DOCUMENT RESUME

ED 081 813 TM 003 174

AUTHOR Shipman, Virginia C., Ed.
TITLE Disadvantaged Children and Their First School Experiences. ETS-Head Start Longtitudinal Study. Technical Report Series.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY Office of Child Development (DHEW), Washington, D. C. Bureau of Head Start and Child Service Programs.
REPORT NO ETS-PR-72-27
PUB DATE Dec 72
NOTE 433p.; Grant No. H-8256

EDRS PRICE MF-$0.65 HC-$16.45
DESCRIPTORS Data Analysis; Data Collection; *Disadvantaged Youth; *Longitudinal Studies; *Measurement Instruments; *Preschool Tests; Tables (Data); *Technical Reports; Testing; Tests
IDENTIFIERS *Project Head Start

ABSTRACT

This volume consists of a series of technical reports on each of the individual test measures administered in both Years 1 and 2 of the Longitudinal Study of Disadvantaged Children and Their First School Experiences. Each report includes: (1) relevant background information concerning the task's purpose and related research findings; (2) general task description and administration procedures, including training cautions where appropriate; (3) scoring and, where necessary, coding procedures; (4) score properties as revealed in various internal analyses; (5) sample performance, including descriptive information by three-month age groups and sex for the total sample tested in Year 1 and in Year 2, and comparative analyses for the longitudinal sample (those children tested in both years of the study); (6) information bearing on construct validity, i.e., the task's convergent and, where possible, discriminant validity, the extent to which the measure has the same meaning in Year 1 and Year 2, the similarity of meaning across groups (i.e., SES); and (7) discussion of methodological aspects affecting valid assessment (e.g., reliability, item characteristics, irrelevant measurement difficulties and appropriateness for field use). (For related documents, see TM 003 175-198.) (Author/DB)

PR 72-27

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman

December 1972

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

YOUNG CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

A LONGITUDINAL STUDY

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Acknowledgments

The writer wishes to express deep appreciation to Walter Emmerich, Robert Linn and William Ward for reviewing this series of reports, providing thoughtful and constructive criticism throughout, and for their support and counsel throughout the study.

Special acknowledgment must be given to the following research staff who contributed substantially to particular reports: Diran Dermen, Walter Emmerich, Karla Goldman, Lynn Gilbert, David Lindstrom, Judith Meissner and William Ward.

I am grateful, too, to Susan Simosko for her assistance in preparing the introductory section and for coordinating so well the various activities necessary for finalizing the production of the entire report.

Particular thanks go to John Barone who coordinated data analysis activities and to Norma Hvasta, Judith Ohls and Emily White who programmed the analyses reported here; thanks, too, are due the women who painstakingly coded all the data under the able supervision of Joan Tyson.

Also, the continuing encouragement, understanding, and constructive criticism provided by the Office of Child Development, particularly by Dr. Lois-ellin Datta, Mr. Richard Orton, Dr. Thelma Zener and Dr. Edward Zigler, and by the Head Start Research Advisory Committee, Dr. Urie Bronfenbrenner, Dr. Boyd McCandless and Dr. Alfred Yankauer, have contributed greatly to the form and substance of the study.

Appreciation is also expressed to Thelma Benton and Judith Ferguson for their care and speed in typing the several drafts and final copy. For assistance with proofing and editing, thanks go to Susan Simosko, Joanne Goldenbaum, Lynn Gilbert, David Lindstrom and Judith Meissner.

Special thanks are due the former Local Coordinators and Technical Directors: Mrs. Lida Campbell and Dr. Raymond Phillips, Lee County, Alabama; Mrs. Verna Shepherd and Dr. Jack Crawford, Portland, Oregon; Mr. Ronald Greeley, Mr. Bobby Westbrooks and Dr. Arthur Littleton, St. Louis, Missouri; and Mr. Conrad McLean and Mrs. Sharon Creech, Trenton, New Jersey. All contributed knowledge of their communities and varied technical and administrative skills that were invaluable for organizing and coordinating testing activities in the field. I owe gratitude as well to the many testers, test center and playroom supervisors, and drivers, without whose efforts data could not have been collected. Their hard work, enthusiasm, and patience were a continuing source of encouragement to those of us who knew the frustrations they experienced working within a complex organizational structure that was not always geared to their needs. In addition to the valuable program and field coordination provided by Joseph Boyd and Samuel Barnett, a large debt must be acknowledged to Anne Bussis, Rosalea Courtney, Karla Goldman, Jean Orost, Patricia Warren and Phyllis Ward who assisted me in tester training. Gratitude must also be expressed for the monitoring and field consultation provided by ETS Regional Office staff: Junius Davis, Roderick Ironside, Chandra Mehrotra, Daniel Norton, Santelia Knight, Robert Lambert and George Temp.

Deepest gratitude, however, goes to the children and their families who participated in the study. Together we hope to contribute to a better understanding of the young child's development.

Virginia C. Shipman

Princeton, New Jersey
December 18, 1972

Introduction

Virginia C. Shipman

This volume consists of a series of technical reports on each of the individual test measures administered in both Years 1 and 2 of the Longitudinal Study of Disadvantaged Children and Their First School Experiences. Each report includes 1) relevant background information concerning the task's purpose and related research findings; 2) general task description and administration procedures, including training cautions where appropriate; 3) scoring and, where necessary, coding procedures; 4) score properties as revealed in various internal analyses; 5) sample performance, including descriptive information by three-month age groups and sex for the total sample tested in Year 1 and in Year 2, and comparative analyses for the longitudinal sample (those children tested in both years of the study); 6) information bearing on construct validity; that is, the task's convergent and, where possible, discriminant validity, the extent to which the measure has the same meaning in Year 1 and Year 2, the similarity of meaning across groups (i.e., SES), and 7) discussion of methodological aspects affecting valid assessment (e.g., reliability, item characteristics, irrelevant measurement difficulties and appropriateness for field use).

The purpose of this introductory section is to provide the necessary context for interpreting these separate reports. It presents an overview of the project and a brief description of the study sample and data collection and analysis procedures relevant to these reports. The individual technical reports offer particular information insofar as sample composition, administration and results are unique to the measure in a given year or site. A more detailed description of the study, including theoretical rationale and measurement considerations, appears in Project Report 68-4 (ETS, 1968).

Some of the critical design and logistic problems are also elaborated in the ETS brochure (RM-69-6) "Untangling the Tangled Web of Education" and ETS Project Report 69-12.

Given the breadth of study questions being addressed, it should be noted that many instruments that were given during this same period are not included in this report--i.e., parent interviews, mother-child structured interaction tasks and, for those children attending a preschool program in Year 2, observations of teacher-child, peer and child-material interactions, teacher and Center Director questionnaires, and teacher ratings. Also, those tests administered for the first time in Year 2 are not described.

Overview

The ETS-Head Start Longitudinal Study was originally conceived as an assessment of early education as it affects the cognitive, personal and social development of the young child. As an inherent part of this assessment, the study was designed to describe and, if possible, evaluate the environmental and background variables that moderate and influence these associations.

The age range chosen for study was the developmental span of approximately four through eight years of age--or from two years prior to entrance into first grade through completion of third grade. This period is thought to be particularly important because it is a time during which many abilities consolidate and the child makes the social transition from familiar home surroundings to the world of school, peers and unfamiliar adults. The first data were collected during the spring and summer of 1969 on over 1800 children, the majority falling between the ages of three years nine months (3-9) and four years eight months (4-8). All were scheduled to be enrolled in first grade in the fall of 1971. Data collection for many of these children and their families, communities and schools is planned to continue through spring of 1974.

The study population was identified and information was gathered prior to the time when the target children were eligible to enter a Head Start program. Decisions about sending or not sending children to Head Start or kindergarten were therefore made without our intervention by the parents involved, after the study was underway. Thus, given lack of control in assigning children to "treatments" or programs, prior information (baseline data) is used to assess the comparability of children receiving different treatments. By following the same children over a number of years, the comparability of beginning grade school experience for both Head Start and non-Head Start youngsters--e.g. the degree to which primary grade curricula are congruent with and capitalize on the child's preschool experience--can also be assessed.

A longitudinal design also affords the opportunity to study variables which might be expected to have long-term rather than short-term effects. Such a strategy has potential value for educational and social planning, theories of child development and development of techniques for assessing young children and their environments.

Sample Characteristics

Four regionally distinct communities were selected which 1) had a sufficient number of children in school and in the Head Start program, 2) appeared feasible for longitudinal study given expressed community and school cooperation and expected mobility rates and 3) offered variation in preschool and primary grade experiences. The study sites chosen were ee County, Alabama; Portland, Oregon; St. Louis, Missouri; and Trenton, New Jersey. Within these communities, elementary school districts with a substantial proportion of the population eligible for Head Start were selected for participation. For the most part, schools in target districts are located near Head Start centers. It is in these school districts that the sample is expected to be enrolled by third grade. in the fall of 1973. In each school district an attempt was made to include all children of approximately 3 1/2- to 4 1/2-years of age in the initial testing and data collection of 1969, although some children were excluded from the sample (e.g., children from families whose primary language was not English, and those with severe physical handicaps).

In 1969 mothers were interviewed and children tested prior to their enrollment in Head Start or any other preschool program; in 1969-70 these measures were repeated and extensive observation of those children attending preschool programs in Portland, St. Louis and Trenton took place. In Lee County, where Head Start is a kindergarten level program, a brief version of the test battery was administered. The following is an overview of the salient demographic characteristics of the initial four-site sample (for a more complete description of this population the reader is referred to Project Report 71-19 [Shipman, 1971]):

1. At least partial data were obtained for a total of 1875 children. However, the number of subjects at each site varies, with Lee County and Portland together constituting 60% of the sample.
2. The sample is 62% black.
3. Boys make up 53% of the sample. For the four sites they compose 54.5% of the black sample and 50.5% of the white sample.

4. For the three sites in which children had the opportunity to attend Head Start in Year 2 of the study (1969-70), 37.2% of the sample attended Head Start, 11% attended other preschool programs and 51.8% had no known attendance in Head Start or other preschool programs. In Lee County, where Head Start is a kindergarten level program, 41.7% of the initial sample attended Head Start, 19.1% attended other preschool programs and 39.3% had no known attendance in Head Start or other preschool programs.

5. Substantially more blacks than whites attended Head Start. While this varies by site, in the total sample, only 5.1% of the children who attended Head Start are white.

6. The parents of the white children generally have had more years of formal education than the black parents, except in St. Louis where the reverse is true.

7. Although the fathers of both blacks and whites tend to be in blue-collar positions, a disproportionately large number of blacks are so classified.

8. Educational and occupational data were obtained for substantially fewer fathers than mothers. Moreover, the percentage difference between the number of fathers and the number of mothers for whom data were obtained was greater for blacks than for whites, and for children who attended Head Start than for others.

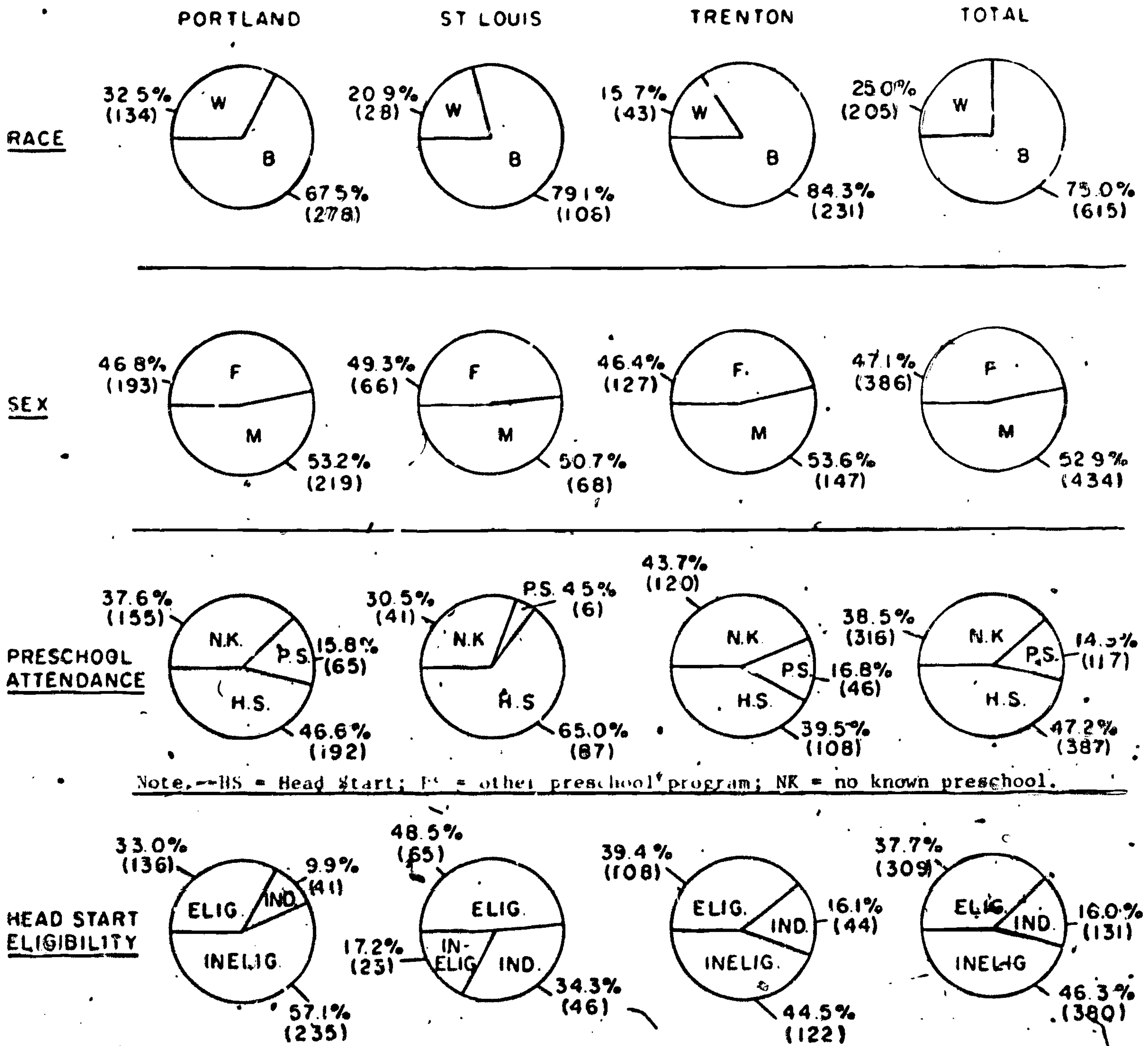
The Year 2 sample included children from four sites: Trenton, Lee County, Portland, and St. Louis. As noted earlier, during Year 2 data-gathering procedures in Lee County were limited to a fraction of the test battery due primarily to limited resources and to the fact that most of the children in Lee County were not enrolled in preschool programs until the third year of the study when Head Start was available. Data from this site have, therefore, been included only for tasks where appropriate.

