. The higher Counting reliability in the Home Start Study and ETS Study may be partially due to the more lenient scoring systems used. The Same Order Matching reliability estimate was the lowest -- and unacceptable -- in all studies. Because of the low internal reliability estimates and poor item/score characteristics, the Same Order subtest was dropped from further HSPV analyses.

TABLE 26

KR-20 RELIABILITIES FOR FALL 1971

ETS ENUMERATION - COUNTING SUBTEST SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	1135	2.205	2.069	.681
Black	467	2.009	1.935	.650
White	486	2.461	2.167	.698
Mexican-American	165	2.024	2.054	.695
Male	588	2.039	2.022	.679
Female	547	2.384	2.104	.682
Young[3]	499	1.353	1.593	.592
Old	632	2.864	2.149	.686
Previous Preschool	275	2.818	2.097	.671
No Previous Preschool	835	2.032	2.032	.680

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

[2]Maximum score = 6.

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

TABLE 27

KR-20 RELIABILITIES FOR FALL 1971

ETS ENUMERATION - TOUCHING SUBTEST SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	1135	3.868	1.823	.622
Black	467	3.597	1.859	.620
White	486	4.008	1.786	.621
Mexican-American	165	4.133	1.736	.618
Male	588	3.660	1.822	.612
Female	547	4.091	1.798	.629
Young[3]	499	3.409	1.882	.627
Old	632	4.217	1.690	.596
Previous Preschool	275	4.120	1.759	.617
No Previous Preschool	835	3.773	1.835	.622

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

[2] Maximum score = 6.

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

TABLE 28

KR-20 RELIABILITIES FOR FALL 1971

ETS ENUMERATION - SAME NUMBER MATCHING SUBTEST SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	1135	3.082	1.774	.508
Black	467	2.732	1.868	.517
White	486	3.288	1.893	.525
Mexican-American	165	3.461	1.495	.232
Male	588	3.092	1.837	.486
Female	547	3.071	1.912	.532
Young[3]	499	2.685	1.753	.449
Old	632	3.380	1.901	.532
Previous Preschool	275	3.251	1.880	.519
No Previous Preschool	835	3.036	1.871	.506

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

[2] Maximum score = 8.

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

TABLE 29

KR-20 RELIABILITIES FOR FALL 1971

ETS ENUMERATION -- SAME ORDER MATCHING SUBTEST SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	1135	2.515	1.365	.354
Black	467	2.257	1.329	.333
White	486	2.638	1.400	.383
Mexican-American	165	2.861	1.154	.152
Male	588	2.522	1.337	.320
Female	547	2.506	1.395	.388
Young[3]	499	2.214	1.348	.358
Old	632	2.741	1.327	.317
Previous Preschool	275	2.571	1.387	.377
No Previous Preschool	835	2.505	1.362	.349

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

[2] Maximum score = 6.

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

TABLE 30

KR-20 RELIABILITIES FOR FALL 1971

ETS ENUMERATION PARTIAL SCORES

(COUNTING + TOUCHING + SAME NUMBER MATCHING SUBTESTS)

	n	mean[2]	S.D.	KR-20
Total[1]	1135	9.155	4.175	.751
Black	467	8.338	3.929	.721
White	486	9.757	4.342	.771
Mexican-American	165	9.618	3.797	.714
Male	588	8.791	4.084	.741
Female	547	9.547	4.236	.760
Young[3]	499	7.447	3.591	.687
Old	632	10.460	4.092	.744
Previous Preschool	275	10.189	4.047	.737
No Previous Preschool	835	8.841	4.175	.753

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

[2] Maximum score = 20.

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

TABLE 31

KR-20 RELIABILITIES FOR FALL 1971

ETS ENUMERATION TOTAL SCORES

	n	mean[2]	S.D.	KR-20
Total[1]	1135	11.670	4.866	.766
Black	467	10.595	4.468	.726
White	486	12.395	5.110	.789
Mexican-American	165	12.479	4.410	.731
Male	588	11.313	4.706	.750
Female	547	12.053	5.005	.781
Young[3]	499	9.661	4.270	.715
Old	632	13.201	4.691	.751
Previous Preschool	275	12.760	4.678	.749
No Previous Preschool	835	11.346	4.883	.769

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

[2] Maximum score = 26.

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

Correlations with Other Tests

Correlations of the ETS Enumeration partial scores and of the three subtest scores (Counting, Touching, and Same Number Matching) with other tests in the Fall 1971 HSPV battery (WRAT, PPVT, 32-item PSI, ITPA Verbal Expression Subtest, Brown, MI-Truck, and Eight-Block Sort) are listed in Table 32. The highest correlation for the partial score was .584 (N = 1073) with the 32-item PSI. Other tests with correlations over .40 were WRAT - Dot Counting (.542), WRAT - Copying Marks (.508), PPVT (.475), ITPA Verbal Expression Subtest (.459), WRAT - Reading Numbers (.446), WRAT - Recognizing Letters (.427), and Eight-Block Sort (.442 - total success, .405 - reason). The highest correlations with the Counting subtest scores were .625 with the 32-item PSI, .620 with the WRAT - Dot Counting, .504 with the WRAT - Copying Marks, and .500 with the WRAT - Reading Numbers. The highest correlations with the Touching subtest scores were .383 with the WRAT - Dot Counting and .382 with the 32-item PSI. None of the correlations with the Same Number Matching Subtest scores were greater than .298 (with the ITPA - Verbal Expression Subtest).

Correlations of ETS Enumeration I with other Year 1 tests in the ETS Study were low; the highest correlation was .34 with the Form Reproduction total score which has a perceptual component. The Touching subtest of ETS Enumeration II in Year 2 also correlated in the .30's

TABLE 32 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	[illegible] (2881)	.375 (2995)															
WRAT-NAME LETTERS	.346 (2881)	[illegible] (2995)	.302 (2995)														
WRAT-READ NUMBERS	.407 (2881)	.412 (2995)	[illegible] (2995)	.630 (2995)													
WRAT-DOT COUNTING	.451 (2881)	.403 (2995)	.419 (2995)	.344 (2995)	.351 (2995)												
PSI (32-item)	[illegible]	.353 (2860)	.451 (2860)	.414 (2860)	[illegible] (2860)	.369 (2860)											
ITPA-VERBAL EXPRESSION	.357 (1147)	.319 (1172)	.371 (1172)	.276 (1172)	.341 (1172)	[illegible] (1172)	.506 (1138)										
ETS ENUMERATION TOTAL[2]	[illegible] (1075)	.508 (1097)	.427 (1097)	.367 (1097)	.445 (1097)	.517 (1097)	.582 (1073)	.359 (1115)									
ETS ENUMERATION COUNTING	.492 (1075)	.504 (1097)	[illegible] (1097)	.359 (1097)	.500 (1097)	[illegible] (1097)	.625 (1073)	.351 (1115)	.761 (1135)								
ETS ENUMERATION TOUCHING	.282 (1075)	.355 (1097)	.273 (1097)	[illegible] (1097)	[illegible] (1097)	[illegible] (1097)	[illegible] (1073)	[illegible] (1115)	.721 (1135)	.390 (1135)							
ETS ENUMERATION SAME # MATCHING	[illegible] (1075)	[illegible] (1097)	[illegible] (1097)	[illegible] (1097)	.176 (1097)	.155 (1097)	.132 (1073)	[illegible] (1115)	[illegible] (1135)	.257 (1135)	.292 (135)						
BROWN-UNADJUSTED	[illegible] (2649)	[illegible] (2753)	[illegible] (2753)	.335 (2753)	[illegible] (2753)	[illegible] (2753)	.323 (2649)	[illegible] (1145)	[illegible] (1073)	[illegible] (1073)	.160 (1073)	.054 (1073)					
BROWN-ADJUSTED	[illegible] (2649)	[illegible] (2753)	[illegible] (2753)	[illegible] (2753)	[illegible] (2753)	[illegible] (2753)	[illegible] (2665)	[illegible] (1145)	[illegible] (1073)	[illegible] (1073)	.134 (1073)	.032 (1073)	.637 (2879)				
MI-TRUCK[3]	[illegible] (637)	[illegible] (625)	[illegible] (625)	[illegible] (625)	[illegible] (625)	[illegible] (625)	[illegible] (608)	[illegible] (637)	.136 (597)	.135 (597)	.047 (597)	.107 (597)	.118 (610)	.109 (610)			
EIGHT-BLOCK PLACEMENT	.304 (1119)	[illegible] (1148)	[illegible] (1148)	[illegible] (1148)	[illegible] (1148)	[illegible] (1148)	[illegible]	[illegible] (1096)	[illegible] (1032)	.413 (1032)	.200 (1032)	.180 (1032)	.212 (1113)	.183 (1113)	.005 (573)		
EIGHT-BLOCK REASONS	.445 (1130)	[illegible] (1148)	.322 (1148)	[illegible] (1148)	[illegible] (1148)	[illegible] (1148)	[illegible] (1070)	.418 (1096)	[illegible] (1032)	[illegible] (1032)	.238 (1032)	.214 (1032)	.178 (1113)	.168 (1113)	.063 (573)	.520 (1211)	
EIGHT-BLOCK SUCCESS TOTAL	.439 (1119)	[illegible] (1148)	.324 (1148)	[illegible] (1148)	.344 (1148)	[illegible] (1148)	[illegible] (1030)	[illegible] (1096)	[illegible] (1032)	.416 (1032)	.266 (1032)	.226 (1032)	.220 (1113)	.209 (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.

with the Form Reproduction score (.35). Correlations of the touching items with the 64-item PSI were also in the .30's in both years. The highest correlation of the ETS Enumeration II in Year 2 was .69 with the 64-item PSI. Other correlations of interest were .53 with the Form Reproduction test, .53 with the PPVT, .49 with the TAMA General Knowledge Test, and .40 with the Eight-Block Sort total score. Correlations of the Enumeration scores with the other numerical test in the battery, Spontaneous Numerical Correspondence, were low and positive in both years (.22, .38) (Shipman, 1972).

Remarks

One of the most appealing aspects of the ETS Enumeration Test is that it systematically attempts to measure the various components of the cognitive process involved in learning mathematical skills. Since this test is relatively new and in the developmental stages, more technical analyses and refinement of this test are needed and encouraged.

All three studies in which this test has been used with preschool children found that the Same Order Matching items have a low internal reliability estimate and a low correlation with the total score. It is therefore recommended, as do the authors of the ETS report (Shipman, 1972) and of the

the Home Start report (Hi-Scope, 1973), that the order subtest be dropped from future studies. The information regarding the Same Number Matching subtest is still conflicting. Since this subtest correlated high with the total score (.70's) and had moderate internal reliability estimates (.50's) in both the ETS Study and the HSPV Study, the Same Number Matching should not be dropped from the test at the present time. Future work with this subtest included must be done before the Home Start authors' recommendation to exclude it can be supported.

Several other questions need to be answered in future analyses:

1. What is the relative effect of using the Counting scoring system used in the HSPV Study? Is it a more accurate description of a child's developmental abilities in enumeration? Would the more strict scoring system used in the ETS Study and the Home Start Study be better?
2. Is there a strong tendency for younger children to select pictures in the middle of the page on the matching subtests?
3. Predictive validity estimates for the test as a whole and the individual subtests are needed.
4. Further item refinement in all the subtests is needed.

References

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

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

Piaget, J. The child's conception of number. New York: Norton, 1965.

Potter, M. C., & Levy, E. L. Spatial enumeration without counting. Child Development, 1968, 39, 265-272.

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. 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, HEW, under grant # H-8526. Princeton, N. J.: Educational Testing Service, 1972.