The comparative analyses that follow are confined to the longitudinal sample, that is, to those children who fulfilled initial qualifications for inclusion in the study and about whom at least one unit of test information had been collected in both 1969 and 1970. Thus, this population is a subsample of the initial four-site sample described above and reported in Project Report 71-19. The three-site urban longitudinal sample consists of 820 children (427 children were seen in Lee County on the abridged test battery, 418 of whom were longitudinal subjects). In some cases data for these children are incomplete. As can be seen in Figure 1, there are some fairly substantial differences in sample size by . The Portland children constitute 50.2% of the urban longitudinal sample, whereas the Trenton and St. Louis sites comprise only 33.4% and 16.4% of this sample, respectively. In contrast to the initial three-site urban sample, Portland and Trenton represent higher percentages of the longitudinal sample (50.2 vs. 45.7 and 33.4 vs. 33, respectively), whereas the percentage of St. Louis subjects is less (16.4 vs. 21.3), due to the much greater sample attrition at that site. Some of the major demographic characteristics of this three-site urban longitudinal sample are presented below, indicating disproportionalities of single and multiple classifications. (See Table 1 for the number of subjects cross-classified by sex, race, preschool attendance and Head Start eligibility.) It should be recognized that the investigators regard these characteristics as demographic variables only and discourage simplistic or stereotypic psychological interpretation of these biological and cultural statuses.

Figure 1

Children in Each Site Classified by Race, Sex, Preschool Attendance and Head Start Eligibility

(Year 1-2 Longitudinal Urban Sample)



Note.--HS = Head Start; PS = other preschool program; NK = no known preschool.

Note.--IND = indeterminate economic eligibility for Head Start.

Table 1

3-Site Urban Longitudinal Sample, Classified by Sex Race, Preschool Attendance and Head Start Eligibility (Year 1-2)

		Boys		Girls		
		White	Black	White	Black	Total
Head Start	Elig.	13	98	8	78	197
	Inelig.	14	55	11	47	127
	Indet.	5	26	6	26	63
No Known Preschool Program	Elig.	8	47	12	31	98
	Inelig.	32	45	44	43	164
	Indet.	7	20	8	19	54
Other Preschool Program	Elig.	1	8	0	5	14
	Inelig.	19	29	16	25	89
	Indet.	0	7	1	6	14
	Total	99	335.	106	280	820

There are, as indicated, a number of disproportionalities in the various classifications of importance: 1) the Portland sample constitutes 50% of the urban longitudinal population, 2) there are three times as many blacks as whites in the sample, 3) 53% of the sample are boys, 4) 61.5% of the children attended preschool programs, 5) a substantially greater percentage of Head Start children are black, 6) approximately two-thirds of the white population (vs. approximately half the black families) are economically ineligible for Head Start, 7) a substantial percent (54.5) of Head Start children are boys, 8) the children in St. Louis are two months older than the Trenton and Portland children, and 9) the mothers of children in Head Start have over two years less formal education than other mothers.

Differences in the number of children in various classifications is a necessary part of the type of design used in the study. It would inevitably be impossible in such a study to identify and select equal or proportional cell sizes because of the very large number of classification variables; but even if the number of classification variables were to be kept small, the differential attrition over the life of the study would still result in an unbalanced sample. Such disproportionalities complicate interpretation of general means, for one must be concerned that an apparent effect is not due to important differences among other variables that are not cancelled out in computing a general mean. Consequently, there is a need for caution in the interpretation of analyses since any factors associated with demographic characteristics are disproportionately represented. Of the disproportionalities explicated above, three appear particularly critical for the comparative analyses of this report: 1) the higher representation of Portland children in the sample, 2) the confounding of race and socioeconomic status as indicated by the higher percentage of ineligible whites and by the lower educational level of black mothers, and 3) the disproportionate number of blacks in Head Start programs.

Data Collection Procedures

Community support and participation were essential if meaningful, useful data were to be obtained. Community leaders and administrators were consulted, and written intents (not merely consents) to participate in the study were sent to ETS by both community agencies and local school boards. Field operations were organized around local staff who served as coordinators, interviewers, testers and observers. During the first year of the study, ETS Princeton staff assumed a major problem-solving role. By the second year, local technical advisors were hired to serve in multiple capacities--advisors, monitors of data collection, and public relations officers. In this way, through the hiring of as many local people as possible, study control and support stemmed more and more from the community.

In Year 1 the first phase of data collection, household canvassing and parent interviews, was subcontracted to the New York firm of Audits and Surveys (A&S) by ETS. A&S' task was first to locate all eligible children within the geographic areas being studied and then to complete a 90-minute interview with each child's mother or mother surrogate. Interviewer supervisors and the local coordinators worked in close cooperation and, where feasible, shared the same field office. (For a detailed description of interviewing procedures, including discussion of some of the logistical problems that arose, the reader is referred to Project Report 70-20 [ETS, 1970].)

The second phase of data collection emphasized individual testing. To this end, local women were preselected by the local coordinator prior to the arrival of the ETS training team. Although the usual educational credentials were not required, previous work experience with young children and the ability to read well and speak with ease were highly desirable. Most of the trainees were housewives who had limited work experience, and most were black.

Once training and evaluations were complete (approximately 4 weeks), each center operated one or two weeks more for a dry run with a trainer from ETS' Princeton office remaining at each center to provide general assistance and additional testing instruction. After actual testing began, monitoring of center operations (except at Trenton) was assumed by ETS regional office personnel with the assistance of Princeton office staff; the Princeton office staff itself monitored Trenton operations.

Since most instruments were not off-the-shelf tasks, and also had never been given by paraprofessional testers, it was important to allow for the flexibility of refining test manuals, formats and procedures to facilitate actual field operations. To this end, the first two sites (Lee County and Portland) were used during training for continued simplification and clarification of testing and scoring procedures based on trainer and trainee experience and suggestions.

Similarly, the grouping of tasks into batteries had been arranged to take into consideration the need to balance type of response (active vs. passive, verbal vs. nonverbal), to maintain constancy of certain sequencing (e.g., Johns Hopkins Perceptual Test before Matching Familiar Figures, since the former involves practice on the responses demanded), to offer a variety of stimuli, and to provide the child with something to take home each day (a photograph, bag of toys, coloring book, Tootsie Roll). In addition, the batteries also had to be representative of the various domains. The first week of dry-run cases in each site piloted the adequacy of the sequencing. After experiences in the first two sites, minor adjustments were made to permit more equivalent testing time and level of test administration difficulty across batteries. Trainees and trainers were encouraged to discuss the merits of the various modifications, and not until it was time to test actual sample children were procedures stabilized for final production of manuals and answer sheets. From such cooperative efforts were derived not only more adequate measurement techniques, but also valuable community-based feedback on research procedures. (Table 2 shows the final order of the tests in the Year 1 batteries.)

Table 2

The Measures and Testing Sequence Used in Year 1

	Estimated Time (in minutes)
Day 1	
*First-Day-of-School Question (mother)	2
*Mother-Child Interaction Tasks:	
Hess & Shipman Toy Sorting Task	15
Hess & Shipman Eight-Block Sorting Task	30
Hess & Shipman Etch-a-Sketch Interaction Task	15
Motor Inhibition Test	10
**ETS Matched Pictures Language Comprehension Task I	5
Battery A	
Preschool Inventory (Caldwell)	20
**Vigor I (Running)	3
**Spontaneous Numerical Correspondence	10
**Massad Mimicry Test I	12
**TAMA General Knowledge I	5
**Risk Taking 1 and 2	20
Picture Completion (WPPSI)	5
Battery B	
Sigel Object Categorizing Test	20
Mischel Technique	2
Johns Hopkins Perceptual Test	10
**Open Field Test	10
**ETS Story Sequence Task, Part 1	10
Seguin Form Board Test	10
Matching Familiar Figures Test	15
Battery C	
Fixation Time	16
**Vigor 2 (Crank-turning)	2
Brown IDS Self-Concept Referents Test	10
Preschool Embedded Figures Test	15
Children's Auditory Discrimination Inventory	10
Peabody Picture Vocabulary Test, Forms A & B	15
**Boy-Girl Identity Task	5
**ETS Enumeration I	7

*Not included in this series of technical reports.
**Tests developed for ETS-Head Start Longitudinal Study.

Testing centers were located in churches or community recreation facilities in or near the districts where the children lived. Each center provided at least six individual testing rooms or partitioned spaces and a larger play and rest area; most also included kitchen facilities. Each center, operating five days a week, was staffed by nine persons--a center supervisor, a play-area supervisor, a driver, and six testers--with each child being scheduled for a four-day testing sequence, usually of 1 1/2 hour duration daily, and the fifth day scheduled for makeups. A rigid schedule was not always possible or desirable, however. For example, centers sometimes operated in the early evenings and on Saturdays for the convenience of working mothers; if necessary, staffs were transferred to new locations to accommodate the children in other school districts within a community; and in the testing situations, testers were instructed to wait until the children were ready, with breaks taken when necessary.

While training procedures in Year 2 were not substantially different from those in Year 1, except that Princeton staff trained local personnel to assume major training responsibility, certain task modifications were made. For example, the Peabody Picture Vocabulary Test was modified to make the items more meaningful for the population under study in terms of racial and regional characteristics and activities pictured. The Sex Role Constancy Task was renamed the Boy-Girl Identity Task, a title less controversial and less technical. Wording in manuals was further simplified, format made more uniform, and special comments referring to frequent errors made in administration and recording included whenever possible. It should be noted that most Year 1 tasks remained essentially the same in Year 2.

Identical items, administration procedures and instructions to the child were used; modifications introduced served primarily to clarify tasks to the testers, thus facilitating training, and, hopefully, increasing similarity of testing procedures across testers and sites.

The only difference in testing procedure in Year 2 was the reduction to a three-day battery (see Table 3). Since only one mother-child interaction task was administered, it became a component of one of these batteries. Each tester was therefore trained on one battery and children were randomly assigned to a battery sequence. Since parent interviews were also being administered by ETS at the testing center, some testers were also trained as interviewers to supplement those women specifically assigned to each center as interviewers. As noted earlier, there was increased monitoring provided by the hiring of a local technical director.

Considering the scope and innovative nature of the study, data collection during the first two years went surprisingly well. Problems arose, of course. (For a discussion of some of these problems, the reader is referred to Project Reports 69-12 and 71-19.) Despite the many difficulties encountered, the study continued operating and reliable data were collected during these first two years.

Data Processing

Various processing operations required for Year 1 and Year 2 child test data included scoring and coding of the raw data, construction and maintenance of the data base, and the design, programming, and execution of various internal and cross-domain analyses. Many of these analyses will not be discussed in this report. Some of these, the initial descriptive analyses of instruments and structural analyses of the Year 1 and

Table 3

The Measures and Testing Sequence Used in Year 2

	Estimated Time (in minutes)
Battery A	
First-Day-of-School Question (mother)	5
Hess and Shipman Eight Block Sorting Task	30
*Vigor I (Hopping)	5
*Stanford Memory Test	10
Boy-Girl Identity Task	10
Children's Auditory Discrimination Inventory	10
Rest-Play	(5)
Preschool Embedded Figures Test	15
**Motor Inhibition Test	10
ETS Story Sequence Tasks, Parts 1 & 2	20
Massad Mimicry II	10
Risk-Taking 2	5
Battery B	
Sigel Object Categorizing Test	20
Vigor 2 (Crank-turning)	5
Fixation Time	20
*Naming Category Instances	15
Rest-Play	(5)
**Peabody Picture Voc. Test, ETS Adaptation, Forms A & B	20
Spontaneous Numerical Correspondence	5
*Gumpgookies	25
Seguin Form Board	5
**Brown IDS Self-Concept Referents Test	15
Battery C	
TAMA General Knowledge Test II	10
**Cooperative Preschool Inventory (Caldwell)	20
Form Reproduction	5
Mischel Technique	2
**Johns Hopkins Perceptual Test	15
**ETS Matched Pictures II	10
Open Field Test	(10)
*Relevant Redundant Cue Concept Acquisition Task	15
*Social Schemata	15
**Matching Familiar Figures Test	15
**Enumeration II	5
*Spatial Egocentrism Task	15

*These tasks were administered for the first time in Year 2 and will be described in later reports.

**Also administered Year 2 in Lee County.

Year 2 test batteries were reported in Project Reports 70-20, 71-19 and 72-18; some were useful as preliminary analyses described in this report but are not of sufficient general interest to be reported in detail; and some will be reported in greater detail in future reports. A detailed account of the design and preparation of the data base is described elsewhere (PR-70-20 and PR-71-19).

Coding.

Typically, data were coded by one person, checked in detail by a second, then spot-checked prior to keypunching. Where appropriate, a reliability check was made by having a small part of the data coded independently by two coders. Each answer sheet was checked for tester error in administration (e.g., allowing the mother to be present, or incurring interruptions on the Fixation Test within a sequence), for recording errors (e.g., not rounding to .2 second on timed tasks or not circling the final response of a multiple response), and for comments that might affect the scoring. Given the inexperience of our testers, considerable time had to be spent preparing the data for coding. Such time, however, was valuable in providing greater familiarity with the actual responses made to a given task and clues to understanding the processes involved.

Analyses of Individual Instruments

In addition to obtaining descriptive statistics on derived scores by age, sex, race and preschool experience breakdowns within and across sites, supplementary analyses have been run on all of the instruments involved in this report. Since instruments differ widely in content, style and presentation, a wide variety of internal analyses were required. For all scores that were composites derived from right-wrong items, tables of item difficulty, biserial correlations of the items with the score, and KR-20 reliability coefficients were computed. For other types of composite scores the alpha coefficient of reliability (KR-20 for dichotomously scored items) was computed. Other secondary analyses were designed by researchers responsible for particular instruments, using such techniques as analysis of variance, product-moment correlations and partial correlations, regression and factor analysis, reliability studies for scores, scorers, and testers, contingency tables, frequency distributions and percentile tables, and several non-parametric rank statistics. Many of these secondary analyses involved transformations of variables, including logarithmic transformations of several positively skewed time scores. The common purpose of these internal analyses was to derive and evaluate comprehensive scores which would represent as well as possible the total information in the test.

In the analysis programs, as well as in the file maintenance program, label checks, data checks, variable checks, program checks, and input control-card checks were all carefully planned to prevent the possibility of incorrect use of data, labels, or programs in a given computer run.

Structural Analyses

Missing data Pearson product-moment correlation tables were constructed for every variable that was placed on the merged file. Correlations were run for the total sample, and for critical breakdowns such as sex, preschool experience, SES level, site, and for subclassifications based upon interactions among these main level factors.

Before using factor analysis techniques it was necessary to reduce the total number of variables from all the instruments, which was approximately

300, to a manageable (and meaningful) subset. The reduced subset was selected by eliminating unreliable variables, subscores and other logically dependent measures. In those cases where two or more scores from a given instrument were logically distinct, not experimentally interdependent, and not very highly correlated with one another, several scores from an instrument were included.

In addition, each variable in the structural analyses was submitted as a dependent variable in two separate sets of ANOVAS. The first used age, sex and SES level as independent variables, analyzing data separately by year. The age data were divided at the median in order to obtain an "older" and a "younger" group. Grouping subjects by mother's educational level provided a rough index of socioeconomic status. The groups consisted of subjects whose mothers had less than 10 years of schooling, at least 10 but not more than 12 years of schooling, and more than 12 years of schooling. To compare data across years, data for subjects tested in both years were subjected to repeated-measures ANOVAS. Since the testing schedule in Year 2 led to some shifting of subjects between the "younger" and "older" categories, no Year 1-Year 2 comparisons of the age data could be made. However, as there was little or no shifting of subjects between SES or sex subgroups, these data could be subjected to repeated-measures analyses of variance performed on their combined Year 1-Year 2 data. The results of these analyses of variance and the results of all the above analyses are reported in each of the following technical reports. It should be noted that the ANOVAS performed separately by year included all subjects having any test data in both years, while the repeated-measures ANOVAS required Ss to have been tested on the same task in both years. Thus, the Ns for the repeated-measures ANOVAS are slightly smaller.

Comments

The initial study report (ETS, 1968) specified a wide variety of measures that we felt would help us describe more adequately the complex interrelationships and structure of children's abilities and characteristics over time and enable us to tease out their interaction effects with particular preschool and primary grade programs. We included measures to encompass the objectives claimed by preschool and primary grade programs, and the aspects of development that social science theory holds as important for human functioning. Measures were also included that would help us to delineate basic cognitive, affective and social processes and their course of development. Whenever possible, multiple sources of information about a particular phenomenon were proposed. We emphasized process rather than static variables. Implicit throughout was our belief that only for the intermediate purpose of structural analysis and derivation of measures within domains could one separate cognitive, perceptual, social and affective domains or study the child without taking his environment into account.

This series of reports describes the interrelationships among certain cognitive, perceptual, affective and social behaviors of the children in the first two years of the study as assessed by the individual test batteries. The questions asked of the data were: To what extent are these indices of the functioning of the preschool child describable in terms of differentiated processes? What is the extent of structural stability and change between Years 1 and 2? Within the particular age period being represented, are differential results obtained by age, sex, social status and/or by their interactions? In addition to contributing to our understanding of the young child, answers to such questions have obvious implications for interpretation of particular test findings obtained in various assessment situations.

It must be emphasized, however, that the data presented here provide only some beginning answers to the questions to which the study is addressed. Further analyses are planned which, it is hoped, will provide a more comprehensive picture of the children in our sample and which will help delineate important sociocultural determinants assessed by measures of home and preschool environments. As noted earlier, the project's focus is on interactions as well as main effects; moreover, the questions being asked must be answered within a framework of repeated measures and observations of the same children (and their parents) over a period of time. Thus, this series of technical reports will be continually updated as such analyses are completed.

The results presented in these reports provide the initial detailed sample description and part of the baseline data for projected future analyses. Only partial answers to questions about the appropriateness of the various measures are provided by these results. Earlier it was pointed out that disproportionalities and confoundings among major classifications make simple interpretations of main effects and inferences about the population hazardous. Much more complex multivariate analyses will need to be performed. Our aim is to delineate the relevant variables so as to understand individual differences and psychologically defined rather than static group differences. Thus age or sex becomes significant only insofar as we understand associated variables that help explain particular interactions, such as those embodied in differential verbal communications between mother and child or the classroom teacher's differential use of praise and blame. Similarly, socioeconomic status is important to the extent that we delineate the component variables associated with the term, and use them as individual predictors within socioeconomic levels. Variables such as socioeconomic status are thus seen as indicators of sets of more basic processes. Established relationships between these indicators and individual behaviors are valuable when they are meaningful summaries or composites of more fundamental process variables or when they suggest hypotheses for more detailed analyses of process variables.

In concluding, certain comments can be made about the data reported here. The most compelling finding from the summary of test results is the wide range of individual differences exhibited in this relatively restricted sample. Low-income youngsters are not a homogeneous group. Youngsters from low-income families span a much wider range of cognitive, perceptual, affective, social, and physical functioning than some would have us believe. Many children performed well in a variety of areas. Conversely, there were very few "untestable" children. Of those problems reported, many were the consequence of the child's interacting with inexperienced testers. Using such a wide variety of tasks, one also becomes more aware of individual differences in the patterning of skills. Knowledge of such patterning of strengths and weaknesses is, of course, a necessary diagnostic tool for the effective planning of educational programs.

In general, tasks in the Year 1 and 2 batteries proved to be appropriate for this age group. They were sensitive to individual differences, were enjoyed by most children, and were relatively easy to administer. Of particular importance for this age group was the fact that the tests were not speeded tests and the administration procedures allowed for great flexibility. Because of the young child's greater susceptibility to situational variables in testing (Sattler & Theye, 1967), the total testing climate was geared toward making the child more comfortable. Time was taken to establish rapport (in some cases, several days), relatively familiar testing rooms in church schools were used, and the tasks were administered by local testers whose dialect and race (wherever possible) were similar to the child's. All of these factors contributed to a congenial and supportive atmosphere. In addition, we attempted to schedule so that each mother could accompany her child on the first testing day. These test conditions, differing as they do from the rigidities of non-essential components of standardized practice, may have contributed substantially to the level of competency observed, as well as to the validity and reliability of measurement.

The data show that research can be done in low-income areas. It is accomplished by making measures as relevant as possible, getting advice from community residents, pretesting tasks in similar communities, and recruiting and training local personnel to carry out most of the operations required. Further, they show that administering individual tests in educational research is not the exclusive prerogative of the educational elite. We have been strengthened in our belief that traditional training models must be questioned: effective training must involve mutual learning and cross-socialization. The local women in our study learned to perform effectively a wide variety of demanding tasks. They managed well under many difficult local situations. Clearly, we have joined many others in discovering a large pool of as yet untapped human resources. Through our continuing joint efforts we hope to provide information that will contribute significantly to the policy-making decisions which affect the well-being of our nation's children and their families.*

*For more specific information for each of the tasks in this series, the reader may contact: Mrs. Susan Simosko
ETS-Head Start Longitudinal Study
Educational Testing Service
Princeton, New Jersey 08540

References

Anderson, S. B. & Doppelt, J. (Jt. Chm.) Untangling the tangled web of education. Symposium sponsored by the National Council on Measurement in Education and presented in conjunction with the conferences of Educational Records Bureau and Educational Testing Service, New York, November 1968. Princeton, N.J.: Educational Testing Service, RM-69-6, 1969.

Educational Testing Service. Disadvantaged children and their first school experiences: Theoretical considerations and measurement strategies. Princeton, N.J.: ETS, December 1968, PR-68-4. Prepared under OEO Contract 4206 and OEO Grant CG-8256.

Educational Testing Service. Disadvantaged children and their first school experiences: From theory to operations. Princeton, N.J.: ETS, August 1969, PR-69-12 (2 Vols.). Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Educational Testing Service. Disadvantaged children and their first school experiences: Preliminary description of the initial sample prior to school enrollment. Princeton, N.J.: ETS, August, 1970, PR-70-20 (2 Vols.). Prepared under Grant CG-8256, Department of Health, Education, and Welfare.

Sattler, J. M. & Theye, F. Procedural, situational, and interpersonal variables in individual intelligence testing. Psychological Bulletin, 1967, 68, 347-360.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N.J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

Technical Report 1

Boy-Girl Identity Task

Walter Emmerich
Karla S. Goldman

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Boy-Girl Identity Task

Background

It has been suggested that achievement of a cognitive judgment of gender identity constancy provides an early and fundamental organizer upon which later sex-role development builds (Kohlberg, 1966b). Inclusion of a measure of sex-role constancy in the Longitudinal Study offered an opportunity to chart its developmental course in an economically disadvantaged group, to relate individual differences in rate of attainment to the child's social environment, and to discover possible relationships between sex-role constancy and other child characteristics, including other cognitive processes and personal-social behaviors. While abilities to identify and verbally label the sexes are known to develop quite early in life (Kohlberg, 1966b), a firm belief that one is a "boy" or "girl" is hypothesized to be one aspect of concrete operational thought; therefore, attainment of this belief is expected to occur by about age six or seven in most children.

Gender identity constancy is assessed using procedures similar to Piagetian techniques for appraising physical constancies and conservations (De Vries, 1969; Kohlberg, 1966b). For example, after presenting the child with a picture of a boy and labeling it as such, the examiner verbally and/or pictorially introduces a sequence of transformations (items) which increasingly resemble the opposite sex. Constancy is signified when the child indicates that the standard stimulus remains a "boy" despite changes suggested by the examiner, and when the child can provide verbal justifications for choices indicative of gender identity constancy.

De Vries (1969) found that bright middle-class boys of three years have some competence on this task and that four-year-olds have attained considerable competence. In our own pilot testing, gender identity constancy was maintained on about half of the present items in a small sample of middle-class four-year-olds.

Task Description and Administration

The present instrument is a refinement of the technique introduced by Kohlberg (1966a) and used by De Vries (1969) in her study of bright middle-class boys. Technical improvements were designed to make the task easier than these earlier versions. The instrument consists of two parallel tasks, each with five items. In Task I a picture and name of a girl are presented to S. Items consist of hypothetical changes introduced by E in which the girl's motives, action, clothing and hair style are modified to resemble these characteristics in boys. For example, Item 1 (Task I) is, "If Janie really wants to be a boy, can she be?" Constancy is indicated when S says that the stimulus remains a "girl" despite the change suggested by E. In Task II, a picture of a boy is presented and named. Items consist of hypothetical changes introduced by E in which the boy's motives, action, clothing and hair style are modified to resemble these characteristics in girls. For example, Item 1 (Task II) is, "If Johnny really wants to be a girl, can he be?" In this case, constancy is indicated when S says that the stimulus remains a "boy" despite the change suggested by E. If the child exhibited constancy on an item, he was also asked to give reasons for his response.