Gumpgookies

Purpose

Gumpgookies is a test designed to measure young children's (ages 3 to 8) motivation to achieve in school. Many educators and psychologists believe that the motivation to achieve in school is crucial in determining academic success. Thus, this variable is a goal of many preschool programs. The basis for investigation in achievement motivation was established in the 1950's by McClelland and his followers (McClelland, 1958; McClelland et al., 1953). Extensive work on the development of a measure of achievement motivation for young children and of a preschool motivation curriculum has been done at the University of Hawaii under the direction of Dorothy Adkins and Bonnie Ballif (1970, 1971).

Description

Each of the Gumpgookie items (27 in the HSPV version) presents a semi-structured story about two imaginary figures called Gumpgookies. After the examiner reads the story and describes the actions of the two Gumpgookies in the item situation, the child chooses the Gumpgookie in the story that is most like him. A summary of the item choices is as follows:

1. Will try later to hit the ball--tries to hit the ball.
2. Cannot tell what this story is about--can tell what this story is.

3. Will forget to do it--will do it.
4. Playing--reading.
5. Cannot find the things it makes--keeps the things it makes.
6. Likes to learn--likes to play all the time.
7. Shows its paintings to others--hides its paintings.
8. One's mother makes it go to school--one wants to to to school.
9. One's house is almost finished--one's house fell down.
10. Stopped trying to win--kept trying to win.
11. Tries to do things well--does not care.
12. Does not like the teacher--likes the teacher.
13. Does not look at the board--looks at the board.
14. Always does its best--does its best when someone is watching.
15. Is working--is looking around.
16. Wants to do well in school--is doing well in school.
17. Tired of school--not tired of school.
18. Helps the teacher--plays with things.
19. Watching--trying to write.
20. Learns one new thing--learns lots of new things.
21. Can point to the letter B--thinks all letters look the same.
22. Thinks school is fun--is tired of school.
23. Will never win--will win someday.
24. Is making another painting--is sitting down.
25. Thinks it will be a good day--thinks it will be a bad day.
26. Is getting tired--is getting smarter.
27. Steps on some ants in the dirt--puts the ants in a bottle.

Before the test is begun, the child is given several practice items in which he must indicate that what he likes to do best is the same as what his Gumpgookie likes to do.

Items are scored either one or zero. One is given when the subject responds in the direction assumed to indicate the presence of an achievement motivation component.

Development of Instrument

Gumpgookies was developed by Adkins and Ballif from a theoretical framework which assumes five components of the

motivation to achieve "(1) an affective component, expressed as positive affect from achievement; (2) a conceptual component, whereby the individual sees himself as an achiever; (3) a purposive component, enabling the individual to establish and respond to future goals; (4) a cognitive component, by means of which the instrumental steps necessary to attain goals are known; and (5) an ethical component, through which the individual can evaluate his own performance (Adkins & Ballif, 1970, p. 138)."

From an original 200 item instrument which was pilot tested with Head Start children in Hawaii, Adkins and Ballif developed three forms of Gumpgookies: (1) a 75-item form, which is individually administered to preschoolers; (2) a 100-item form, which is group administered to non-reading elementary children; and (3) a 100-item form, which is group administered to elementary children who can read. Factor analyses of these preliminary forms showed that scores were partially determined by the position of the item in the test (primacy effects--near the beginning, and recency effects--near the end), the position of the correct alternative on the item page (i.e., right or left), and the order in which the two alternatives were presented to the child. Because of the effects of these extraneous variables, a new format for each of the three forms was created so that the order in which the pictures were presented to the child was randomized.

The new randomized 75-item individual form was given

to 10 different ethnic subgroups of 1067 preschool children, geographically scattered across the United States. The appropriate randomized group forms were given to 668 children in the first, second and fourth grades in Hawaii. Five factors paralleling the main components of the theory were found for the elementary school group. The two strongest factors across grade levels were the conceptual and the affective. Only tentative interpretations of factors based on the combined preschool group could be made. According to Adkins and Ballif, it should not be surprising that factors differ with age groups, since much evidence indicates that motivation represents learned behaviors and factor structures may become more complex as the child grows older.

These analyses also showed that, even with randomized forms, extraneous influences on scores were consistently found. Young children seem to be more influenced by the position of the items on the page and the order in which answers are presented, while the older children seem to be more influenced by primary/recency effects. Thus, Adkins and Ballif conclude that "there seems to be no escape from the fact that total scores for individual children may be distorted by their idiosyncratic proclivities to be affected by irrelevant tendencies (1970, p. 140)."

Since it was felt that both the 60-item version, developed for use in the Follow Through evaluation, and the 75-item version were too long for use with children in the HSPV sample, a 27-item experimental version was created. The 27 items selected

had the highest loadings on the first four factors (instantaneous activity, school enjoyment, self evaluation, purposiveness) found by Adkins (1972) in a factor analysis of 1800 four-and five-year-old children.

Reliability

KR-20 reliability coefficients for Gumpgookies range from the low to mid .80's for the 75-item form and from the upper .80's to low .90's for the 100-item form. Test-retest reliabilities for comparable forms are not available. However, the test-retest reliability for the 100-item non-randomized form with 75-item randomized form was .66 after three weeks (n=44) (Adkins & Ballif, 1970).

Internal consistency reliability coefficients (KR-20's) for the Spring 1972 HSPV sample are listed in Table 1. The KR-20 for the total sample (n = 1885) was .730. The estimates for approximately 90 subsamples with a sample size greater than 20 ranged from .625 for young black females with no pre-school experience (n = 160) to .904 for male Mexican-American children with previous preschool experience. Approximately one-sixth (17%) of the KR-20 estimates were over .800. Since these reliability estimates were computed from the spring sample, they should be considered cautiously and not compared to KR-20's of other tests given in the fall of the same year before the Head Start Planned Variation experience had begun.

TABLE 1

KR-20 RELIABILITIES FOR SPRING 1972 GUMPGOOKIES SCORE

	n	Mean[2]	S.D.	KR-20
Total[1]	1885	19.389	4.385	.730
Black	857	20.801	3.998	.717
White	752	20.614	4.813	.800
Mexican-American	244	20.373	4.407	.762
Male	976	20.437	4.475	.764
Female	909	20.910	4.272	.757
Young[3]	72	19.389	4.385	.730
Old	1108	21.548	4.164	.763
Previous Preschool	482	21.006	4.539	.785
No Previous Preschool	1357	20.547	4.334	.752

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

[2]Maximum score = 27.

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

Validity.

One demonstration of the test's validity is that teachers' ratings of motivation, based on several different procedures, tend to substantiate that Gumpgookie scores do discriminate between children rated as high in motivation and those rated as low in motivation (Adkins & Ballif, 1970).

Significant correlations between Gumpgookies and other tests were obtained for an Hawaiian Head Start sample during the 1968 national evaluation study: .24 with the Stanford-Binet, .31 with the PSI, and .23 with the Psycholinguistic Age score of the ITPA. The correlation of the 100-item randomized form of Gumpgookies with the Children's Self Concept Index, which was used in the Westinghouse/Ohio Head Start evaluation, was .43 for a sample of 104 second graders in Hawaii. On a sample of first, second, and fourth graders in Hawaii, Gumpgookies correlations with several tests of academic achievement, such as the Metropolitan Readiness Test and the California Reading Test, were low but statistically significant.

Correlations of Gumpgookies with other tests in the HSPV battery are not reported here since Gumpgookies was only given in the Spring. Any correlations would be confounded by treatment effects.

Head Start Planned Variation Score Characteristics

The distribution of the Gumpgookies scores for all children in the Spring 1972 sample is presented in Table 2.

TABLE 2

DISTRIBUTION OF GUMPGOOKIES SCORES FOR ALL CHILDREN IN THE SPRING 1972 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	5	18.200	4.707
39-41	2	22.000	2.000
42-44	10	17.400	2.835
45-47	39	18.641	4.306
48-50	150	19.233	4.303
51-53	273	19.392	4.108
54-56	290	19.662	4.632
57-59	282	20.429	4.164
60-62	251	21.398	5.183
63-65	181	21.144	4.591
66-68	175	22.303	3.612
69-71	152	22.645	3.671
72-74	64	23.469	3.648
75-77	3	23.000	2.160
78-80	2	22.000	4.000
TOTAL	1879	20.668	4.387

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

[2]Maximum score = 27.

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 = 1879) was 20.668 (S.D. = 4.387).

The distributions of Gumpgookies scores in the spring for all planned variation children and all non-planned variation children are negatively skewed (see Tables 3 and 4). From Table 3 it can be seen that there is a ceiling effect for scores of the planned variation children; approximately 10.6% of the children scored at the top two scores while approximately 21.4% scored at the top three scores. Table 4 shows the negatively skewed distribution of scores for the non-planned variation children; approximately 15.4% scored at the top three scores.

Norms

Two sets of age norms have been developed for Gumpgookies. One set of norms is for the 55-item version of Gumpgookies that was used in the spring of the 1968-69 national evaluation of Head Start. These norms, based upon pretest scores of 1485 children in the 1968-69 Head Start sample, range from 43 to 61 months (Adkins & Ballif, 1970). The other set of norms is for the 75-item randomized individual form of Gumpgookies. These norms represent a total of 1588 children ranging in age from 39 to 76 months (Adkins & Payne, 1971).

TABLE 3

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

Score	# of Children	
0	1	
1	0	
2	0	
3	1	
4	3	Each x represents 5 children
5	2	
6	3	
7	6	x
8	10	xx
9	7	x
10	11	xx
11	10	xx
12	22	xxxx
13	25	xxxxx
14	39	xxxxxxxx
15	42	xxxxxxxx
16	53	xxxxxxxxxxx
17	52	xxxxxxxxxxx
18	72	xxxxxxxxxxxxxx
19	96	xxxxxxxxxxxxxxxxxx
20	109	xxxxxxxxxxxxxxxxxxxx
21	107	xxxxxxxxxxxxxxxxxxx
22	135	xxxxxxxxxxxxxxxxxxxxxxxxxx
23	133	xxxxxxxxxxxxxxxxxxxxxxxxx
24	155	xxxxxxxxxxxxxxxxxxxxxxxxxxxxx
25	150	xxxxxxxxxxxxxxxxxxxxxxxxxxxx
26	109	xxxxxxxxxxxxxxxxxxxx
27	38	xxxxxxxx

Total N = 1391

TABLE 4

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

Score	# of Children	
0	0	
1	0	
2	0	
3	0	
4	2	xx
5	0	
6	0	
7	3	xxx
8	1	x
9	4	xxxx
10	1	x
11	6	xxxxxx
12	5	xxxxx
13	12	xxxxxxxxxxxx
14	13	xxxxxxxxxxxxx
15	11	xxxxxxxxxxx
16	20	xxxxxxxxxxxxxxxxxxxx
17	33	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
18	36	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
19	27	xxxxxxxxxxxxxxxxxxxxxxxxxx
20	48	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
21	36	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
22	51	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
23	52	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
24	57	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
25	40	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
26	26	xxxxxxxxxxxxxxxxxxxxxxxxx
27	10	xxxxxxxxxx

N = 494

Remarks

Even though there is an impressive amount of technical analysis that has been done on all the longer forms of Gumpgookies, it is not clear exactly what the test is measuring. This seems especially true for the longer preschool forms since the extraneous "item location" variables definitely affect the scores and factor analyses yield no conclusive results. To say that Gumpgookies measures achievement motivation seems premature since the relationship between Gumpgookies and achievement measures needs to be explored in further studies. For instance, are the low correlations of Gumpgookies with other achievement measures, such as the California Reading Test and the PSI, due to the fact that intelligence was not controlled or that Gumpgookies does not measure achievement motivation or to some other reason?