The stimuli were coloring-book-type ink drawings (7 1/2" high) printed on pale blue oaktag (8 1/2" x 11") with clothes and hair colored in by hand. By using these drawings on a blue background, it was hoped to provide race-free stimuli. For Task I, the girl drawing and the boy drawing are hinged together like a book. The girl drawing is on top, slit horizontally, separating the head from the body, and the hypothetical changes are demonstrated by flipping either portion. For instance, "If Janie had her hair cut short like this . . ." is demonstrated by flipping the top portion and revealing the head of the boy drawing underneath. For Task II the drawings are in the reverse order with the boy drawing on top, and the questions are parallel to those in Task I (e.g., "If Johnny let his hair grow long, like this . . .").

The task is quite simple to administer, requiring approximately five minutes. The tester must remember to ask "why?" when a constancy response is given and must be skillful but cautious in the use of probes, so that the child does not think his response is being reinforced (negatively or positively). The tester must also learn to clarify an ambiguous response without influencing it. During actual testing, both testers and children enjoyed the task.

Scoring

Responses indicating constancy were scored 1.0 and those indicating lack of constancy were scored 0.0. If the child's final response to an item (after probing) remained ambiguous, the item was scored .5. Scores reported here are based upon the children's choices only. Future content analyses of children's explanations of their (constancy) choices may yield additional scores.

Score Properties

An underlying assumption in deriving a total sex-role constancy score is that items are sufficiently homogeneous to constitute a unidimensional measure of this construct. As seen in Tables 1 (Year 1) and 2 (Year 2), however, this assumption was unfulfilled in both Years 1 and 2. In both years items I-1 and II-1 had low correlations with the remaining items, although these two items were correlated with each other (r = .45 in Year 1,

Table 1

Year 1 Item Intercorrelations
(N = 1318 to 1374)

		Task I 1	2	3	4	5	Task II 1	2	3	4	5
	1										
	2	.09									
Task I	3	.08	.35								
	4	.05	.22	.27							
	5	-.02	.21	.26	.36						
	1	.45	.10	.07	.07	.04					
	2	.04	.21	.12	.08	.08	.13				
Task II	3	-.01	.07	.12	.02	.06	.04	.31			
	4	-.06	.03	.06	.08	.08	.03	.27	.37		
	5	-.03	.04	.06	.05	.11	.03	.23	.28	.35	

Table 2

Year 2 Item Intercorrelations
(N = 882 to 895)

		Task I 1	2	3	4	5	Task II 1	2	3	4	5
	1										
	2	.14									
Task I	3	.11	.24								
	4	.08	.08	.22							
	5	.03	.19	.28	.26						
	1	.69	.19	.12	.11	.11					
	2	.08	.35	.17	.15	.24	.16				
Task II	3	.02	.15	.23	.08	.13	.08	.33			
	4	-.02	.11	.16	.15	.23	.01	.24	.33		
	5	-.03	.15	.13	.11	.33	.05	.20	.26	.26	

.69 in Year 2). Moreover, item sets I 2-5 and II 2-5 had some internal consistency but tended to be uncorrelated with each other, especially in Year 1. These outcomes indicate that a total sex-role constancy score based upon the sum of the ten items does not constitute an index of a reality judgment based upon concrete operational thought in this sample at these age levels. Consequently, derivation of such a total score is not recommended for use with similar samples.

However, the above correlational patterns suggest that it is possible to derive three distinct subscores representing different kinds of "pseudo-constancy" that might have interesting age-specific (preoperational) relationships with other variables (Emmerich, 1973, in press). The first such score is the sum of items I-1 and II-1, indicative of constancy on the "wish" to be a girl or boy, respectively. The second is the child's summed score on items I-2 through I-5, signifying constancy of the girl stimulus despite suggested changes in activity, clothes, and hair style. The third index is the child's summed score on items II-2 through II-5, indicating constancy of the boy stimulus despite suggested changes in activity, clothes, and hair style.

Table 3 gives internal consistency estimates (coefficient alpha) for the subscores and intercorrelations among the three subscores within and between Years 1 and 2. In examining the stability coefficients, it should be noted that since the measures were not yet consolidated developmentally as indexes of a unidimensional construct, changes in individual ranks may be expected.

The above scoring paradigm is based upon the assumption that a similar correlational structure was present in Years 1 and 2. Inspection of Tables 1-3 suggests that this assumption is reasonable, but there are also indications that the multidimensional structure of Year 2 is less strong than that of Year 1. For example, parallel items across tasks I and II were more highly correlated in Year 2 than in Year 1; indeed, the correlations between I 2-5 and II 2-5 increased between Years 1 and 2 from .16 to .43 (Table 3). Such evidence for increasing unidimensionality with development would be consistent with the theoretical analyses of Flavell and Wohlwill (1969) and Emmerich (1973), and will be evaluated in future years of the study.

Table 3

Subscore Reliabilities and Intercorrelations Within and Between Years 1 and 2 (N = 665-1092)

	Year 1			Year 2			Coefficient
	I-1+II-1	I 2-5	II 2-5	I-1+II-1	I 2-5	II 2-5	Alpha
Year 1							
I-1+II-1	--	.13	.05	.18	.02	-.02	.62
I 2-5	.13	--	.16	.00	.17	.14	.59
II 2-5	.05	.16	--	-.06	.05	.14	.63
Year 2							
I-1+II-1	.18	.00	-.06	--	.23	.10	.81
I 2-5	.02	.17	.05	.23	--	.43	.49
II 2-5	-.02	.14	.14	.10	.43	--	.59

Subscale Meanings

As already noted, the presence of a similar multidimensional item structure in Years 1 and 2 suggests that the three subscales measure distinct preoperational sex-role beliefs, attitudes, or stereotypes. Future analyses are expected to clarify the meanings of these three subscales. For example, it should be possible to study each subscale's developmental (mean) trend, its distinctive pattern of correlates within age periods, and its changing pattern of correlates with age (including later age periods in the Longitudinal Study). These analyses will be guided by a competence-performance model of cognitive structure formulated by Flavell and Wohlwill (1969) and applied by Emmerich (1973, in press) to the sex-role domain.

To illustrate, it is expected that certain items now constituting the subscales will, with development, split off from their respective "pseudo-constancy" clusters and become parts of a reconstituted unidimensional structure indicative of concrete operational thought (genuine sex-role constancy). Paradoxically, this new meaning for an item could initially increase its difficulty for the child, especially if the item is judged particularly difficult in the concrete operational sense. In the Boy-Girl Identity Task, items in the first subscale (I-1 and II-1) were designed to be less difficult than items in the second (I 2-5) and third (II 2-5) subscales. Our model predicts that the second and third subscale mean scores (and perhaps also the first) will decrease developmentally before they increase developmentally. Preliminary findings indicate that the second and third subscale means did decrease significantly for the total longitudinal sample between Years 1 and 2 ($ps < .001$). On the other hand, the first subscale mean increased significantly between Years 1 and 2 ($p < .001$), suggesting that this cluster may have tapped sex-role constancy in the concrete operational sense. These and other implications of the developmental model will be tested in future analyses. Until these fuller implications are more fully tested, however, the underlying meanings of the three subscales remain unknown.

Sample Performance

A repeated-measures analysis of variance (age x sex x SES) performed on the longitudinal sample (those children tested both in Year 1 and Year 2) showed SES, indexed by mother's education (below 10th grade, 10th-12th grade, above 12th grade) to be significant ($F = 13.92$, $df = 2/640$, $p < .001$) for the first subscale with Year 1 and Year 2 scores combined. Children whose mothers had more schooling had higher scores on this subscale. SES thus appears to be linked with the early development of sex-role constancy in the concrete operational sense. For the second and third subscales, these analyses did not reach significance. However, examining the difference between scores in Year 1 and Year 2, sex was a marginally significant variable ($F = 4.61$, $df = 1/618$, $p < .03$) for the third subscale. Boys showed a greater difference across years than girls. Boys had higher scores than girls in each year and also had more rapidly decreasing scores from Year 1 to Year 2. This is consistent with the pseudo-constancy interpretation.

An additional analysis of variance (age x sex x SES) performed on the longitudinal sample separately in each year revealed no significant age trends in Year 2. In Year 1, however, the "wish" subscale demonstrated age to be marginally significant ($F = 4.13$, $df = 1/1022$, $p < .05$) with older Ss having higher scores. For the "boy" subscale (II 2-5) age also was significant ($F = 13.64$, $df = 1/978$, $p < .001$), as was an age x sex interaction ($F = 5.45$, $df = 1/973$, $p < .02$). Younger Ss had higher scores with younger boys having the highest, again consistent with the pseudo-constancy interpretation.

Relationship with Other Measures

Correlations with other measures in the study were low. It is useful, though, to look at the correlations between each of the three subscales and the Preschool Inventory which in a series of structural analyses had the highest loadings on the first factor which was defined as general information-processing skills (see Shipman, 1971, 1972, for detailed presentation of these results). In Year 1 the correlations were .18, -.04, and -.24, indicating some relationship with the cognitive factor only in the case of the first score, consistent with the above interpretation. In Year 2 the correlations were .21, .18, -.19, indicating little change. Similar correlations were also found over the two years with the Peabody Picture Vocabulary Test, which had high loadings on the first factor.

Correlations with other Piagetian measures in the study were low, but highest for the first subscale. In Year 1, correlations for the three subscales with Enumeration I were .16, -.04, and -.03; and with Spontaneous Numerical Correspondence (total correct), they were -.07, -.08, and -.08. In Year 2, correlations with Enumeration II (pointing) were .05, -.01 and -.06; and with Spontaneous Numerical Correspondence they were .25, -.06, and -.19.

Summary

Since the ten items did not constitute a homogeneous scale either in Year 1 or 2, their sum is not recommended as an index of gender identity constancy in this population during the preschool years.

Similar multidimensional item structures were found in Years 1 and 2. From these it was possible to derive three reasonably independent subscales having some internal consistency. The first of these subscales may index the beginnings of sex-role constancy based upon concrete operational thought. The other two subscales apparently measure preoperational sex-role beliefs and attitudes. While these subscales are expected to become reconstituted as a single homogeneous dimension when our sample enters the period of concrete operations, they may have meaning in their own right during Years 1 and 2. Future analyses will attempt to establish these meanings.

A model formulated by Flavell and Wohlwill (1969) and applied by Emmerich (1973, in press) suggests a number of hypotheses for the Boy-Girl Identity Task. These hypotheses refer to developmental changes in (a) item structure, (b) mean levels, (c) stability coefficients, (d) experiential correlates and (e) cognitive correlates in the child. Preliminary evidence bearing on some of these developmental changes was presented above, but fuller tests await analyses which include the children's responses to the instrument at a later period of their development; i.e., during the transition to concrete operational thought.

References

De Vries, R. Constancy of generic identity in the years three to six. Monograph of the Society for Research in Child Development, 1969, 34, (3, Serial No. 127).

Emmerich, W. Socialization and sex-role development. In P. B. Baltes and K. W. Schaie (Eds.), Life-span developmental psychology: Personality and socialization. New York: Academic Press, 1973, in press.

Flavell, J. H. & Wohlwill, J. F. Formal and functional aspects of cognitive development. In D. Elkind and J. H. Flavell (Eds.), Studies in cognitive development: Essays in honor of Jean Piaget. New York: Oxford University Press, 1969. Pp. 67-120.

Kohlberg, L. Stages in children's conceptions of physical and social objects in the years four to eight. Unpublished manuscript, 1966. (a)

Kohlberg, L. A cognitive-developmental analysis of children's sex-role concepts and attitudes. In E. E. Maccoby (Ed.), The development of sex differences. Stanford, Calif.: Stanford University Press, 1966. Pp. 82-173. (b)

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

Technical Report 2
Brown IDS Self-Concept Referents Test

Virginia C. Shipman
Lynn E. Gilbert

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Brown IDS Self-Concept Referents Test

Purpose

An underlying assumption about the development of children is that their potential for learning is enhanced when they feel secure, are able to relate well to others, and have a generally positive self-concept. Self-concept is closely related to motivational variables, which have been shown to influence achievement, including performance on intelligence tests (see Zigler & Butterfield, 1968). There has been considerable documentation of the negative impact of very severe assaults on self-esteem, a situation to which economically disadvantaged, and particularly minority children, are more likely to be exposed. For these reasons a favorable self-concept is in itself one of the specified goals of the Head Start program; thus it was considered crucial to include measures of self-esteem in the Longitudinal Study test battery.

As early as 1943, Horowitz attempted to measure nursery-age children's ideas about themselves. She used pictures of figures varying in age, size, manner of dress, and number of family members and asked, "Which one is you?" Yet, in contrast to numerous studies of language development and cognitive functioning, there are very few studies of the emergence and development of the self-concept in young children (Wylie, 1961). Several reasons for this are offered by Brown (1966), including limited ability of young children to conceptualize and verbalize feelings about themselves, instability of the self-concept at a young age, and lack of appropriate measures.

More recently, however, a number of investigators have attempted to assess self-concept among preschool children. Long and Henderson (1968), using a task in which children of mixed SES selected and arranged symbols to represent themselves in relation to others, found low SES Ss to have a lower self-concept, and low self-esteem to be associated with immature classroom behavior. McAdoo (1971) assessed self-concept by having children respond to adjectives in reference to their own photographs. In her sample (southern rural and northern urban black Ss enrolled in a year-long Head Start program), there was no correlation between self-concept and IQ as measured by the Peabody Picture Vocabulary Test; for the self-concept measure southern children and boys scored higher, and there was no significant difference between children from intact and non-intact homes.

Several studies have attempted to relate self-concept to cognitive skills and school achievement. To assess whether low self-concept leads to poor performance or whether poor performance generates a low self-image, Wattenberg and Clifford (1962) collected data on children from two socioeconomic groups when they were in kindergarten and again in the second grade. Self-concept was measured through teacher ratings, sentence completions, and tape recordings of Ss' remarks as they drew pictures of their families. They found self-concept more predictive of reading achievement than was mental ability, and the self-concept and intelligence measures were not highly correlated. Moreover, when they separated self-concept in reference to competence and to good behavior they found them to be relatively independent.

Clark, Ozehosky, Barz and O'Leary (1967) studied self-concept and vocabulary development in black and white preschoolers. Children were asked to select "the real you" in each of 50 bipolar drawings (the U-scale), and were given a picture vocabulary test. White Ss exhibited superior vocabulary skills but were not significantly different in their self-concept scores.

Also investigating the relationship between self-concept and verbal ability, Williams (1969) assessed self-concept by means of the U-scale and teacher ratings and administered a picture vocabulary test. His subjects were black preschool children, some in a suburban, integrated program and others in an urban segregated one. There were no significant differences between urban and suburban boys, while suburban girls were superior both in self-concept and vocabulary scores. There was no significant sex difference for children from father-absent homes.

In two related studies, Ozehosky and Clark (1970, 1971) examined the relationship of self-concept to kindergarten achievement and also explored the possibility of differential effectiveness of verbal and non-verbal measures. They administered a sentence-completion test and their pictorial U-scale as the two measures of self-concept. Metropolitan Readiness Tests and Grade Point Average, as well as the U-scale, discriminated (at the .01 level) Ss previously rated by their teachers as having high and low self-concepts. However, the verbal test was a poor discriminator for boys and differentiated girls only at the .05 level; approximately 30% of the responses were unscorable. These results led the investigators to conclude that the verbal measure "may reflect differences in verbal fluency rather than differences in their phenomenal selves [1971, p. 199]."

Most self-concept studies compare various groups in the way they perceive themselves or relate self-concept to achievement variables, with little or no attention paid to other variables that may be producing these differences (e.g., confounding of teacher attitudes affecting both self-concept ratings and grades, other differences in sample characteristics). Also, there are little data concerning developmental trends in the young child's emerging self-image. It was hoped that inclusion of a self-concept measure in the Longitudinal Study would provide information about the child's growing ability to make differentiated judgments about himself and those family and school variables associated with such differentiation.

The Brown IDS Self-Concept Referents Test, developed in 1966, is a technique for assessing self-concept using a photograph of the young child to induce him to take the role of another toward himself. It is based on the notion of "self" as rooted in social experience and interaction, of "self-concept" as formed by an individual's own perceptions of how others perceive him. Looking at the photograph, the child answers questions in terms of his own feelings about himself and then in terms of what his mother, teacher, and classmates think. This measure was included in the present study because it was one of the few available at the time relating to the child's evaluation of "self-as-subject" and "self-as-object" which had reliability data and evidence of validity for use with four-year-old disadvantaged children. In Brown's (1966) sample of black lower-class and white middle-class preschool children, the self-perceptions of the black children were significantly less favorable than those of the white children, and black children perceived their teachers as viewing them less favorably. However, black and white children did not differ in their perception of either their mothers' or their peers' evaluations. Test-retest reliability for the self-referent responses was .71 for blacks and .76 for whites. These findings were later replicated by Brown (1967). Walker, Bane and Bryk (1973) refer to a modified version of the Brown (2 additional self items combined with 5 teacher-referent items) administered to kindergarten and first-grade children in the Fall 1971 Follow Through evaluation. The KR-20 test reliability coefficient was .82, and the test-retest coefficient for 632 Ss after a 2-3 week interval was .55.

Task Description and Administration

A full-length color Polaroid photograph is taken of the child, with no instructions to "smile" so that a spontaneous facial expression may be obtained. After the tester ascertains that the child recognizes himself in the picture, the child is asked to respond to 15 bipolar items (e.g., Is (child's name) happy or is he sad?; Is (child's name) afraid of a lot of things or is he not afraid of a lot of things?), each time with reference to the photograph. All items are presented in an "either-or" format, and positive choices are randomly assigned first and second position. Since data with preschool children indicate that they may have difficulty understanding the difference between "self" and "other" referents, only the self-referent part of the test was administered in the first year of the study. In Year 2 the teacher-referent was also presented to those children attending a preschool program.

Testers need particular training in operating the camera and in learning how to respond appropriately with a nonverbal child or one who has difficulty understanding the task. Administration takes about 10 minutes. This is generally an enjoyable task for the children since they look forward to having their pictures taken.

Scoring

In the present study, each item was scored as positive (1), negative (0), refused, indeterminate (e.g., mulitple answers), or "don't know." Total number of omitted items, that is, ones to which the child did not make a differential response, and an adjusted total self-concept score were the two scores used in the present analyses. The self-concept score was adjusted in order to account for omitted items and was computed as the percentage of positive responses for those items clearly answered in either a positive or negative way. (Maximum score is 100%, based on the first 14 items. Item 15, regarding number of friends, was not included in the computation since it was not part of the original test. It was added because of the possibility of relating this information to actual observations of peer interaction in later study years.) In addition each child was judged as smiling or not smiling in the photograph.

In Year 2 supplementary scoring was introduced to reflect the child's initial response pattern to each item: whether he verbalized one specified alternative; verbalized both (gave a multiple answer); qualified his answer (e.g., "sometimes"); said "yes" or "no" only; nodded or shook head only; gave both a verbal and nonverbal response; gave no response; gave other task-related responses (e.g., "good" when asked if smart or stupid); or gave irrelevant responses. These results will be examined in future analyses and hopefully will provide supplementary cues regarding changes in self-concept. For example, older children and those showing other evidence of greater cognitive complexity may be expected to have more differentiated feelings about themselves and be more likely to qualify their answers. Such data will be valuable in highlighting problems in assessment using self-report measures.

Score Properties

Table 1 presents the alpha coefficient of reliability for total score and number of omits for Years 1 and 2. Alphas ranged between .59 and .91. Since the adjusted total score is based on different items for different Ss the alpha coefficient was not appropriate. Preliminary analyses suggest that an overall reliability statistic may disguise substantial variations among subgroups on this task. For example, in Year 2 the alpha for Total Omits on the self-referent task was .87 for boys and .79 for girls, whereas on the teacher-referent task the alphas were .71 and .89 for boys and girls, respectively. (Alpha coefficients for the total sample were .84 and .85, respectively.)

Table 1

Score Reliabilities* for Years 1 and 2

Score		N	Year 1 Reliability	N	Year 2 Reliability
Total Score:	Self-Referent	1372	.64	1299	.59
Total Score:	Teacher-Referent	--	--	717	.69
Total Omits:	Self-Referent	1441	.91	1314	.84
Total Omits:	Teacher-Referent	--	--	723	.85

*Coefficient alpha

R-biserials for the unadjusted total score were generally high. In Year 1 they ranged from .42 to .64, with nine over .60. For the self-referent task in Year 2 the range was .40 to .79 with 11 items over .60; and for the teacher-referent task the biserials were between .41 and .86, with eleven of the 14 items exceeding .65. The item with the lowest biserial in each case was item 6 (like to have other kids' things/own things). There were too few omitted items to produce meaningful biserial correlations for the total omits score.

The correlation between the adjusted and unadjusted total scores was .83 in Year 1. In Year 2, as number of refusals and indeterminate responses decreased, it was .93 for the self-referent score and .97 for the teacher-referent score. The correlation across years on the self-referent task was .22 and .33 for the adjusted total and total omits scores, respectively. To the extent that the total omits score reflects cognitive competencies, the difference in stability coefficients is evidence for the greater consistency of cognitive than affective and social behaviors during this age period, a general finding for this test battery. Table 2 presents the correlations of scores within and across years for the self-referent task. Correlations for the teacher-referent scores are shown in Table 3.

Table 2

Intercorrelations of Brown Self-Concept Test Scores, Within and Across Years (Self-Referent only)

	Total Score		Adjusted Total Score		Total Omits		Smiling	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
Total Score, Year 1		.27	.83	.24	-.67	-.16	.12	.16
Year 2			.23	.93	-.22	-.56	.06	.19
Adjusted Total Score, Year 1				.22	-.15	-.09	.08	.13
Year 2					-.15	-.22	.04	.16
Total Omits, Year 1						.33	-.11	-.17
Year 2							-.07	-.12
Smiling, Year 1								.22
Year 2								

Note.--*N*s range between 1085 and 1143.
$r_{.001}$ = .104 for *N* = 1000.

Table 3

Intercorrelations of the Self-Referent and Teacher-Referent Scores in Year 2

	Self-Referent			Teacher-Referent	
	Total Score	Adjusted Total Score	Total Omits	Total Score	Adjusted Total Score
Total Score, Teacher	.64	.61	-.34		
Adjusted Total Score, Teacher	.62	.62	-.24	.97	
Total Omits, Teacher	-.32	-.15	.70	-.44	-.23

Note.--Correlations are based on data for the total Year 2 sample (*N* = 708-714).

Sample Performance

Tables 4, 5 and 6 present the score distributions for adjusted and unadjusted total scores and number of omits in Years 1 and 2. As has been found in previous studies, the distribution of the self-concept score was markedly skewed (M = 82.0% in Year 1 and 86.2% in Year 2 for the self-referent task), indicating the strong tendency for the child to select positive attributes. Figures for the teacher-referent task in Year 2 were quite similar, with M = 85.8%. The similar means, as well as comments reported on test protocols, suggest that children at this age were not clearly differentiating between the self- and teacher-referents.

Analyses of Variance (age x sex x SES*) were performed on the longitudinal sample, those children tested in both Year 1 and Year 2.

Table 4

Distribution of Scores: Self-Referent, Year 1

Group	Total Score N	Total Score M	Total Score SD	Adjusted Total Score N	Adjusted Total Score M	Adjusted Total Score SD	Total Omits N	Total Omits M	Total Omits SD
42-44 mo.	72	9.9	2.69	72	76.5%	15.2	83	2.7	4.28
45-47 mo.	296	10.5	2.66	296	81.3%	15.0	313	1.8	3.17
48-50 mo.	327	10.7	2.45	327	82.1%	14.4	339	1.4	2.72
51-53 mo.	357	10.8	2.28	357	82.1%	14.2	373	1.3	2.72
54-56 mo.	261	11.1	2.35	261	84.7%	14.1	273	1.4	2.94
57-59 mo.	58	10.4	2.16	58	78.8%	14.9	59	1.0	2.16
Boys	717	10.6	2.45	717	81.8%	14.7	756	1.6	3.05
Girls	654	10.7	2.45	654	82.2%	14.4	684	1.5	2.88
Total	1371	10.7	2.45	1371	82.0%	14.6	1440	1.5	2.97

*Mother's education was used as the index of SES: below 10th grade, grades 10 to 12, above 12th grade. Although the children were classified into these three groups for analysis purposes, almost all were of relatively low SES.