Furthermore, special caution must be used with the shortened version of Gumpgookies. Since no previous analyses have been done on the non-randomized 27-item test and it was only used in the spring of the HSPV Study, it must be considered as an experimental version in preliminary stages of development. Adkins (1972) recommended against use of this shortened version because it had not been used in advance and analyzed.

Several other questions concerning the use of Gumpgookies need to be explored further in future analyses with this test:

1. Do children tend to respond to items in "socially-accepted" ways? How much is Gumpgookies a measure of socialization?

2. What does it mean when a child's identified Gumpgookie likes to do something different from what he likes to do?

3. Do children relate to the questions with the physical objects in them? Is there a difference in their understanding and responses to these items?

4. Does the way a tester reads the questions influence a child's response?

5. Does the fact that the word "gumpgookie" is a tongue-twister have any effects on the child's test performance?

6. Are there problems in administering the test?

7. At what point (after how many items) does a child stop paying attention to the task?

References

Adkins, D. C. Personal communication, 1972.

Adkins, D. C., & Ballif, B. L. Final report--motivation to achieve in school. Honolulu: University of Hawaii, Head Start Research Center, 1970.

Adkins, D. C., & Ballif, B. L. Preschool motivation curriculum. Honolulu: Education Research and Development Center, University of Hawaii, 1971.

Adkins, D. C., & Payne, F. Age-normed "Z" scores for the 75-item randomized individual form of Gumpgookies. Honolulu: Center for Research in Early Childhood Education, College of Education, University of Hawaii, 1971.

McClelland, D. C. Methods of measuring human motivation. In J. W. Atkinson (Ed.), Motives in fantasy, action and society. New York: Van Nostrand, 1958.

McClelland, D. C., Atkinson, J. W., Clard, R. A., & Lowell, E. L. The achievement motive. New York: Appleton-Century-Crofts, 1953.

Hertzig-Birch Scoring System

Purpose

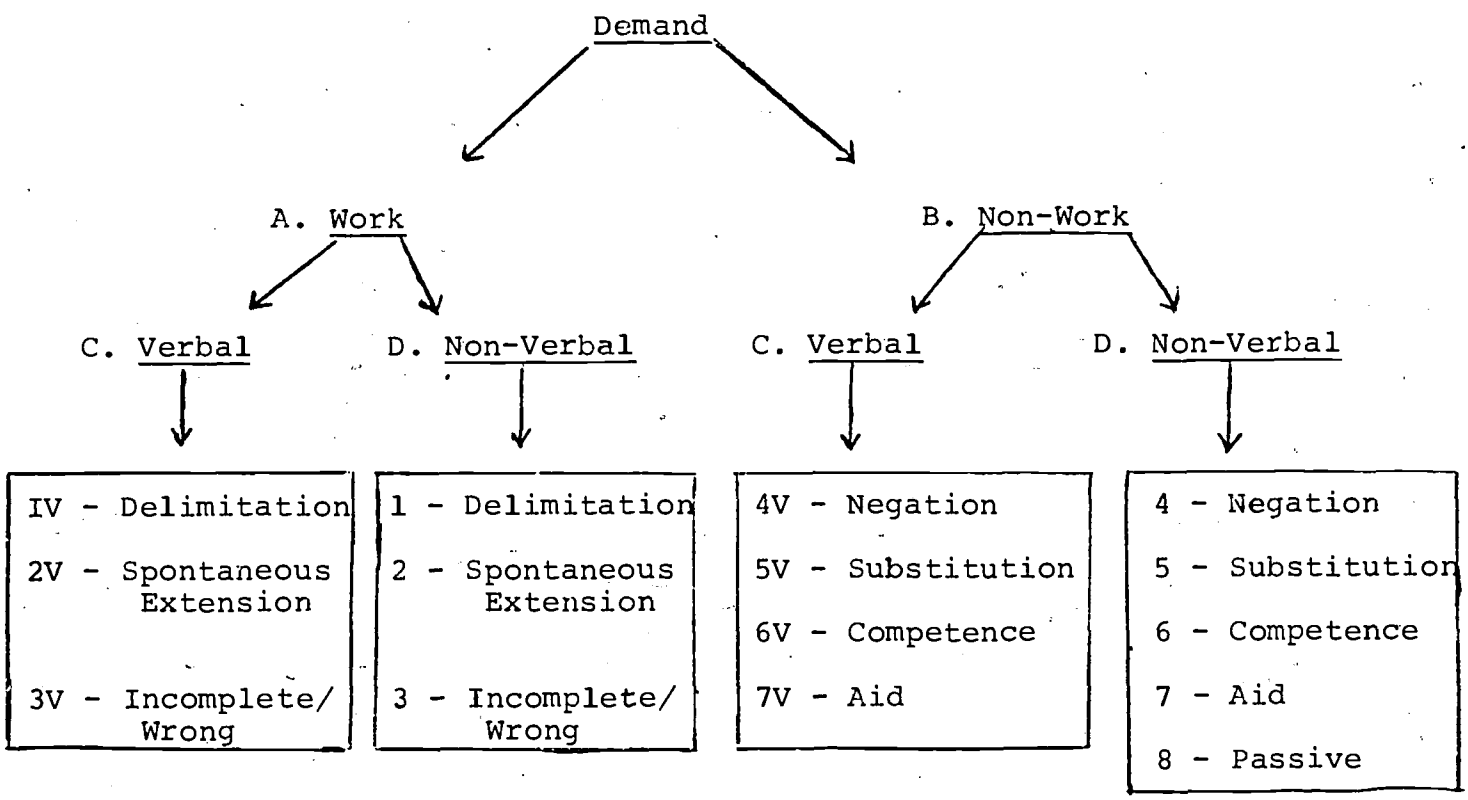
The Hertzig-Birch Scoring System is designed to assess a child's style of responding to cognitive demands in test taking situations. In addition to recording a child's response to a test item as either right or wrong, this system focuses on coping mechanisms and styles. What the child does actively or passively to handle, organize, accept or influence the environment around him is as important to his future growth and success as whether he knows the correct response to a task demand.

Description

The examiner classifies the child's behavior as a particular type of verbal or non-verbal, work or non-work response (see Fig. 1). A work response (delimitation or correct, spontaneous extension or extra, incomplete/wrong) is one in which the child is engaged in doing the task, regardless of his success or failure. A delimitation response is one in which the child responds to the tester's demand correctly and does nothing else. Spontaneous elaboration is scored if the child gives unsolicited elaboration to an item after he has completed the required task correctly. For example, a child may respond to the question, "How many toes do you have?", that he has "ten--five on one foot and five on the other foot". "O.K., yes, here, there, here it is, and

Fig. 1.

BIRCH RESPONSE STYLE CODING CATEGORIES



this one" are not coded as spontaneous extensions. An incorrect or wrong response is one where a child fails to do the required task after he has worked on it and done nothing else.

A non-work response (negation, substitution, competence, aid, and passive) is one where the child fails to work on the task presented. A negation response is a direct refusal to do work; examples are "I won't", "I don't want to", and a shaking of the head or turning away from the task. A substitution response is one in which the child does something else which is irrelevant to the task required. Examples are "I want to get a drink now" or getting up to play with other toys in the room. A competence response is one in which a child states a limitation in his ability to perform the requested task. Examples would be "I don't know how to", and "I can't" or "I'm too little to do it". An aid response is one where the child makes a direct request for help from the tester. The request must be one in which the child asks the tester to help him solve the task and not one in which the child asks the tester for clarification or explanation of the task. Examples of aid are verbal requests such as "Show me how to do it" and "Tell me what the answer is". A passive response is a "no response" category, meaning that the child does nothing at all when the tester presents the task.

A verbal response is one in which the child uses words for any purpose. The words do not have to pertain to the task required. If the child does not use any words, the response is scored non-verbal, regardless of whether the appropriate task demand was for a verbal or non-verbal response. All of the work responses and three of the non-work responses (negation, substitution and competence) can be scored verbal or non-verbal. The non-work response--aid--is usually verbal, while the non-work response--passive--is always non-verbal.

The Hertzig-Birch scoring system was used with the Stanford-Binet in the first two years of the HSPV Study and with the PSI in the final year of the HSPV Study. In both instances, the tester (and not an independent observer) used the Hertzig-Birch for the last response of the child to each task demand. In the case of the Stanford-Binet, the tester scored "+" or "-" in the "pass-fail" column before an item and used Hertzig-Birch codes in the subtest spaces for each test.

From Fall 1971 to Spring 1972 the procedure for recording the Hertzig-Birch scoring was modified. A code for spontaneous elaboration after a wrong answer was added. In the original scoring system outlined above, spontaneous elaborations were noted only for correct answers. In the fall, the tester recorded the codes by circling one or more of the following letters

for each item:

C - correct
W - wrong
E - extra
R - refusal
S - substitution
DK - don't know
A - aid
NR - no response
V - verbal

In both Fall 1971 and Spring 1972, the tester was asked to write out the child's responses and record probes that were used.

Scoring

The Hertzig-Birch scores from the first year of the Planned Variation Study were analyzed by SRI using two measures: (1) spontaneous extension, and (2) passivity/substitution (SRI, 1971). The Hertzig-Birch measure of spontaneous extension was defined as the number of elaborations (verbal and non-verbal) divided by the total number of correct responses--i.e., (2 + 2V) ÷ (1 + 1V + 2 + 2V). The Hertzig-Birch measure of passivity/substitution was defined as the ratio of passive and substitution (verbal and non-verbal) responses to all incorrect responses--i.e., (5 + 5V + 8) ÷ (3 + 3V + 4 + 4V + 5 + 5V + 6 + 6V + 7 + 7V + 8).

The scores used in analyzing the Hertzig-Birch codes for the second year Stanford-Binet data and for the third year PSI data are the frequency counts for each code.

A detailed description and analysis of the Hertzig-Birch scores for the 1970-71 HSPV data are available in another Huron Institute report (Featherstone, 1973).

Development of Scoring System

The Hertzig-Birch scoring system was adapted from the system used by Hertzig et al. (1968) in their longitudinal study of three-year-old Puerto Rican (n=60) and middle class (n=116) children in New York City. This original system of scoring responses during a Stanford-Binet testing session was developed so that a child's style of responding to a cognitive demand could be closely examined. In this original study detailed protocols of each child's verbalizations and behavior during the testing session were made by an independent observer. Later, the protocols were scored according to the work/non-work, and verbal/non-verbal distinctions described above. In this original study, all of the child's behavior was coded. Spearman rank order correlations on 30 records rescored after 8 months ranged from .93 to .97 for each individual category.

In general, Hertzig et al. found that the middle class children used a significantly higher initial proportion of work responses (.81 vs. .72), and a significantly larger proportion of initial non-work responses followed by work responses (.53 vs .42) than did the Puerto Rican children. The most frequent type

of non-work response among middle class children was competence while that of the Puerto Rican children was either substitution or aid. Even when the IQ of the children was held constant, the differences in response styles were found.

Hertzig-Birch Code Characteristics

When Used with the Stanford-Binet. Examination of frequency distributions for a selected subsample (PV children with no previous preschool experience who took the Stanford-Binet) of the 1969-70 HSPV sample reveals that about 85% of all responses were coded 1 (delimitation) or 3 (incomplete/wrong). When answering correctly, children did not generally go beyond the requirements of the task; if unable to answer correctly they still generally made a relevant "work" response. Table 1 gives mean and median frequencies of each response (per child) for the fall and spring Stanford-Binet testing, plus the percentage of children having responses of that category. The only category which changed greatly from fall to spring was substitution (code 5); the percentage of children with substitution responses decreased by almost one-half from fall to spring.

The intercorrelations of the four categories--extension, substitution, competence, and passivity--are

TABLE 1

HERTZIG-BIRCH CODE CHARACTERISTICS
WITH THE STANFORD-BINET

	Fall 1969			Spring 1970		
Codes	Mean	Median	% Children Making Responses	Mean	Median	% Children Making Responses
1	66.381	64.0	100.00	62.410	59.0	100.00
2	2.473	0.0	48.25	.940	0.0	31.43
3	37.965	36.0	100.00	38.651	38.0	94.92
4	.270	0.0	10.16	.270	0.0	9.52
5	3.165	1.0	56.83	1.044	3.3	24.44
6	4.717	3.0	69.21	7.797	3.0	66.67
7	.162	0.0	9.21	.276	1.0	12.06
8	2.984	0.0	47.62	2.959	0.0	39.05
(3-8)	49.263	47.0	100.00	50.997	50.0	100.00

Codes 1 - delimitation
2 - elaboration (correct)
3 - wrong
4 - negation
5 - substitution
6 - competence
7 - aid
8 - passive

Sample includes all PV children with no previous preschool experience.

n = 315

included in Table 2. Request for aid and refusal were not included since only 10% of the children had any responses at all in these two categories. All of these intercorrelations, even the significant ones, were small enough to justify the independence of the variables.[1]

Table 3 presents the mean, median and standard deviation for each of the codes used with the Stanford-Binet for the Fall 1970 HSPV total sample (n = 613). Most of the responses were coded 1 (delimitation) or 3 (incomplete/wrong); all other codes had medians of zero.

When Used with the Preschool Inventory. The mean and standard deviations for the Hertzig-Birch codes used with the 32-item Preschool Inventory in the third year of the Head Start Planned Variation Study are listed in Table 4 for both the Fall 1971 and Spring 1972 total HSPV sample (N = 2972). As was true for the codes used with the Stanford-Binet, codes 1 (delimitation) and 3 (wrong) were used most often.

The intercorrelations of the seven codes used with the PSI in the Fall 1971 battery are included in Table 5. All the correlations for the total HSPV sample were small (less than .30) except for the correlation between code 1 (delimitation) and code 3 (wrong) which was -.91.

---

[1]For further discussions of the score characteristics for the 1969-70 HSPV sample, as well as for correlations of several Hertzig-Birch variables with selected background variables, see Featherstone (1973).

TABLE 2

INTERCORRELATIONS OF 4 HERTZIG-BIRCH CODES

FALL 1969 (n = 315)

codes	2	5	6	8
5	.213*	1.000		
6	.063	-.043	1.000	
8	-.009	.007	-.114*	1.000

*significant at .025 level

Codes:

2 - elaboration (correct)
5 - substitution
6 - competence
8 - passive

Sample includes all PV children with no previous preschool experience.

TABLE 3

MEAN, MEDIANS AND STANDARD DEVIATIONS FOR HERTZIG-BIRCH CODES WITH THE STANFORD-BINET FOR FALL 1970 HSPV SAMPLE[1]

Codes[2]	Mean	Median	S.D.
1	28.688	27.0	13.442
2	.155	0.0	.900
3	16.837	16.0	9.029
4	.548	0.0	2.128
5	1.007	0.0	2.342
6	1.605	0.0	3.826
7	.021	0.0	.175
8	2.494	0.0	4.814

[1] N = 613. Sample includes PV and non-PV children with adequate information.

[2] Codes: 1 - delimitation
2 - elaboration (correct)
3 - wrong
4 - negation
5 - substitution
6 - competence
7 - aid
8 - passive

TABLE 4

MEAN AND STANDARD DEVIATIONS OF HERTZIG-BIRCH CODES WITH THE PSI FOR FALL 1971 AND SPRING 1972 HSPV SAMPLE

	Fall 1971		Spring 1972	
Codes[2]	Mean	S.D.	Mean	S.D.
1	14.116	6.128	18.592	5.964
2	.469	.809	.526	.920
3	15.523	5.596	11.426	5.401
wrong/extra	----	----	.125	.471
4	1.048	1.779	.032	.267
5	.313	1.161	.107	.544
6	.052	.344	.627	1.181
7	.373	1.025	.020	.209
8	.035	.244	.408	1.099

[1]N = 2972

[2]Codes: 1 - delimitation
2 - elaboration (with correct)
3 - wrong
4 - negative-refusal
5 - substitution
6 - competence - don't know
7 - aid
8 - passive - no response

TABLE 5

INTERCORRELATIONS OF HERTZIG-BIRCH CODES FOR FALL 1971 TOTAL HSPV SAMPLE[1]

Codes[2]	PSI Total	1	2	3	4	5	6	7
1	.99							
2	.11	-.02						
3	-.91	-.91	-.07					
4	-.23	-.22	-.07	-.07				
5	-.29	-.29	-.04	.10	-.03			
6	-.05	-.05	.00	-.02	.03	.04		
7	-.24	-.23	-.09	.05	.04	.07	.00	
8	-.08	-.08	-.05	.03	-.02	.10	-.02	.04

[1]Sample includes all planned variation and non-planned variation children not in Level I sites. N = 2986

[2]Codes:
1 - delimitation
2 - elaboration (correct)
3 - wrong
4 - negation/refusal
5 - substitution
6 - competence/don't know
7 - aid
8 - passive

The correlations of the Hertzig-Birch codes with the other tests in the Fall 1971 HSPV battery (PSI, PPVT, WRAT, Brown, MI, ITPA, ETS Enumeration, Eight-Block Sort Task) are listed in Table 6 for the total sample. All of the codes, except for 1 (delimitation) and 2 (correct/extra), correlate negatively with the other tests. Most of these negative correlations are very small. The only sizeable negative correlations are those of code 3 (wrong) with other tests: -.906 (PSI total), -.578 (PPVT), -.572 (ETS Enumeration - Counting Subtest), -.542 (ETS Enumeration - Partial Score) and -.531 (WRAT - Copying Marks Subtest).

SRI Follow Through Evaluation

In the Fall 1971 Follow Through evaluation (Emrick, 1972) the Hertzig-Birch codes with the 29-item PSI were included in a supplementary battery given to kindergarten and entering first grade children in 17 projects (n = 651). The codes used were essentially the same as those used with the Stanford-Binet in the first two years of the HSPV Study and with the 32-item PSI in the fall of the third year of the HSPV Study. There were two correctness codes (correct and correct with elaboration) and six incorrectness codes (wrong, refusal, substitution, "don't know", aid, and no response). There was no code for elaboration after wrong answers. All

TABLE 6 CORRELATIONS OF HERTZIG-BIRCH CODES WITH OTHER TESTS IN FALL HSPV BATTERY[1]

Hertzig-Birch Codes	PSI TOTAL	PPVT TOTAL	WRAT COPY MARKS	WRAT RECOG. LETTERS	WRAT NAME LETTERS	WRAT READ NUMBERS	BROWN UNADJ.	BROWN ADJ.	MI TRUCK[2]	ITPA	ETS COUNT.	ETS TOUCH.	ETS SAME # MATCH	ETS PART A SCORE[3]	EIGHT-BLOCK PLACE.	EIGHT-BLOCK REASON	EIGHT-BLOCK TOTAL
1-correct[4]	.991	.653	.552	.474	.409	.504	.315	.256	.168	.491	.620	.373	.228	.576	.294	.438	.432
2-correct-extra	.100	.130	.024	.078	.058	.050	.074	.032	-.032	.126	.050	.072	.034	.072	.087	.049	.074
3-wrong	-.906	-.578	-.531	-.414	-.387	-.477	-.259	-.225	-.160	-.431	-.572	-.362	-.215	-.542	-.249	-.385	-.375
4-refusal	-.228	-.189	-.079	-.115	-.106	-.115	-.129	-.047	-.056	-.149	-.173	-.047	-.063	-.135	-.106	-.157	-.155
5-substitution	-.291	-.230	-.112	-.215	-.080	-.097	-.139	-.136	-.005	-.188	-.165	-.147	-.083	-.184	-.129	-.127	-.146
6-do not know	-.054	-.001	-.035	-.044	-.015	-.041	-.012	-.016	-.037	-.015	-.064	.005	-.007	-.033	-.001	.001	.000
7-aid	-.242	-.215	-.115	-.138	-.082	-.092	-.130	-.071	.011	-.192	-.102	-.084	-.024	-.099	-.117	-.107	-.127
8-no response	-.082	-.068	-.042	-.096	-.026	-.039	-.061	-.061	-.011	-.097	-.071	-.046	-.084	-.093	-.024	-.043	-.040
	n=2986	n=2855	n=2860				n=2689		n=608	n=1138	n=1073				n=1090		

[1]Sample includes all children with adequate information, not in Level I sites.
[2]MI Scores are log transformations of slow times.
[3]Partial score = sum of counting, touching and same number matching subtest scores.
[4]Note: correct score (code 1) score - PSI total score.

responses were also coded either verbal or non-verbal. Results showed that 58% of the responses were correct (31% verbal, 27% non-verbal) and 34% of the responses were wrong (17% verbal, 17% non-verbal). Only 7% of all responses were non-work responses; about one-half of these (4% of the total responses) were no response at all. Essentially no responses were coded as non-verbal elaborations, refusals, or requests for aid. If the no response category is excluded, the percentage of verbal responses of all categories (51%) exceeds the percentage of non-verbal responses (45%).

Reliability

To estimate the reliability of several Hertzig-Birch variables, Featherstone (1973) used the 1969-70 HSPV data to assess the importance of tester differences. Testers included in the analyses had to have tested more than six children, while sites were included only if more than two testers met this criterion. For the nine sites where tester differences in IQ score and in frequencies of certain Hertzig-Birch codes could be estimated, Featherstone found that tester differences for two variables (number of extensions per child, number of passive responses) were significant above the .05 level in six out of nine sites, while they were significant in only one site for the variable-substitutions. These results suggest that some of the Hertzig-Birch variables may not be reliable.

Remarks

The inclusion of the Hertzig-Birch scoring system in a test battery allows one to look at ways in which a child copes with cognitive demands. Since there is little technical information available at this time on the scoring system, the procedure must be considered as an experimental measure. Some of this needed information will be available in a future report on the ETS Longitudinal Study which also used this scoring system.

In future evaluations the effects of having the tester, and not an independent observer, code the child's behaviors should be investigated. It is hard for an inexperienced tester to give the Stanford-Binet test well and also record the Hertzig-Birch codes. All the Binet testers in the HSPV Study were experienced testers and were given special training on the Hertzig-Birch system. It is easier for a tester to record both the test answers and the Hertzig-Birch codes for the PSI.

It should be noted that only the last response of the child was recorded. Sequences of behavior noted by the original Hertzig-Birch system are not available for analysis in the HSPV data. A suggestion for future use with the Hertzig-Birch system, especially with the Stanford-Binet, is to score at least the child's first and last response.

In her interactional analyses with the first two years of HSPV data, Featherstone (1973) found that the passive and competence responses of a child seemed to be useful in predicting interactions with the preschool model. Specifically, children high in competence responses and/or low in passive responses appeared to do better in less-directive models, while the opposite seemed true for more directive models. Even though these patterns were not "overwhelmingly strong," Featherstone "suggests that variables relating to cognitive style may be quite useful in predicting which children will make substantial gains within a particular model (p. 50)." More studies and analyses exploring these interaction hypotheses are encouraged.