Table 5

Distribution of Scores: Self-Referent, Year 2

Group	Total Score			Adjusted Total Score			Total Omits		
	N	M	SD	N	M	SD	N	M	SD
51-53 mo.	76	11.5	2.09	76	85.0%	13.6	76	0.4	0.94
54-56 mo.	296	11.6	2.09	296	86.0%	13.0	303	0.8	1.82
57-59 mo.	303	11.8	1.87	303	85.7%	12.4	308	0.4	1.44
60-62 mo.	346	11.9	2.02	346	86.7%	13.0	349	0.4	1.35
63-65 mo.	244	11.9	2.01	244	86.9%	12.9	243	0.4	1.42
66-69 mo.	20	11.6	1.64	20	83.8%	11.8	20	0.2	.37
Boys	679	11.7	2.04	679	85.7%	13.3	688	0.5	1.61
Girls	606	11.8	1.95	606	86.7%	12.4	611	0.5	1.33
Total	1285	11.8	2.00	1285	86.2%	12.8	1299	0.5	1.48

Table 6

Distribution of Scores: Teacher-Referent, Year 2

Group	Total Score			Adjusted Total Score			Total Omits		
	N	M	SD	N	M	SD	N	M	SD
51-53 mo.	49	11.3	2.51	49	83.4%	17.0	49	0.5	.10
54-56 mo.	156	11.5	2.50	156	84.1%	16.3	157	0.5	1.59
57-59 mo.	158	11.7	2.31	158	84.7%	15.7	159	0	1.17
60-62 mo.	196	12.0	2.05	196	87.0%	13.8	198	.3	1.12
63-65 mo.	136	12.2	1.98	136	87.9%	13.6	138	0.3	1.20
66-69 mo.	13	12.2	1.57	13	86.8%	11.2	13	0.0	0.00
Boys	381	11.9	2.24	381	86.0%	15.3	382	0.2	0.83
Girls	327	11.8	2.23	327	85.5%	14.7	332	0.5	1.60
Total	708	11.8	2.24	708	85.8%	15.0	714	0.3	1.25

For the adjusted total score, using a median-age split for each year separately, only marginally significant age differences were found in Year 1 ($F = 5.37$, $df = 1/1032$, $p < .025$). In a repeated-measures ANOVA SES was significant beyond the .001 level ($F = 23.96$, $df = 2/999$) when data were combined across years, with low SES children obtaining the lowest mean self-concept scores and high SES children, the highest. There was also a significant age x SES interaction ($F = 7.09$, $df = 2/994$, $p < .001$); among low and middle SES children, the younger children scored lower, whereas among the high SES subjects the younger ones scored slightly higher. This suggests a curvilinear relationship paralleling developmental changes in the meaning of the scale (e.g., lack of comprehension versus emerging differentiation both leading to an indeterminate response). No significant sex differences occurred.

The correction for unscorable responses would have spuriously inflated the self-concept score to the extent that they reflected a defensive response rather than the child's lack of differentiation with regard to a particular item. However, the distinct drop in the mean number of unscorable items (1.5 in Year 1 and 0.5 in Year 2) and in the percentage of indeterminate responses per item suggests that these reflected instead poorer comprehension of the task in the earlier study year.

For the total omits score, the repeated-measures ANOVA with data combined across years showed SES to be a significant variable ($F = 18.16$, $df = 2/1052$, $p < .001$); low SES children omitted the most items and high SES children, the least. An SES x Year interaction was also significant ($F = 8.27$, $df = 2/1052$, $p < .005$), with low SES subjects having the largest difference in number of omits from Year 1 to Year 2, and high SES subjects having the smallest difference. No significant sex differences were found.

ANOVA performed for each year separately to determine age effects revealed marginally significant age differences ($F = 5.10$, $df = 1/1087$, $p < .025$) in Year 1, with younger children omitting more items.

Item intercorrelations were generally quite low (few reached .20) and did not reveal any patterning either within or across years. The only two item pairs with moderate correlations were items 9 and 11 ($r = .39$ in Year 1 and .30 in Year 2) and items 10 and 14 ($r = .39$ in Year 1 and .35 in Year 2); in both cases, the items in the pair differ by only one word (scared of a lot of things/people--not scared; likes the way clothes/face looks--doesn't like). As the child matures and his self-image becomes more differentiated, there may be a more distinct clustering of items. The child's self-concept may also become increasingly influenced by sex role stereotypes. For example, the analyses of variance revealed no significant sex differences for the total score; however, on most items girls tended to choose the positive attribute more often than did boys. The most notable exception was the strong-weak category, consistent with expectations concerning sex role identification. These data are congruent with the usual finding that girls are more advanced in their understanding of socially desirable responses.

The majority of children (66.6% in Year 1, 50.7% in Year 2) did not smile for their picture. This may reflect general lack of familiarity with having their pictures taken and lack of ease in the testing situation. Smiling in the photograph correlated only .13 in Year 1 and .24 in Year 2 with the happy (rather than sad) response to item 1. Thus, children did not seem to be responding primarily on the basis of immediate stimulus cues.

Relationship with Other Measures

Correlations of the Brown Self-Concept Test (adjusted total score) with other tasks were generally quite low, ranging between .01 and .19 for the three-site longitudinal sample in Year 1. It correlated .19 with the Peabody Picture Vocabulary Test (PPVT), Form A, tapping receptive language skills; and .17 with the Preschool Embedded Figures Test. In Year 2, the highest correlation of the Brown was with Form B of the PPVT, measuring productive language (*r* = .15). It correlated .14 with the PPVT Form A, Spontaneous Numerical Correspondence configuration matching score, and Massad Mimicry real words score. This low communality with other tasks was perhaps due to the paucity of measures in the test battery tapping affective behaviors.

The correlations of the omit and smiling scores were generally low with other tasks for the three-site longitudinal sample in both years. In Year 1 the omit score correlated highest with the Preschool Inventory (*r* = -.28) and with the Preschool Embedded Figures Test (*r* = -.25); in Year 2, given a more restricted range, highest correlations were with the Seguin Form Board test, time to quickest solution (*r* = .22), the PPVT, Form B (*r* = -.21), and the Preschool Inventory (*r* = -.18). Smiling in Year 1 correlated highest (*r* = .16) with the PPVT, Form A, and Vigor 2 (*r* = .15) and in Year 2 with fastest time to completion on the Seguin Form Board test (*r* = -.16), PPVT, Form A (*r* = .15) and Vigor 2 (*r* = .14).

Factor analysis of the child test data revealed a factor defined by the self-concept score. In Year 1 this factor correlated .25 (Promax 13-factor solution) with the factor representing general information-processing skills, suggesting a cognitive task component. Either the self-concept score at that age is partly a measure of understanding or brighter children are more self-confident or both; the present data do not enable us to determine which is the more appropriate interpretation. In Year 2, the self-concept factor correlated .22 with the general abilities factor, a finding which may reflect the reduced extent to which comprehension confounds performance on this task in an older sample. In Year 2 extension analyses, the Gumpgookies Test, measuring achievement motivation (including one's feelings of competence as a learner), correlated .30 with the self concept factor. (See Shipman, 1971, 1972 for further discussion of these structural analyses.)

Summary

The present data support Ozehosky and Clark (1971) in their suggestion that among preschool children a test of this nature reflects verbal facility as well as self-concept. In Year 1 there was a mean of 8.8% indeterminate responses per item, and across years children of lower SES were significantly higher in their frequency of omitted items. Because of the high degree of verbal comprehension required and because of ceiling problems, Stanford Research Institute (1972) concluded that the Brown, in spite of adequate internal reliability, should not be used in large scale evaluations of preschool children. Inspection of the test protocols also indicates the uneven difficulty in the wording of items and suggests that the use of implied double negatives is particularly to be avoided in testing young children.

There is evidence in the present data, however, that the importance of verbal comprehension was reduced among older children. The mean number of omitted items dropped from 1.5 to 0.5 from Year 1 to Year 2, and the self-concept factor obtained from factor analyses of the child test data correlated less with the general intellectual competency factor in Year 2.

Results are generally consistent with previous research (Brown, 1966; Clark et al., 1967; Walker, Bane & Bryk, 1973) in that most children did tend to select the socially desirable attribute. Self-concept in very young children is not easily defined and may become more clear with increasing age. One may hypothesize a developmental pattern consisting of three stages: first, a self-concept not clearly differentiated and therefore not readily assessed by verbal-report instruments; second, a period in which the self is considered in global terms along a "good-bad" dimension; and third, a differentiated self-concept in a variety of personality dimensions. Accordingly, scores would increase and then decrease with maturation and experience. Thus, older children would be expected to have more differentiated self-feelings which could result in lower scores, and sex differences may be evidenced as responses increasingly reflect sex-role stereotypes.

In subsequent study years, self-concept scores will be related to school achievement and to peer interaction data obtained through classroom observations. Moreover, we will investigate the extent to which the young child's generally high self-esteem is maintained as his experience with others increases, noting differences among economically disadvantaged Ss associated with emphases on ethnic pride in the community, home and school. Of particular interest will be the extent of congruence over time in parallel measures of self esteem, e.g., in an individual testing situation and in observation of school settings.

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. N. Y.: New York University, N. Y. Institute for Developmental Studies, Office of Education (DHEW), Washington, D. C., April 1966. (Shorter version presented as paper at meeting of the Eastern Psychological Association, New York, April 1966.)

Brown, B. Personal communication, 1967.

Clark, E., Ozehcsky, R., Barz, A. & O'Leary, J. Self-concept and vocabulary development in Negro and white preschool children. Paper presented at the meeting of the Eastern Psychological Association, Boston, April 1967.

Horowitz, F. A pictorial method for study of self-identification in preschool children. Journal of Genetic Psychology, 1943, 62, 135-148.

Long, B. & Henderson, E. Self-social concepts of disadvantaged school beginners. Journal of Genetic Psychology, 1968, 113, 41-51.

McAdoo, H. Racial attitudes and self-concepts of black preschool children. Dissertation Abstracts International, 1971, 31 (8-A), 3963.

Ozehosky, R. & Clark, E. Children's self-concept and kindergarten achievement. Journal of Psychology, 1970, 74 (2), 185-92.

Ozehosky, R. & Clark, E. Verbal and non-verbal measures of self-concept among kindergarten boys and girls. Psychological Reports, 1971, 28 (1), 195-199.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Stanford Research Institute. Results of fall 1971 Follow Through supplemental testing. Washington, D. C.: Bureau of Elementary and Secondary Education, Department of Health, Education, and Welfare, 1972.

Walker, D., Bane, M. J. & Bryk, T. The quality of the Head Start Planned Variation data. Cambridge, Mass.: Huron Institute, 1973, in press.

Wattenberg, W. & Clifford, C. Relationship of the self-concept to beginning achievement in reading. Detroit, Mich.: Wayne State University, 1962. (Final Report of Cooperative Research Project #377.)

Williams, D. E. Self-concept and verbal mental ability in Negro preschool children. Dissertation Abstracts, 1969, 29, 3475B.

Wylie, R. The self-concept. Lincoln, Nebraska: University of Nebraska Press, 1961.

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

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

A LONGITUDINAL STUDY

Technical Report 3

Children's Auditory Discrimination Inventory

Virginia C. Shipman
Karla S. Goldman

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Children's Auditory Discrimination Inventory

Background

Although there appears to be a relationship between auditory discrimination and certain cognitive skills, measures for determining auditory discrimination are often confounded by differences in speech patterns and vocabulary level. Most of these measures use verbal stimuli which require the child to differentiate words spoken in standard English. Based on this type of measure, many studies have revealed marked socio-ethnic and age differences.

Observing 24 black first, third, and fifth graders in New York City, Deutsch (1964) found a positive relationship between Wepman Auditory Discrimination Test (WADT) scores, reading, and age. It was found that the scores of good readers were higher than those of poor readers, and older children were superior to younger Ss. Research by Stern (1966) indicated that children's auditory discrimination may be assessed by asking children to identify pictures which represent terms, both nonsense and meaningful, presented orally. Nonsense terms were included as a set of terms equally unfamiliar to all subjects. Elenbogen and Thompson (1972) investigated the relation of familiarity of word pairs on the WADT to task performance. Thirty middle-class and 15 lower-class kindergarten Ss were individually administered the WADT and a modified version of the WADT. In the modified version of the WADT, phonemes were exchanged between pairs to create nonsense syllables. There was no significant class difference on the modified version of the WADT; however, middle-class Ss scored significantly higher than lower-class Ss on the standard WADT. These results suggest that tests which include meaningful verbal stimuli may measure a vocabulary factor in addition to auditory discrimination. Goetzinger, Dirks and Baer (1960), using middle-class 10- to 17-year-old boys, suggested that auditory memory may also be an influencing variable.

Hendrix (1968) found auditory discrimination differences between an advantaged and disadvantaged group of preschool children. The tests used to measure auditory discrimination ability were the Boston University Speech Sound Discrimination Picture Test, Wepmar Auditory Discrimination Test, Templin Sound Discrimination Test, and Wilson Environmental Sounds Test. The advantaged preschoolers scored significantly higher than the disadvantaged preschoolers on all four tests. No significant difference between males and females was found. Similar findings have been reported by other researchers (Clark & Richards, 1966; Deutsch, 1965; Oakland, 1968; Stern, 1966). Deutsch (1964) attributed auditory discrimination deficit in disadvantaged children to limited exposure to meaningful speech, and to excessive amounts of noise in their environment. Stressing the importance of experience and exposure to adequate stimuli, she claimed that the signal-to-noise ratio is influential in the stimulus perceived and in the response evoked. The higher the ratio (i.e., the greater the amount of "signal" as compared to "noise") the more likely will be the accurate perception of the signal. The link between this and Gestalt formulations regarding figure-ground relationships in visual perception is clear.

The relatively poor performance on auditory discrimination tests among disadvantaged children could be due to the instruments' insensitivity to differences between black dialect and standard English speech. Labov's (1969) work suggests many pronunciation differences between the nonstandard English of black disadvantaged children and standard English. Working with black disadvantaged children in New York City, he found that there were words pronounced as homonyms by black dialect speakers which were spoken contrastingly by standard English speakers. Thus, knowledge of a child's speech background must be considered before his auditory discrimination ability can be meaningfully evaluated. Using three groups of 40 first-grade boys consisting of black dialect-speakers and black standard English speakers, Gottesman (1972) investigated differences in auditory discrimination. The auditory discrimination test was composed of word pairs pronounced as contrasting words by all Ss. Findings revealed no significant group differences in performance on these contrasting pairs. However, on those word pairs pronounced as homonyms in black dialect but as contrasting words in standard English, standard English-speaking Ss scored significantly higher than black dialect-speaking Ss.