Finally, further investigations using the Hertzig-Birch system should explore whether a child's way of responding to cognitive demands in a test-taking situation can be generalized to non-testing situations.

References

Emrick, J.A. Results of Fall 1971 Follow Through Supplemental Testing. Stanford Research Report under Contract #OEC-O-8-522480-4633. Washington, D.C.: Bureau of Elementary and Secondary Education, Office of Education, Department of Health, Education and Welfare, 1972.

Featherstone, H. Cognitive effects of preschool programs on different types of children. Prepared for Office of Child Development, Grant #H1926. Cambridge, Mass. Huron Institute, 1973.

Hertzig, M. E., Birch, H.G., Thomas, A., & Mendez, O.A. Class and ethnic differences in the responsiveness of preschool children to cognitive demands. Monograph of the Society for Research in Child Development, 1968, 33, No. 1.

Stanford Research Institute. Implementation of planned variation in Head Start: Preliminary evaluations of planned variation in Head Start according to Follow Through approaches (1969-70). Washington, D. C.: Office of Child Development, Department of Health, Education and Welfare, 1971.

Illinois Test of Psycholinguistic Abilities
Verbal Expression Subtest

Purpose

The purpose of the Illinois Test of Psycholinguistic Abilities (ITPA) Verbal Expression Subtest is to measure a child's ability to express himself verbally. The ITPA, from which this subtest was taken, is a diagnostic test of cognitive functioning designed to measure intraindividual differences on language, perception and short-term memory abilities.

Description

This subtest asks the child to "tell me all about this," as he is handed four familiar objects one at a time. When the practice item (a nail) is presented to the child, the tester attempts to make clear by his questions that the possible range of correct responses include object name, color, shape, composition, uses, major parts and other physical characteristics. For each of the four test items (a ball, a block, an envelope, and a button), the tester begins by saying, "tell me about this," and may prompt the child by saying, "tell me something else," or "tell me more about it." When the child stops talking or repeats himself three times or changes the subject, the tester goes on to the next object. The test items differ from the practice item in that

the tester does not ask the child specific questions about the object.

The tester writes the child's exact words in the appropriate place. Later the child's responses are scored in ten categories: (1) name, (2) color, (3) shape, (4) material, (5) use, (6) major parts, (7) number (8) other physical characteristics, (9) comparison, (10) person, place or thing commonly associated with the object or with some action of that object. A detailed description of the ten categories, which are mutually exclusive and collectively exhaustive, is presented in the Examiner's Manual for the revised edition (Kirk et al., 1968). A child's score is the sum of the number of times each category occurs for all the objects.

The coding reliabilities calculated for this subtest in the fall of 1971 were very high (Appendix D). In the standardization study of the Revised Edition of the ITPA, interscorer reliabilities for this subtest with preschool age children were very high for both experienced examiners (.98 to 1.00) and novice examiners (.97 to .99) (Paraskevopoulos & Kirk, 1969).

Development of Instrument

The ITPA, developed by McCarthy and Kirk in 1961, had nine subtests devised to measure three postulated psycholinguistic processes--receptive, expressive and organizing--at two levels of organization. The two levels are the representational level, "which requires the mediating process

of utilizing symbols which carry the meaning of an object," and the automatic level, "in which the individual's habits of functioning are less voluntary but highly organized and integrated" (Paraskevopoulos & Kirk, 1969, p. 14). The revised edition of the ITPA, developed by Kirk, McCarthy, and Kirk, has twelve subtests: auditory reception, visual reception, visual sequential memory, auditory association, auditory sequential memory, visual association, visual closure, verbal expression, grammatic closure, manual expression, auditory closure and sound blending. The Verbal Expression Subtest assesses the expressive process at the representational level. In the original ITPA battery, the Verbal Expression Subtest was called the verbal encoding subtest. Since its development the ITPA has been used in a large number of studies[1] (Buros, 1972). The only time the Verbal Expression Subtest has been used alone as a measure is in the third year of the HSPV Study and in the Fall 1971 Follow-Through Evaluation (SRI,1972).

Standardization

The most comprehensive standardization of the ITPA was done on a sample of 962 middle class children from Midwestern cities. Because the sample included only 42 non-whites, comparisons with the HSPV sample are inappropriate. In this sample the correlation between the Verbal Expression Subtest and a measure of social class was -.13 (significant at the .05 level).

---

[1]Including the Westinghouse-Ohio evaluation of Head Start.

Norms for the ITPA Verbal Expression subtest scores for the Fall 1971 HSPV sample are presented in Tables 1-8. These norm tables, based on three month age intervals (fifteen groups from 36-38 months to 78-80 months), give the number of children, the mean score and the standard deviation at each age level for the following samples: total (Table 1), males (Table 2), females (Table 3), children with no previous preschool (Table 4), children with previous preschool (Table 5), black children (Table 6), white children (Table 7), and Mexican-American children (Table 8).

Reliability

The reliability estimates for the Verbal Expression Subtest calculated on the standardization sample are fairly high, considering it is a subtest (Paraskevopoulous & Kirk, 1969). The median internal consistency coefficients (corrected for restricted intelligence range) were .80 (3-7 to 4-1 years), .86 (4-7 to 5-1 years), and .72 (5-7 to 6-1 years). Five-month stability coefficients (a test-retest reliability estimate) were .74 for the four-year-olds and .63 for the six-year-olds.

In the fall of 1971 the Verbal Expression Subtest was included in a test-retest/inter-tester reliability study conducted by Huron Institute and SRI. The details of this study using two sites in the HSPV sample are reported in Appendix A. In general, the test-retest reliability coefficients for a two-week period for approximately 20 children were high.

TABLE 1

DISTRIBUTION OF ITPA VERBAL EXPPESSION

SUBTEST SCORES FOR ALL CHILDREN

IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	5.333	5.437
39-41	1	6.000	-----
42-44	12	8.250	4.380
45-47	33	8.424	3.627
48-50	108	9.287	3.925
51-53	194	9.387	4.185
54-56	197	10.162	4.562
57-59	189	11.101	4.252
60-62	134	12.239	4.721
63-65	100	12.930	5.545
66-68	116	14.629	5.442
69-71	74	13.473	6.931
72-74	42	14.548	5.508
75-77	1	18.000	-----
78-80	---	-----	-----
TOTAL	1204	11.278	5.163

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

[2]A child's score is the number of times each category occurs for all objects.

TABLE 2

DISTPIBUTION OF ITPA VERBAL EXPRESSION

SUBTEST SCORES FOR MALES

IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	1	6.000	-----
42-44	5	7.600	3.720
45-47	19	8.000	3.784
48-50	56	8.696	3.375
51-53	110	8.600	3.631
54-56	104	10.125	4.741
57-59	92	10.902	4.176
60-62	58	12.345	4.729
63-65	48	12.813	5.648
66-68	63	14.968	5.933
69-71	37	13.270	6.966
72-74	27	13.000	4.776
75-77	1	18.000	-----
78-80	---	-----	-----
TOTAL	621	10.981	5.185

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

[2]A child's score is the number of times each category occurs for all objects.

TABLE 3

DISTRIBUTION OF ITPA VERBAL EXPRESSION SUBTEST SCORES FOR FEMALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	5.333	5.437
39-41	---	-----	-----
42-44	7	8.714	4.742
45-47	14	9.000	3.317
48-50	52	9.923	4.354
51-53	84	10.417	4.617
54-56	93	10.204	4.354
57-59	97	11.289	4.315
60-62	76	12.158	4.713
63-65	52	13.038	5.445
66-68	53	14.226	4.761
69-71	37	13.676	6.889
72-74	15	17.333	5.641
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	583	11.595	5.120

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

[2] A child's score is the number of times each category occurs for all objects.

TABLE 4

DISTRIBUTION OF ITPA VERBAL EXPRESSION SUBTEST 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	3	5.333	5.437
39-41	1	6.000	-----
42-44	12	8.250	4.380
45-47	26	7.923	3.025
48-50	87	8.977	3.451
51-53	151	9.583	4.431
54-56	170	10.147	4.536
57-59	158	10.911	4.246
60-62	86	11.523	4.764
63-65	58	11.621	4.905
66-68	75	14.360	5.137
69-71	43	14.186	6.845
72-74	28	14.714	6.005
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	898	10.878	4.972

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

[2]A child's score is the number of times each category occurs for all objects.

TABLE 5

DISTRIBUTION OF ITPA VERBAL EXPRESSION SUBTEST 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	---	-----	-----
45-47	7	10.286	4.861
48-50	18	10.722	5.596
51-53	37	8.919	3.070
54-56	22	10.818	4.896
57-59	26	11.808	4.123
60-62	46	13.435	4.431
63-65	41	14.756	5.930
66-68	39	15.205	5.979
69-71	29	12.621	6.784
72-74	14	14.214	4.329
75-77	1	18.000	-----
78-80	---	-----	-----
TOTAL	280	12.639	5.550

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

[2]A child's score is the number of times each category occurs for all objects.

TABLE 6

DISTRIBUTION OF ITPA VERBAL EXPRESSION SUBTEST SCORES FOR BLACK CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	3	5.333	5.437
39-41	---	-----	-----
42-44	8	8.125	2.472
45-47	19	8.895	3.892
48-50	53	9.453	4.342
51-53	79	9.139	3.818
54-56	82	10.049	4.155
57-59	87	11.103	3.991
60-62	46	11.978	4.019
63-65	49	11.490	5.043
66-68	45	13.333	3.950
69-71	28	11.393	4.593
72-74	14	12.143	4.565
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	513	10.655	4.401

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

[2]A child's score is the number of times each category occurs for all objects.

TABLE 7

DISTRIBUTION OF ITPA VERBAL EXPRESSION SUBTEST SCORES FOR WHITE CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score[2]	S.D.
36-38	---	-----	-----
39-41	1	6.000	-----
42-44	4	8.500	6.727
45-47	14	7.786	3.121
48-50	41	9.220	3.619
51-53	81	9.926	4.348
54-56	88	10.477	4.947
57-59	72	11.042	4.309
60-62	63	12.651	5.124
63-65	34	14.441	6.006
66-68	44	16.659	6.223
69-71	33	15.970	7.740
72-74	27	16.148	5.240
75-77	1	18.000	-----
78-80	---	-----	-----
TOTAL	503	12.028	5.754

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

[2]A child's score is the number of times each category occurs for all objects.

TABLE 8

DISTRIBUTION OF ITPA VERBAL EXPRESSION SUBTEST 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	12	9.500	2.693
51-53	33	8.576	4.466
54-56	24	9.208	4.368
57-59	27	10.926	4.906
60-62	20	12.200	4.686
63-65	16	13.750	4.841
66-68	25	13.480	5.201
69-71	13	11.615	6.878
72-74	---	-----	-----
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	170	10.971	5.182

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

[2]A child's score is the number of times each category occurs for all objects.

They ranged from .569 (paraprofessional A-paraprofessional B) to .882 (paraprofessional B-paraprofessional A). In addition, there were significant tester effects at the .04 level; these were attributed entirely to one of the two sites. In other words, when individual subject differences were held constant, there were significant differences attributable to individual tester's frames of reference.

In the Fall 1971 Follow Through Evaluation (Emrick, 1972) the ITPA Verbal Expression Subtest was included in a supplementary battery given to kindergarten and entering first grade children in 17 projects. In general, the mean total response score of the four items was about six months below the normative data reported by Paraskevopoulos and Kirk. The measures of internal consistency (coefficient alpha) were high for the test and retest given two to three weeks later. The range of alphas was .739 to .887 for the test condition and .722 to .877 for the retest condition. The test-retest coefficient for the entire sample (n = 620) after a 2-3 week interval was .608. Correlations with a 29-item experimental version of the PSI were .566 (test) and .517 (retest); correlations with the Brown were .248 (test) and .314 (retest). Even though the reliability estimates of the test were acceptable, the test variance and inter-project variance were quite large, making the interpretation of the data difficult. Because of the large variances, overall low mean response rates and test administration problems, SRI concluded that the test in its present form not be used in future large scale evaluations.

Correlations with Other Tests

Using the middle class standardization sample, intercorrelations between the Verbal Expression Subtest and the other subtests of the ITPA ranged from .09 with the Auditory Sequential Subtest to .36 with the Manual Expression Subtest and .40 with the Auditory Association Subtest. For the same sample, correlations with the Stanford-Binet mental age (MA) and Stanford-Binet IQ score were low (.23 to .31) for the three age groups between 3 years-7 months to 6 years-1 month (Paraskevopoulos & Kirk, 1969).

In the Fall 1971 Follow Through supplemental testing study with kindergarten and entering first grade children (Emrick, 1972), correlations of the ITPA Verbal Expression Subtest with the 29-item version of the Preschool Inventory were .566 (test) and .517 (retest); correlations with the Brown were .248 (test) and .310 (retest).

Correlations of the subtest with other tests in the Fall 1971 HSPV battery are presented in Table 9 for the total HSPV sample. The largest correlation (.506) is with the 32-item Preschool Inventory (n = 1138). Other correlations over .40 are .487 (Peabody Picture Vocabulary Test), .422 (Eight-Block Sort Total Success) and .418 (Eight-Block Sort Reason Success).

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)	.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	.475 (1075)	.508 (1087)	.427 (1097)	.307 (1097)	.446 (1097)	.542 (1097)	.584 (1073)	.459 (1115)									
ETS ENUMERATION COUNTING	.492 (1075)	.504 (1097)	.422 (1087)	.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	.257 (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)	.066 (625)	.164 (608)	.032 (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)	.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)	.418 (1096)	.405 (1032)	.402 (1032)	.258 (1032)	.211 (1032)	.178 (1113)	.168 (1113)	.063 (573)	.320 (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.

Item Characteristics

Table 10 contains a frequency distribution of responses, both in terms of number of children and percent of children, for each of the ten categories for each of the four items of the ITPA Verbal Expression Subtest for the Fall 1971 HSPV sample. Table 11 presents the frequency data for the subtest for the total response to all four items for each category. For example, Table 11 shows that 83 children (6%) used the name category once during the entire subtest, while the largest number of children (43%) used the name category three times during the test. It can be seen from Tables 10 and 11 that many of the ten categories were used infrequently. Those categories which have a large number of no responses were number (92%), comparison (90%), other characteristics (77%), major parts (74%), shape (73%), and material (63%). (See Table 11). From Table 10 it appears that about one-quarter of the children gave one response to each item in two categories: color, and person, place, thing. The majority of children gave one name for at least three out of four objects (Tables 10 and 11). In addition, the majority of children gave one or more uses for each object; this category (use) was the one with the largest number of two or more responses (Table 11).

TABLE 10

ITPA VERBAL EXPRESSION SUBTEST
FREQUENCY DISTRIBUTION OF RESPONSES
(NUMBER OF CHILDREN AND PERCENT OF CHILDREN)
IN TEN CATEGORIES FOR EACH ITEM

Category	Item	Responses 0	1	2	3	4+
Name	Ball	98	1108	4	0	0
		8%	81%	0%	0%	0%
	Block	367	832	11	0	0
		30%	68%	0%	0%	0%
	Envelope	704	493	13	0	0
		58%	40%	1%	0%	0%
	Button	215	995	0	0	0
		17%	82%	0%	0%	0%
Color	Ball	878	332	0	0	0
		72%	27%	0%	0%	0%
	Block	712	497	1	0	0
		58%	41%	0%	0%	0%
	Envelope	190	300	0	0	0
		75%	24%	0%	0%	0%
	Button	902	308	0	0	0
		74%	25%	0%	0%	0%
Shape	Ball	1039	170	1	0	0
		85%	14%	0%	0%	0%
	Block	1037	171	2	0	0
		85%	14%	0%	0%	0%
	Envelope	1169	39	1	1	0
		96%	3%	0%	0%	0%
	Button	1043	163	4	0	0
		86%	13%	0%	0%	0%

(cont'd)

TABLE 10 (cont'd)

Category	Item	Responses 0	1	2	3	4+
Material	Ball	1073	137	0	0	0
		88%	11%	0%	0%	0%
	Block	1102	108	0	0	0
		91%	8%	0%	0%	0%
	Envelope	864	346	0	0	0
		71%	28%	0%	0%	0%
	Button	1168	42	0	0	0
		96%	3%	0%	0%	0%
Use	Ball	326	384	289	154	57
		26%	31%	23%	12%	4%
	Block	649	428	120	13	0
		53%	35%	9%	1%	0%
	Envelope	536	432	176	48	18
		44%	35%	14%	3%	1%
	Button	733	360	97	19	1
		60%	29%	8%	1%	0%
Major Parts	Ball	1147	63	0	0	0
		94%	5%	0%	0%	0%
	Block	1122	78	9	1	0
		92%	6%	0%	0%	0%
	Envelope	1150	55	3	2	0
		95%	4%	0%	0%	0%
	Button	978	200	26	6	0
		80%	16%	2%	0%	0%
Number	Ball	1205	5	0	0	0
		99%	0%	0%	0%	0%
	Block	1178	31	1	0	0
		97%	2%	0%	0%	0%
	Envelope	1198	10	2	0	0
		99%	0%	0%	0%	0%
	Button	1138	71	1	0	0
		94%	5%	0%	0%	0%

(cont'd)

TABLE 10 (cont'd)

Category	Item	Responses 0	1	2	3	4+
Other Characteristics	Ball	1125	82	3	0	0
		92%	6%	0%	0%	0%
	Block	1114	89	6	0	1
		92%	7%	0%	0%	0%
	Envelope	1130	74	6	0	0
		93%	6%	0%	0%	0%
	Button	1125	79	6	0	0
		92%	6%	0%	0%	0%
Comparison	Ball	1185	25	0	0	0
		97%	2%	0%	0%	0%
	Block	1172	37	1	0	0
		96%	3%	0%	0%	0%
	Envelope	1200	10	0	0	0
		99%	0%	0%	0%	0%
	Button	1140	63	7	0	0
		94%	5%	0%	0%	0%
Person, Place, Thing	Ball	1067	136	7	0	0
		88%	11%	0%	0%	0%
	Block	1020	182	7	1	0
		84%	15%	0%	0%	0%
	Envelope	368	518	239	65	20
		30%	42%	19%	5%	1%
	Button	495	656	57	1	1
		40%	54%	4%	0%	0%

N = 1210 children (902 PV children and 308 non-PV children)

TABLE 11

ITPA VERBAL EXPRESSION SUBTEST FREQUENCY DISTRIBUTION (NUMBER AND PER CENT OF CHILDREN) IN TEN CATEGORIES FOR ALL ITEMS[1]

Category	0	1	2	3	4	5	6	7	8	9	10	11	12
Name	32	83	236	521	324	14	0	0	0	0	0	0	0
	2%	6%	19%	43%	26%	1%	0%	0%	0%	0%	0%	0%	0%
Color	623	190	112	115	170	0	0	0	0	0	0	0	0
	51%	15%	9%	9%	14%	0%	0%	0%	0%	0%	0%	0%	0%
Shape	894	167	79	46	21	3	0	0	0	0	0	0	0
	73%	13%	6%	3%	1%	0%	0%	0%	0%	0%	0%	0%	0%
Material	765	329	60	40	16	0	0	0	0	0	0	0	0
	63%	27%	4%	3%	1%	0%	0%	0%	0%	0%	0%	0%	0%
Use	188	182	156	177	136	103	118	72	33	28	12	4	1
	15%	15%	12%	14%	11%	8%	9%	5%	2%	2%	0%	0%	0%
Major Parts	906	195	64	22	12	8	2	0	0	0	1	0	0
	74%	16%	5%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Number	1114	76	13	6	0	1	0	0	0	0	0	0	0
	92%	6%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Other Characteristics	938	201	53	12	2	3	0	0	1	0	0	0	0
	77%	16%	4%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Comparison	1094	91	18	4	3	0	0	0	0	0	0	0	0
	90%	7%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Person, Place, Thing	201	303	305	215	119	44	19	2	2	0	0	0	0
	16%	25%	25%	17%	9%	3%	1%	0%	0%	0%	0%	0%	0%

N = 1210 children (902 PV children and 308 non-PV children)

[1]There are four items in the total test: ball, block, envelope, button.

Remarks

Even though the ITPA is a successful diagnostic tool, it is not yet clear what the Verbal Expression Subtest means when it is used alone. This is substantiated by the SRI Fall 1971 Follow Through Study on the supplemental battery (Emrick, 1972).

Perhaps if more items were included and the scoring system extended this subtest would yield more valuable information. The present score gives only an estimate of the quantity of correct concepts used in expression. An extended scoring system which reflects the diversity of expression (i.e., number of different concepts used) or the grammatical content of expression (i.e., number of words used, etc.) might reveal more about the process of verbal expression. Another set of responses that could be scored and studied in order to give a better idea of the cognitive processes of young children is the set of wrong responses given by the children.

Finally, the child's willingness to discuss the attributes and functions of an object for the tester may be one of the most significant aspects of test performance. The child is asked to describe the test objects for no purpose other than to please himself and the tester. In addition, the subtest is difficult, since it asks the child to describe the object without context. An answer of "I can't say any more about it," which is often heard from young children, may mean nothing more than "I don't want to play," "I'm bored," or

"Why do you keep asking me for more?" A child might be able to elaborate at length with some feeling of purpose if he were given more context for the item or if there were more apparent relevance for the task. In addition, it is very possible that many children--especially the younger ones--do not fully understand what is expected of them. The practice item, showing the possible item attributes, is too long for many young children; furthermore, its relevance to responses on later items may be missed altogether by many children.

References

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

Emrick, J. A. Results of Fall 1971 Follow Through Supplemental Testing. Stanford Research Institute Report under Contract #OEC-0-8-522480-4633. Washington, D.C., Bureau of Elementary and Secondary Education, Office of Education, Department of Health, Education and Welfare, 1972.

Kirk, S. A., McCarthy, J. J., & Kirk, W. Examiner's manual: Illinois Test of Psycholinguistic Abilities (rev. ed.). Urbana, Illinois: University of Illinois Press, 1968.

Paraskevopoulos, J. N., & Kirk, S. A. The development and psychometric characterisitcs of the revised Illinois Test of Psycholinguistic Abilities. Urbana, Illinois: University of Illinois Press, 1969.

Motor Inhibition Test

Purpose

The Motor Inhibition Test (MI) was designed to measure a child's ability to inhibit movement when the task demands it. Besides measuring one aspect of the impulsivity dimension, the test measures psychomotor functioning in the areas of hand-eye coordination, large motor coordination and small motor coordination. At this point in the development of psychological theory it is unclear how either psychomotor functioning and/or the dimension of impulsivity is related to the cognitive performance of young children. Studies (Maccoby et al., 1965; Massari et al.,1969; Shipman et al., 1971; Ward, 1968) indicate that the ability to inhibit a response may be either a constituent of general intelligence or a style which contributes to intellectual performance.