In other research, however, it has been shown that black children have a keener awareness of standard/nonstandard differences (Politzer & Hoover, 1972). This may be the result of training, or of greater exposure to both standard and nonstandard speech. Using the Standard Discrimination Test (SDT) consisting of phonological and morphological variables, Politzer and Hoover studied lower- and lower-middle-class black and white children in grades 2, 4, and 6. An analysis of variance of the SDT scores showed that test scores increased with age; girls performed better than boys; and black children performed better than white children, Correlations were high between SDT scores and standardized reading achievement scores for black children, but for the whites, correlations were significant only at the sixth grade level.

It appears logical to infer from these studies that many of the tests used to measure auditory discrimination ability have been inadequate. When comparing social classes or ethnic groups, socio-cultural differences must be taken into account. Confounding by differences in language usage and verbal ability must be eliminated in order to assess the child's level of auditory discrimination.

Task Description and Administration

The Children's Auditory Discrimination Inventory (CADI), developed by Stern (1966), is an individually administered measure to evaluate the child's ability to identify, between two pictures that have been given oral word equivalents, the picture that represents the orally presented stimulus word. This ten-minute task uses 38 cards with two pictures on each and 38 pairs of stimulus words. One picture in each pair is a real picture representing a familiar word and the second picture is a nonsense design paired with a nonsense word. In order to minimize the effects of positional responding, the real and nonsense words are randomly located on the right or left side of the cards. (The real word, however, is always presented first.)

The tester presents each pair of pictures, naming each one as she points to it. Following presentation of each pair, the child is asked to point to the picture named by the tester. The first two items are for practice, and if the child does not respond correctly to these the procedure is explained again. If, after a repetition, the child still does not respond correctly, the test is discontinued. During the test an item may be repeated only once. Also, to prevent lip reading, the tester covers his mouth slightly when saying the stimulus words. Training involves considerable practice in pronouncing the words clearly and correctly. It is recommended that a tape recorder be used for training and for screening out testers with even minor speech impediments.

Scoring

The tester records whether the child's response was correct or incorrect, whether the item was repeated, and whether there was a need to probe for the final answer. The score is the total number correct (range = 0 to 38). For the present study, subscores were also used for real word items (0 to 19) and nonsense word items (0 to 19).

Score Properties

Using the Kuder-Richardson (formula 20) estimate of reliability, internal consistency for the test was found to be .81 (N = 1455) in Year 1, and .82 (N = 895) for Year 2. Internal consistency in Year 1 for the real word items was .76 and for the nonsense word items was .85; in Year 2 it was .76 and .86, respectively. The correlation between performance on real word items and nonsense word items in both years was .03, while the uncorrected part-whole correlation with total score was .53 in Year 1 and .50 in Year 2 for real words and .86 in Year 1 and .88 in Year 2 for nonsense words. Across years performance on real word items correlated .19; nonsense words correlated .45. The lack of correlation between subscores and the difference in stability coefficients suggest that separate scores for real words and nonsense words should be used instead of the total score.

Sample Performance

As can be seen in Table A in the Appendix, although approximately 98% of the sample attempted to respond to each item, those items involving nonsense words had a smaller proportion passing each item than did the items involving real words. The range of percent passing each nonsense word item was 43 to 95 in Year 1, and 44 to 97 in Year 2, with only five item scores in each year above 80%. However, for real word items the range was 72 to 98 in Year 1 and 78 to 100 in Year 2, with eighteen (of nineteen) item scores above 80% in each year. For the Year 1 biserials, only one score was below .50, and 5 were below .60. For the nineteen real words the mean number of correct responses was 16.6 (SD = 2.69) in Year 1 and 17.2 (SD = 2.37) in Year 2. For the nineteen nonsense words the mean was 12.2 (SD = 4.49) in Year 1 and 12.8 (SD = 4.41) in Year 2. Thus, real words were less discriminating and appeared easier, although this may also be due to bias in choosing real over nonsense stimuli. Also, those nonsense items requiring discrimination of multiple consonants were more difficult than those involving simple consonants.

Tables 1, 2, and 3 present means, standard deviations and range of responses for the total sample by three-month age breakdowns and by sex. For this sample the data indicate that children's auditory discrimination increases with age for children between 42 and 69 months, and sex differences appear slightly in favor of girls. A comparison of Tables 2 and 3 again demonstrates that the real words were not as difficult and showed less variability than the nonsense words, possibly due to the fact that the children were more likely to point to the familiar picture.

A repeated-measures analysis of variance (sex x age x SES) performed on the longitudinal sample (those children tested in both years) showed SES, as measured by mother's education (below 10th grade, 10th-12th grade, above 12th grade), to be significant when scores were summed across years: Real Words, $F = 12.81$, $df = 2/697$, $p < .001$; Nonsense Words, $F = 32.88$, $df = 2/692$, $p < .001$; Total Correct, $F = 41.06$, $df = 2/690$, $p < .001$. Sex was significant favoring girls on Nonsense Words ($F = 5.54$, $df = 1/692$, $p < .02$) and Total Score ($F = 8.52$, $df = 1/690$, $p < .004$). Also, there was a significant increase in mean scores from Year 1 to Year 2: Real Words, $F = 15.03$, $df = 1/701$, $p < .001$; Nonsense Words, $F = 58.79$, $df = 1/696$, $p < .001$; Total Score, $F = 92.57$, $df = 1/694$, $p < .001$. In age x sex x SES ANOVAS within each year, age (using a median split) was not significant in Year 2, but in Year 1 it was significant for Nonsense Words ($F = 9.57$, $df = 1/1090$, $p < .005$) and for Total Correct ($F = 11.47$, $df = 1/1088$, $p < .001$).

Relationship with Other Measures

The striking differences between performance on the real word items and the nonsense word items make it somewhat difficult to interpret the total

Table 1

Total Score*: Means, Standard Deviations and Range by Sex and Age in Year 1 and Year 2

Group	N	Mean	SD	Range
Year 1				
42-44 mo.	84	27.69	4.92	18-37
45-47 mo.	310	28.42	5.10	16-38
48-50 mo.	327	28.70	5.57	3-38
51-53 mo.	382	28.94	5.39	15-38
54-56 mo.	274	29.20	5.36	16-38
57-59 mo.	61	29.84	4.28	21-38
Boys	758	28.39	5.38	3-38
Girls	680	29.22	5.19	11-38
Total	1438	28.79	5.31	3-38
Year 2				
51-53 mo.	67	29.75	5.45	18-38
54-56 mo.	219	29.75	5.34	13-38
57-59 mo.	196	29.46	4.69	19-38
60-62 mo.	249	30.25	5.13	14-38
63-65 mo.	155	30.54	4.95	18-38
66-69 mo.	8	29.13	4.58	23-35
Boys	468	29.37	5.16	13-38
Girls	426	30.60	4.92	18-38
Total	894	29.96	5.08	13-38

*Range = 0-38.

Table 2

Real Words as Correct Response*: Means, Standard Deviations and Range by Sex and Age in Year 1 and Year 2

Group	N	Mean	SD	Range
Year 1				
42-44 mo.	84	16.17	2.56	9-19
45-47 mo.	312	16.42	2.64	5-19
48-50 mo.	333	16.72	2.58	1-19
51-53 mo.	383	16.57	2.89	3-19
54-56 mo.	275	16.77	2.71	5-19
57-59 mo.	61	16.64	2.32	10-19
Boys	762	16.43	2.87	1-19
Girls	686	16.77	2.47	5-19
Total	1448	16.59	2.69	1-19
Year 2				
51-53 mo.	67	16.72	2.62	9-19
54-56 mo.	219	17.18	2.56	7-19
57-59 mo.	196	17.06	2.30	8-19
60-62 mo.	249	17.28	2.41	5-19
63-65 mo.	155	17.34	2.34	6-19
66-69 mo.	8	18.25	.71	17-19
Boys	468	17.02	2.51	6-19
Girls	426	17.36	2.20	5-19
Total	894	17.19	2.37	5-19

*Range = 0-19.

Table 3

Nonsense Words as Correct Response*: Means, Standard Deviations and Range by Sex and Age in Year 1 and Year 2

Group	N	Mean	SD	Range
Year 1				
42-44 mo.	85	11.55	4.06	3-18
45-47 mo.	310	12.00	4.21	1-19
4 -50 mo.	327	12.00	4.58	2-19
51-53 mo.	383	12.38	4.65	0-19
54-56 mo.	275	12.46	4.71	1-19
57-59 mo.	61	13.20	3.75	3-19
Boys	760	11.98	4.54	0-19
Girls	681	12.47	4.42	1-19
Total	1441	12.21	4.49	0-19
Year 2				
51-53 mo.	67	13.03	4.63	2-19
54-56 mo.	219	12.57	4.52	1-19
57-59 mo.	196	12.40	4.10	3-19
60-62 mo.	249	12.96	4.54	2-19
63-65 mo.	155	13.20	4.32	0-19
66-69 mo.	8	10.88	4.42	5-16
Boys	468	12.35	4.50	0-19
Girls	426	13.24	4.27	1-19
Total	894	12.77	4.41	0-19

*Range = 0-19.

(combined) score. It is possible that the familiar words primarily measure vocabulary skill and not auditory discrimination, thus favoring those children with more extensive language and educational backgrounds. Further evidence for this confounding may be suggested by CADI's correlations (total score) with other tasks in the battery. Its highest correlation* in Year 1 was with the Peabody Picture Vocabulary Test (PPVT), Form A, .61, followed by the Preschool Inventory, .51; and by the PPVT, Form B, .48. Although also tapping auditory discrimination, Massad Mimicry scores showed only moderate correlations (nonsense words, r = .31; meaningful words, r = .30). In light of this evidence the total inventory cannot be recommended as a test of auditory discrimination.

Correlations for the real words and nonsense words separately with the other scores in the Year 1 structural analysis were highest with the Preschool Inventory (.29, .44, respectively), and PPVT, Form A (.27, .52), and PPVT, Form B (.16, .41). Other correlations were only moderately high: Form Reproduction total score (.20, .29), ETS Enumeration, total correct (.26, .28), ETS Enumeration, counting (.19, .28), Motor Inhibition, walking (.11, .28), Motor Inhibition, drawing (.09, .25), Story Sequence I, receptive (.16, .27), and the Johns Hopkins Perceptual Test, total correct (.18, .22). As can be seen, correlations in general were substantially higher with the nonsense wo[illegible] score.

Factor analyses of the Year 1 child data for both the total sample and the three-site longitudinal sample revealed a distinct information-processing factor (Shipman, 1971, 1972). As might be expected, the most general test in the battery, the Preschool Inventory, had the highest loading. But all of the following had loadings of .50 or higher: perceptual measures (CADI, Johns Hopkins Perceptual Tes[illegible] Matching Familiar Figures, errors), verbal

---

*All correlations in Year 1 are for the total sample.

measures (PPVT, Sigel Object Categorization, Toy Sorting, Eight-Block Sorting), and perceptual-motor measures (Seguin Form Board, Form Reproduction). When the CADI subscores were used, nonsense words also had a high loading on this factor, but real words did not. This first component seemed to be best defined as "g" or general information-processing skills which contribute to level of performance. Possibly this is evidence of the importance of auditory abilities for the development of certain cognitive skills and abilities; at this age, however, it is more likely that the child's ability to comprehend the task and willingness to follow directions were the primary variables being assessed.

In Year 2 for the four-site longitudinal sample, CADI total score again correlated highest with the Preschool Inventory (r = .58) and PPVT, Form A, (r = .53). The highest correlation for the nonsense word score was with the Preschool Inventory (r = .53), followed by PPVT, Form A, .48; TAMA, .44; ETS Matched Pictures, .42; PPVT, Form B, .40; and Matching Familiar Figures, errors, -.40. Performance on the real word items also correlated highest with the Preschool Inventory and PPVT, Form A, but these were low: r = .25 for each. Similarly, in the factor analysis of the three-site longitudinal sample data, nonsense words had a loading of .63 on factor 1 (general information-processing) in the 13-factor Varimax solution, while real words had a loading of only .22 (Shipman, 1972). The real word score also clustered with other scores relating to word familiarity (correct labels on the Sigel Object Categorization Test; Mimicry, real words), defining a separate factor in this same solution.

Summary

The nonsense word score appears to be the more adequate indicator of auditory discrimination at this age in this population. It demonstrated higher internal consistency, greater stability and higher loadings on a general competency factor. In addition, nonsense words are equally unfamiliar to all Ss whereas the real word items involve vocabulary skill to the extent that children have differential familiarity with the meanings of the real words. As suggested by Shontz (1971), however, the real word subscore may be useful as an index of impulsive selecting of the concrete or familiar.

Although significant sex differences were found in favor of girls, these differences in absolute terms were quite small. Older children in Year 1 obtained higher scores, and children's mean performance increased from Year 1 to Year 2 and with level of their mothers' schooling, reflecting maturational and experiential influences. Given the paucity of measures directly assessing auditory discrimination in the present study, it is not known to what extent the SES differences obtained reflected specific differences in the child's verbal environment.