Description

There are three inhibition of movement tasks in the test: the Draw a Line Slowly task, the Walk Slowly task, and the Truck task. The Draw a Line Slowly task consists of a picture of two telephone poles with a wire missing that the child is to draw in with the use of a ruler. The Walk Slowly task requires the child to walk down a six-foot walkway (five inches wide) that is marked off with tape on the floor. The Truck test requires the child to wind the

crank on a toy tow truck. The child is asked to perform all of these tasks twice: first, at his own speed, and second, in compliance with the instructions to do it as "slowly" as possible. The results of the test yield six scores which represent the time taken to do each task at a normal speed and the time taken to do each task when instructed to go "as slowly as possible."

All three parts of the MI were used in the first year of the HSPV Study. The difference between the slow and fast times for each task was computed and summed to give a final score for each child. A difference score was used in the analysis by Stanford Research Institute (1971, p. 55) "to compensate for the fact that a child may get a high 'slow' score by being slow -- not by inhibiting his response." In the second year of the HSPV Study all three tasks of the test were given. In the third year, only the Truck task was given. The toy truck data for the second and third years is more reliable than that of the first year since administration of the task was more standardized. The trucks used in the last two years were easier to wind up. In the analyses of the second and third year data, log transformations on the "slow" times of each subtest are being used as scores. In analyzing the 1971-72 results, a log transformation of the sum of the slow walk and slow draw times is also being used.

Development of Instrument

The Motor Inhibition Test, devised by Hagen and Degerman, was first used by Maccoby and her associates (1965)

in a study with 42 middle class preschoolers. A total score on inhibition of movement was obtained by adding "slow" scores of the three subtests. Since the girls were more consistent in performance from task to task, their total scores were more reliable. Test-retest reliabilities on small samples for each subtest were .77 (Draw a Line Slowly), .81 (Walk Slowly), and .89 (Truck). Inter-correlations among the subtests were as follows: Draw a Line Slowly vs. Walk Slowly (boys, .51; girls, .69); Draw a Line Slowly vs. Truck Test (boys, .39; girls, .53); and Walk Slowly vs. Truck Test (boys, .42; girls, .71). Finally, scores of the MI were positively correlated with Stanford Binet IQ scores ($r = .44$ for the sexes combined; $r = .38$ for boys only; $r = .50$ for girls only). [The Stanford-Binet scores for this sample ranged from 95 to 154 (mean = 135).] In addition, there was a positive correlation tendency between the MI score and the Children's Embedded Pictures Test ($r = .23$ for boys only; $r = .34$ for girls only), although this correlation was not quite significant.

Related Studies

The finding of the Maccoby et al. study that more intelligent children are more able to inhibit movement when engated in a task that requires it has been replicated in several other studies (Massari et al., 1969; Loo & Wenar, 1971; Ward, 1968). Ward found that the MI score (as defined in the Maccoby et al. study) was positively correlated

(.34) with the WISC IQ score for seven and eight-year-old boys in a summer recreational program. Loo and Wenar (1971) found that a combined raw score of the Draw-a- Line task and the Walk-a-Line task correlated significantly (.38, $p < .02$) with the Primary Mental Abilities Test IQ Score for 40 upper middle class children in kindergarten in Columbus, Ohio. In a study with 33 white disadvantaged five-year-olds, Massari, Hayweiser and Meyer (1969) showed that the Maccoby et al. findings were true for a sample with lower intelligence scores (mean IQ = 90). Massari et al. found that the Stanford-Binet IQ correlated positively with the Draw-a-Line Slowly task (pre-test, .45; posttest, .56) and with the Walk Slowly task (pretest, .44; posttest, .60). Further findings from this study show that there is no correlation between IQ and the ability to do a movement "as fast as possible" and that impulse control of motor activity is independent of ability to understand instructions.

One study has shown contradictory findings to the original Maccoby et al., study. Mumbauer and Miller (1970) found no significant correlation between the Maccoby MI score and the Stanford-Binet IQ for 32 advantaged and 32 disadvantaged five-and-one-half-year-olds in Nashville, Tennessee. There was a trend towards a significant correlation at the .05 level for the advantaged group, which was more like Maccoby et al.'s sample.

The Draw-a-Line Slowly task is used by Banta (1970) as a measure of impulse control in the Cincinnati Autonomy Test

Battery (CATB). The impulse control subtest of the CATB asks the child to draw a line (freehand) between two "X's" eight inches apart. After the child does it once at a normal speed, he is asked to draw the line three more times each time doing it "slower" than the last. A score is obtained by averaging the rate scores (the length of the line divided by the time in hundredths of a second) for each of the three times. On a sample of approximately 80, Banta found that the impulse control score correlated positively with the Stanford-Binet (.24), reflectivity scores (.37), intentional learning (.31), kindergarten prognosis (.31), persistence (.28), resistance to distraction (.27), task competence ratings (.25), and innovative behavior (.23). From his study of the 14 scores of the CATB, Banta concludes that impulse control is an "important developmental variable affecting a variety of behaviors relevant to problem-solving ability (1970, p. 475)." Reliability coefficients for the impulse control subtest of the CATB are as follows: test-retest reliabilities are .41 (n = 33 after one month) and .43 (n = 33 after two months); internal consistency reliability coefficients range from .66 to .69 (n = 32) and from .47 to .80 (n = 74); and inter-rater reliability is .90 (n = 30).

ETS Head Start Longitudinal Study

The Motor Inhibition Test is also being used in the ETS Head Start Longitudinal Study (Shipman et al., 1971; Shipman,

1972). Even under the "slow" direction, the children in the samples of the first two years did the tasks relatively quickly. The mean number of seconds was 5.9 in Year 1 and 7.8 in Year 2 for the Drawing task, 6.4 for Year 1 and 7.7 for Year 2 for the Walking task, and 50.0 in Year 1 for the Truck task. All the three-and-one-half-year-olds followed the instructions and performed the task the second time more slowly than they had the first time. Increases in mean times under slow directions over mean practice times were 23% for the Truck task, 36% for the Walking task, and 54% for the Drawing task in Year 1.

Slow scores, transformed by log (x + 1), are reported in the latest ETS technical report (Shipman, 1972) for Year 1 and Year 2 Head Start samples. Mean scores for Walking were .87 (S.D. = .21) in Year 1 and .94 (S.D. = .22) in Year 2; mean scores for Drawing were .84 (S.D. + .29) in Year 1 and .95 (S.D. = .30) in Year 2. The mean score for the Truck subtest in Year 1 was 1.71 (S.D. = .18); this task was not used in Year 2.

Correlations between Walking Slowly and Drawing Slowly were moderately high (.50 in Year 1 and .53 in Year 2), while the correlations of the Truck subtest with each of the other two subtests were low (around .25) in Year 1. Shipman (1972) hypothesizes that the lower correlations for the Truck subtest many have been due to a combination of greater demands made by this subtest on a child's coordination and a tester's skill.

Because of these correlations and the distribution of scores, Shipman concludes that the best MI score is the average of the standardized (and log transformed) "slow" times from the Walking and Drawing tasks. The Truck task was eliminated after Year 1.

Using only the Head Start longitudinal sample, the composite score of Walking and Drawing scores correlated positively with the 64-item Preschool Inventory (.36 in Year 1, .37 in Year 2) and with the Peabody Picture Vocabulary Test, Form A (.34 in Year 1, .36 in Year 2). There was no relation to other measures in the impulsivity domain; the correlations with the Matching Familiar Figures Test, the Preschool Embedded Figures Test and Sigel's Object Categorization Test were less than .15.

Huron-SRI Reliability Study

In the fall of 1971 the Truck task of the Motor Inhibition Test was included in a test-retest/inter-tester reliability study conducted by Huron Institute. Details of this study using 2 sites of the HSPV sample are reported in Appendix A. In general, the test-retest reliability coefficients for the "slow" times (log transformed) after two weeks for approximately 20 children ranged from .302 (expert-paraprofessional B) to .710 (expert-expert). There were significant tester effects at the .001 level of significance. This means that there were significant differences attributable to individual tester's frames of reference (i.e. style of administering the test)

when individual subject differences were held constant.

Head Start Planned Variation Sample Characteristics

Means and standard deviations for the slow, log transformed scores of the three MI subtests are listed in Tables 1 - 3 for ten three-month age intervals (from 42-44 months to 69-71 months) for the Fall 1970 HSPV sample. For the total sample (n=1086), the mean scores for the three subtests were 4.344 (S.D.=.730) for the Drawing subtest, 3.863 (S.D. = .917) for the Walking subtest, and 6.133 (S.D. = .347) for the Truck subtest. Table 4 includes the means and standard deviations for each of the three subtests for selected Fall 1970 subsamples: males, females, black children, white children children with previous preschool experience, and children with no previous preschool experience.[1]

The distribution of Truck slow log scores for 15 three-month age intervals (from 36-38 months to 78-80 months) are listed in Tables 5 - 12 for several Fall 1971 HSPV samples: total (Table 5), males (Table 6), females (Table 7), children with previous preschool experience (Table 8), children with no previous preschool (Table 9), white children (Table 10), black children (Table 11) and Mexican-American children (Table 12). The mean score for the total sample was 3.855 (S. D. = .350, n = 634). The mean scores for the other subsamples were very similar.

---

[1] See Smith (1973) for an extensive analysis and discussion of MI characteristics as they relate to the various HSPV models.

TABLE 1

DISTRIBUTION OF MI DRAW SUBTEST SLOW LOG SCORES FOR ALL CHILDREN IN THE FALL 1970 SAMPLE[1]

Age (Months)	N	Mean Score	S.D.
42-44	1	2.300	----
45-47	19	4.250	.7148
48-50	74	4.188	.6731
51-53	137	4.086	.7201
54-56	156	4.197	.6016
57-59	130	4.184	.7638
60-62	154	4.449	.7101
63-65	143	4.541	.7156
66-68	118	4.579	.7095
69-71	124	4.572	.6962
TOTAL	1086	4.344	.7205

[1]Includes all children not in Level I sites, Oraibi, or Fresno; who had adequate age information.

TABLE 2

DISTRIBUTION OF MI WALK SUBTEST SLOW LOG SCORES FOR ALL CHILDREN IN THE FALL 1970 SAMPLE[1]

Age (Months)	N	Mean Score	S.D.
42-44	1	4.060	-----
45-47	19	4.139	.6239
48-50	74	4.203	.7507
51-53	137	3.774	1.339
54-56	156	3.763	1.186
57-59	160	3.733	1.223
60-62	154	3.855	.6705
63-65	143	3.928	.4959
66-68	118	3.887	.3975
69-71	124	3.922	.4072
TOTAL	1086	3.863	.9172

[1]Includes all children not in Level I sites, Oraibi, or Fresno; who had adequate age information.

TABLE 3

DISTRIBUTION OF MI TRUCK SUBTEST SLOW LOG SCORES FOR ALL CHILDREN IN THE FALL 1970 SAMPLE[1]

Age (Months)	N	Mean Score	S.D.
42-44	1	5.500	------
45-47	19	6.173	.2854
48-50	74	6.165	.2763
51-53	137	6.176	.3430
54-56	156	6.113	.3063
57-59	160	6.129	.2968
60-62	154	6.102	.4003
63-65	143	6.161	.3819
66-68	118	6.159	.3229
69-71	124	6.076	.3890
TOTAL	1086	6.133	.3465

[1]Includes all children not in Level I sites, Oraibi, or Fresno; who had adequate age information.

TABLE 4

MEANS AND STANDARD DEVIATIONS FOR MI SUBTESTS SLOW SCORES FOR SELECTED FALL 1970 HSPV SAMPLES[1]

	Draw Subtest		Walk Subtest		Truck Subtest		
	Mean	S.D.	Mean	S.D.	Mean	S.D.	n
Males	4.354	.7314	3.862	.9284	6.151	.3253	550
Females	4.333	.7274	3.865	.9057	6.114	.3661	536
Blacks	4.273	.7451	4.139	.6710	6.057	.3553	495
Whites	4.404	.7111	3.644	1.1272	6.193	.3332	437
Children w/previous preschool	4.523	.7479	3.824	.5605	6.090	.4057	168
Children with no previous preschool	4.313	.7273	3.878	.9781	6.139	.3342	882

[1]Includes all children, not in Level I sites, Oraibi, or Fresno, who had adequate age information.

TABLE 5

DISTRIBUTION OF MI TOW TRUCK SLOW LOG SCORES FOR ALL CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score	S.D.
36-38	---	-----	-----
39-41	1	4.170	-----
42-44	6	3.837	0.361
45-47	18	3.708	0.592
48-50	51	3.855	0.387
51-53	93	3.845	0.279
54-56	106	3.877	0.318
57-59	105	3.847	0.385
60-62	81	3.820	0.311
63-65	52	3.867	0.363
66-68	50	3.852	0.320
69-71	44	3.878	0.342
72-74	26	3.981	0.360
75-77	1	3.610	-----
78-80	---	-----	-----
TOTAL	634	3.855	0.350

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

TABLE 6

DISTRIBUTION OF MI TOW TRUCK SLOW LOG SCORES FOR MALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score	S.D.
36-38	---	-----	-----
39-41	1	4.170	-----
42-44	1	3.730	-----
45-47	12	3.869	0.331
48-50	27	3.811	0.417
51-53	48	3.803	0.306
54-56	59	3.842	0.336
57-59	50	3.754	0.426
60-62	34	3.798	0.317
63-65	30	3.879	0.348
66-68	27	3.780	0.307
69-71	23	3.875	0.343
72-74	17	3.972	0.386
75-77	1	3.610	-----
78-80	---	-----	-----
TOTAL	330	3.826	0.357

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

TABLE 7

DISTRIBUTION OF MI TOW TRUCK SLOW LOG SCORES FOR FEMALES IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score	S.D.
36-38	---	-----	-----
39-41	---	-----	-----
42-44	5	3.858	0.392
45-47	6	3.385	0.823
48-50	24	3.905	0.343
51-53	45	3.893	0.238
54-56	47	3.921	0.288
57-59	55	3.931	0.320
60-62	47	3.835	0.305
63-65	22	3.850	0.381
66-68	23	3.913	0.324
69-71	21	3.880	0.340
72-74	9	3.998	0.304
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	304	3.886	0.339

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

TABLE 8

DISTRIBUTION OF MI TOW TRUCK SLOW LOG SCORES

FOR ALL CHILDREN WITH PREVIOUS PRESCHOOL EXPERIENCE

IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score	S.D.
36-38	---	-----	-----
39-41	---	-----	-----
42-44	---	-----	-----
45-47	5	3.320	0.876
48-50	12	3.780	0.403
51-53	18	3.802	0.267
54-56	13	3.906	0.261
57-59	14	3.814	0.308
60-62	27	3.834	0.281
63-65	20	3.906	0.414
66-68	15	4.017	0.349
69-71	15	3.883	0.388
72-74	7	4.024	0.345
75-77	1	3.610	-----
78-80	---	-----	-----
TOTAL	147	3.854	0.388

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

TABLE 9

DISTRIBUTION OF MI TOW TRUCK SLOW LOG SCORES FOR ALL CHILDREN WITH NO PREVIOUS PRESCHOOL EXPERIENCE IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score	S.D.
36-38	---	-----	-----
39-41	1	4.170	-----
42-44	6	3.837	0.361
45-47	13	3.857	0.333
48-50	38	3.874	0.383
51-53	71	3.875	0.270
54-56	92	3.873	0.327
57-59	88	3.841	0.392
60-62	27	3.834	0.281
63-65	20	3.906	0.414
66-68	15	4.017	0.349
69-71	15	3.883	0.388
72-74	7	4.024	0.345
75-77	1	3.610	-----
78-80	---	-----	-----
TOTAL	473	3.857	0.338

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

TABLE 10

DISTRIBUTION OF MI TOW TRUCK SLOW LOG SCORES FOR WHITE CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score	S.D.
36-38	---	-----	-----
39-41	1	4.170	-----
42-44	1	4.200	-----
45-47	6	3.957	0.287
48-50	17	4.088	0.323
51-53	42	3.901	0.232
54-56	39	4.061	0.263
57-59	42	3.963	0.318
60-62	39	3.835	0.306
63-65	21	4.003	0.323
66-68	24	3.908	0.272
69-71	22	4.051	0.313
72-74	18	4.084	0.260
75-77	1	3.610	-----
78-80	---	-----	-----
TOTAL	273	3.970	0.311

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

TABLE 11

DISTRIBUTION OF MI TOW TRUCK SLOW LOG SCORES FOR BLACK CHILDREN IN THE FALL 1971 HSPV SAMPLE[1]

Age (Months)	N	Mean Score	S.D.
36-38	---	-----	-----
39-41	---	-----	-----
42-44	5	3.764	0.353
45-47	12	3.583	0.663
48-50	16	3.760	0.344
51-53	38	3.803	0.232
54-56	52	3.711	0.289
57-59	50	3.824	0.329
60-62	30	3.797	0.272
63-65	21	3.811	0.374
66-68	15	3.931	0.359
69-71	12	3.730	0.298
72-74	7	3.843	0.389
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	268	3.779	0.342

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

TABLE 12

DISTRIBUTION OF MI TOW TRUCK SLOW LOG 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	7	3.667	0.437
51-53	12	3.782	0.277
54-56	13	3.952	0.218
57-59	10	3.480	0.622
60-62	9	3.933	0.336
63-65	9	3.662	0.307
66-68	10	3.596	0.241
69-71	10	3.672	0.239
72-74	---	-----	-----
75-77	---	-----	-----
78-80	---	-----	-----
TOTAL	80	3.728	0.383

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

HSPV Correlations with Other Tests

Correlations of the individual MI Subtest (using log transformations of the "slow" times) with the CPSCS, the 64-item PSI, the NYU Booklets 3D and 4A, the Eight-Block Sort Success scores, and the Stanford-Binet IQ and MA were computed by Huron Institute for the total Fall 1970 sample (Table 13) and several subsamples (blacks, whites, young, old, previous preschool experience, and no previous preschool experience). In every case, the correlations were low. For the total sample, the correlations with the 64-item PSI were .279 for the Walking subtest, .356 for the Drawing subtest, and .165 for the Truck subtest; the correlations with the CPSCS were all close to zero; the correlations with the Eight-Block Sort scores were between .067 and .212; the correlations with the NYU Booklet 3D were .275 for the Walking subtest; .298 for the Drawing subtest, and .136 for the Truck subtest; and correlations with the NYU Booklet 4A were .142 for the Walking subtest, .142 for the Drawing subtest, and .106 for the Truck subtest. Unlike other studies, the correlations with IQ were low: .152 with the Walking subtest, 229 with the Drawing subtest, and .120 with the Truck subtest. Correlations with MA for the total sample were .032 for the Truck subtest, .259 for the Walking subtest, and .436 for the Drawing subtest.

Correlations of the Truck subtest with the two other subtests were generally in the .20's. Correlations of the Truck subtest with the Drawing subtest ranged from .151 for

TABLE 13

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	.054 (1024)	.275 (1073)	.142 (1072)	.279 (1074)								
MI DRAW	.078 (1026)	.298 (1076)	.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)	.156 (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)	.306 (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).

blacks (n = 490) to .249 for whites (n = 436) (.223 for the total sample); correlations of the Truck subtest with the Walking subtest ranged from .202 for females (n = 518) to .308 for males (n = 540) (.255 for the total sample). Correlations of the Walking and Drawing subtests were higher (.459 for the total sample); they ranged from .377 for young children (less than 58 months) (n = 408) to .557 for children with previous preschool experience (n = 203). These correlations between the three MI subtests are very similar to those reported by Shipman (1972) for the Head Start Longitudinal sample.

Correlations of a combined score from the Walking and Drawing subtests (log transformation of the sum of the slow scores) with the other tests in Table 13 were very close to the higher of the two individual subtests' correlations with the other tests.

The correlations of the Truck subtest with the other tests in the Fall 1971 HSPV battery (PPVT, WRAT subtests, 32-item PSI, ITPA Verbal Expression Subtest, ETS Enumeration Test, Brown Self-Concept Test, and Eight-Block Sort Task) are reported in Table 14. All of the correlations were very low; the highest was .174 (with the Peabody Picture Vocabulary Test).

Remarks

Motor Inhibition scores have been reported to correlate with a large number of developmental factors such as impulse control, psychomotor functioning, coordination, impulsivity,

TABLE 14 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)	.503 (2860)	.589 (2860)											
ITPA-VERBAL EXPRESSION	.487 (1147)	.339 (1172)	.371 (1172)	.276 (1172)	.341 (1172)	.388 (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)	.356 (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)	.175 (2753)	.270 (2753)	.323 (2689)	.261 (1145)	.228 (1073)	.271 (1073)	.160 (1073)	.054 (1073)					
BROWN-ADJUSTED[3]	.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	.174 (607)	.061 (625)	.048 (625)	.083 (625)	.121 (625)	.006 (625)	.164 (606)	.032 (637)	.136 (597)	.135 (597)	.047 (597)	.107 (597)	.118 (610)	.10: (610)			
EIGHT-BLOCK PLACEMENT	.304 (1113)	.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)	.428 (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.

ability to follow instructions, and performance on intelligence tests. It may be that MI changes with age are evidence for the presence of both a physical and psychological developmental factor. More studies and analyses are needed to determine which of these developmental factors is most measured by the MI. Future investigations are also needed to clarify the relationship between intelligence and motor inhibition.

The interrelationship among the three subtests is puzzling. Because of test administration problems and the low correlations with the other MI subtests and other tests in the HSPV batteries, the Truck subtest yields less valid and valuable information than the other two subtests. It is therefore recommended, that the Truck subtest be dropped from future large-scale evaluations.

References

Banta, T. J. Tests for the evaluation of early childhood education: The Cincinnati Automomy Test Battery (CATB). In J. Hellmuth (Ed.), Cognitive studies. Vol. 1. New York: Brunner/Mazel, 1970.

Loo, C. & Wenar C. Activity Level and Motor Inhibition: their relation to intelligence - test performance in normal children. Child Development, 1971, 42, 967-71.

Maccoby, E. E., Dowley, E. M., Hagen, J. W., & Degerman, R. Activity level and intellectual functioning in normal preschool children. Child Development, 1965, 36, 761-770.

Massari, D., Hayweiser, L., & Meyer, W. Activity level and functioning in deprived preschool children. Developmental Psychology, 1969, 1, 286-90.

Mumbauer, C. C., & Miller, J. O. Socioeconomic background and cognitive functioning in preschool children. Child Development, 1970, 41, 471-80

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. J.: Educational Testing Service, 1971.

Shipman, V. (Ed.) Disadvantaged children and their first school experiences: Technical report series. Princeton, N.J. Educational Testing Service, 1972.

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.

Stanford Research Institute. Implementation of planned variation in Head Start: Preliminary evaluations of planned variation in Head Start according to Follow Through approaches (1969-70). Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, 1971.

Ward, W. C. Reflection-impulsivity in kindergarten children. Child Development, 1968, 39, 737-754.