The present findings exemplify the difficulty of obtaining an auditory discrimination index for young children which eliminates confounding with language usage and verbal ability. Future data collection will enable analysis of relationships to performance on the Wepman and measures of academic achievement, particularly reading skills.

Appendix

Table A

Percent Passing Each Item in Year 1 and Year 2

Item Number	Percent Passing		Biserials*	
Real Words	Year 1	Year 2	Year 1	Year 2
1 Girl, Hujuj	98	99	.35	.23
4 Clock, Koopay	97	100	.63	.20
5 Sleeping, Sagrole	94	99	.62	.39
7 Boat, Boatch	72	78	.63	.63
8 Dog, Dob	88	88	.54	.59
10 Duck, Dup	80	81	.67	.70
12 Scissors, Frissors	88	91	.67	.76
15 Jump, Dump	86	91	.61	.84
18 Church, Schurch	84	88	.77	.84
21 Shirt, Sirt	85	89	.83	.99
24 Bus, Bush	88	90	.76	.73
26 Cow, Tow	85	90	.81	.85
29 Table, Pable	86	89	.70	.93
30 Fish, Fith	91	93	.89	.93
31 Bed, Bej	87	90	.75	.86
33 Ball, Gall	92	93	.71	.90
34 Children, Tildren	86	89	.80	.70
35 Dress, Dreth	87	90	.84	.86
36 Falling, Thalling	84	88	.62	.66
Nonsense Words				
2 Phone, Volvap	95	97	.51	.56
3 Horse, Ulna	92	96	.51	.66
6 Wagon, Zagon	83	93	.57	.60
9 Sun, Thun	74	82	.65	.70
11 Egg, Edd	63	67	.67	.80
13 Hat, Hap	65	64	.72	.71
14 Shoes, Thoes	80	88	.66	.74
16 Plane, Plame	50	48	.66	.66
17 Valentine, Thalentine	46	48	.66	.69
19 Book, Dook	81	87	.68	.74
20 Mouse, Mouf	51	57	.81	.88
22 Leaf, Leath	45	44	.67	.67
23 Coat, Poat	67	66	.79	.81
25 Door, Goor	70	74	.82	.90
27 Stove, Stothe	46	46	.71	.70
28 Read, Reab	59	57	.64	.66
32 Money, Noney	61	70	.82	.86
37 Sock, Sot	43	45	.74	.76
38 Brush, Brutch	48	49	.75	.78

*Correlation of each item with the subscore total, i.e., of each Real Word item with the Real Word total and of each Nonsense Word item with Nonsense Word total.

References

Clark, A. & Richards, C. Auditory discrimination among economically disadvantaged and nondisadvantaged preschool children. Exceptional Children, 1966, 33, 269-272.

Deutsch, C. Auditory discrimination and learning: Social factors. Merrill-Palmer Quarterly, 1964, 10, 277-296.

Deutsch, M. The role of social class in language development and cognition. American Journal of Orthopsychiatry, 1965, 35, 78-88.

Elenbogen, E. & Thompson, G. A comparison of social class effects in two tests of auditory discrimination. Journal of Learning Disabilities, 1972, 5 (4), 36-39.

Goetzinger, C., Dirks, D. D. & Baer, C. J. Auditory discrimination and visual perception in good and poor readers. Annals of Otology, Rhinology and Laryngology, 1960, 69, 121-136.

Gottesman, R. L. Auditory discrimination in Negro dialect-speaking children. Journal of Learning Disabilities, 1972, 5 (2), 38-45.

Hendrix, L. J. Auditory discrimination differences between culturally deprived and nondeprived preschool children. Dissertation Abstracts, 1968, 28 (7-A), 2556.

Labov, W. Some sources of reading problems for Negro speakers of nonstandard English. In J. Baratz and R. Shuy (Eds.), Teaching black children to read. Washington, D.C.: Center for Applied Linguistics, 1969. Pp. 29-68.

Oakland, T. D. Social class and performance on phonemic and nonphonemic auditory discrimination tests. Dissertation Abstracts, 1968, 28 (11-A), 4490.

Politzer, R. L. & Hoover, M. R. The development of awareness of the black standard/black nonstandard dialect contrast among primary school children: A pilot study. Stanford: Stanford Center for Research and Development in Teaching, 1972. Research and Development Memorandum, No. 83.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N.J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shontz, D. A comparative analysis of agricultural migrant and inner city Negro disadvantaged preschool children on several psychoeducational dimensions. (Abstract) Florida State University, Tallahassee, Florida, 1971.

Stern, C. Language competencies of young children. Young Children, 1966, 22 (1), 44-52.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

YOUNG CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

A LONGITUDINAL STUDY

Technical Report 4

ETS Enumeration I and II

Judith A. Meissner
Virginia C. Shipman
Edward A. Chittenden

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

ETS Enumeration I and II

Background

ETS Enumeration I and II were developed specifically for the ETS-Head Start Longitudinal Study to measure various precursors of quantitative performance. Enumeration I tests the ability to itemize each member of a set by using a task that requires neither counting nor reciting the names of numerals. The purpose of the task is to assess the child's ability to organize a field of figures and to keep track of two shifting sets--a set of figures "pointed at" and a set "not-yet-pointed at." Enumeration II also measures this pointing ability while incorporating types of items found in primary-grade tests of quantitative ability (i.e., items requiring counting and the matching of equal numbers and spatial orderings of objects).

Enumeration I and the Pointing items of Enumeration II are patterned closely after a procedure used by Potter and Levy (1968). Preschool children ages 2 1/2 to 4 were asked to touch each figure on a stimulus page "just once." Pages of figures varied in number, arrangement, meaningfulness, and heterogeneity of figures. It was found that (1) success on touching items was correlated with age, (2) arrays with the smallest number of figures were easiest, and (3) random arrangements of figures were most difficult. Potter and Levy concluded that their enumeration technique "may be of value to investigators of early cognitive development [1968, p. 272]" as a method of studying attentional processes thought to be important in the development of concepts of quantity.

Many of the school readiness tests, such as the Preschool Inventory (ETS, 1970), the Stanford Early School Achievement Test (Madden & Gardner, 1969), and the Tests of Basic Experiences (Moss, 1970), have a mathematics or number subtest, which generally taps such skills as dot counting, number conservation and the reading of numbers. Many preschool programs for disadvantaged children emphasize the development of these number skills, and several evaluation studies such as those of Head Start Planned Variation and Project Follow-Through (Bissell, 1972; Walker, Bane & Bryk, 1973) are currently using preschool and primary achievement tests with number subtests as part of their test battery; however, very little technical information is presently available from disadvantaged samples. One exception is a recent study by Washington and Teska (1970) using 96 disadvantaged children ranging in age from 5-7 to 7-5, where the arithmetic section of the Wide Range Achievement Test was found to correlate .84 with the California Achievement Test (Primary Form) arithmetic section, .70 with Binet MA, and .68 with total score on the Illinois Test of Psycholinguistic Abilities.

A modification of ETS Enumeration II was used in the Head Start Planned Variation evaluation (Walker, Bane & Bryk, 1973). Test-retest reliability coefficients over pairs of testers for a two-week time period (for a sample of approximately 20 children) were moderate, ranging from .11 to .81. Analyses of variance revealed significant time differences across all subjects for the total score. No significant difference between types of Head Start programs was found. One of the most appealing aspects of the ETS Enumeration II test, according to Walker, Bane & Bryk (1973), is its systematic attempt to measure the various components in learning mathematical skills, and it was for this reason that the test was originally developed and used in the present study.

Task Description and Administration

The ETS Enumeration tasks consist of a number of colored circles or small pictures, each on an 8 1/2" x 11" page in a loose-leaf notebook. The

Examiner shows the child each page, one at a time, and asks the child either to point to or to count each circle on that page or to point to the correct stimulus picture. Sample items are provided to orient the child to the nature of the task, and practice items precede each new item type. In orienting the child to the task, testers are cautioned to be sure that the child understands the response procedures required for each type of item (i.e., pointing to each circle only once, counting the circles or pointing to the correct picture).

(a) Enumeration I. In the first year of the study, the Enumeration task consisted of one practice item and twelve test items. Only circles were used in the Year 1 version, the number of circles on each test page varying from six to nine. The child was requested to point once, and once only, at each circle on each test page. A thirteenth item, on which the child was asked to count aloud a line of seven circles, was also included to aid in preparing a Counting subtest for the second year of the study.

(b) Enumeration II. The Year 2 version of Enumeration was divided into four parts. In the Counting section of the task (items 1-4), each item consisted of a page of either six or nine circles and the child was asked to count the circles o. page. The second section of the test, Pointing (items 5-8), also consisted of circles. After a practice item the child was asked to point to each circle on a page only once. This group of items was a subset of those administered in Enumeration I. In the last two parts of the test, the child was asked to choose which one of three pictures in a row was the "same" as a stimulus picture at the top of the page. Items 9-16 tested the concept of "Same Number" and items 17-21 were concerned with appropriate "Order."

Scoring

(a) Enumeration I. The enumeration items (items 1-12) were scored "correct" if the child pointed to each circle in the array once, neither omitting nor repeating a circle. If the performance was incorrect, the nature of the errors was coded: "Omits," "Repeats" or "Omits and Repeats." The possible range for total correct scores was 0-12. The counting item (item 13) was scored "correct" and given one point only if the child counted aloud to "7" correctly. The counting item was not included in the total correct score, but was analyzed separately.

(b) Enumeration II. The Counting items (items 1-4) of Enumeration II each received a possible two points of credit. One point was given for correctly counting aloud the circles in a sequence, as credit for accurate itemization of the circles. Another point was given for stating the correct total number in response to the question "how many," as credit for understanding cardinal value. All other items in the test were given one point credit. The Pointing items (items 5-8) required the child to point to each circle on the test page once without repeating or omitting any item. The pictorial items (items 9-21) were scored correct if the child pointed to the picture having the same number (items 9-16) of objects as did the stimulus picture or the appropriate requested order (items 17-21). Thus, the possible range of total scores for the test was from 0 to 25. For the present analyses, Total Score and subscores for Counting, Pointing, Same Number and Order were computed.

Score Properties

(a) Enumeration I. Table 1 gives the item difficulties and coefficient alpha reliability for the twelve test items of Enumeration I. This table indicates that accuracy of performance was systematically influenced by the number and arrangement of circles on a stimulus page, a finding congruent with the conclusions of Potter and Levy (1968) in their enumeration study with nursery school children. As indicated in Table 1, the twelve items can be described as six pairs of items, since the members of each pair are alike in number and arrangements of circles. As shown, mean scores were higher for items with six circles (items 1-6) than for items with nine circles (items 7-12). Within these two groups of items, difficulty was related to the

Table 1

Enumeration I
Percent Passing Each Item and Item Biserials* (N = 1459)

Item Number and Description	Percent Passing	Biserial
1. 6 two-colored circles in a line	74.7	.79
2. 6 three-colored circles in a line	68.5	.79
3. 6 two-colored circles in two rows	60.4	.80
4. 6 three-colored circles in two rows	56.3	.78
5. 6 two-colored random circles	50.0	.76
6. 6 three-colored random circles	50.2	.79
7. 9 two-colored circles in a line	46.7	.78
8 9 three-colored circles in a line	44.2	.80
9. 9 two-colored circles in three rows	38.2	.82
10. 9 three-colored circles in three rows	37.3	.81
11. 9 two-colored random circles	26.5	.74
12. 9 three-colored random circles	26.8	.76

Alpha = .85.
*Correlation of each item with the total score.

arrangements of circles, single lines of circles (items 1, 2, 7 and 3) being easier than rows of circles (items 3, 4, 9 and 10), which in turn were easier than random arrangements (items 5, 6, 11 and 12). As can be seen from the table, the percent of children passing an item steadily declined from 74.7

to 26.5. The biserials, which ranged in magnitude from .74 to .82, gave evidence for the homogeneity of the items, as did the alpha reliability of .85. Alphas within the range of .81 to .86 were found on all analyses of consistency for age and sex subgroups.

Item intercorrelations are shown in Table 2. These phi coefficients were highest for pairs of items of like type (i.e., 1 and 2, 3 and 4, etc.). The data in Tables 1 and 2 suggest that the item-pairs could form a scale of difficulty which reflects the effect of increasing number and complexity of arrangement of the circles on a stimulus page.

Table 2

Correlation* Matrix for Enumeration I Items (N = 1459)

	1	2	3	4	5	6	7	8	9	10	11	12
1.	.	0.45	0.35	0.31	0.26	0.32	0.34	0.31	0.29	0.29	0.22	0.22
2.	0.45	.	0.34	0.34	0.28	0.34	0.35	0.33	0.29	0.29	0.25	0.23
3.	0.35	0.34	.	0.46	0.36	0.34	0.30	0.30	0.35	0.29	0.23	0.26
4.	0.31	0.34	0.46	.	0.38	0.30	0.30	0.31	0.34	0.33	0.22	0.25
5.	0.26	0.28	0.36	0.38	.	0.40	0.30	0.26	0.31	0.30	0.29	0.27
6.	0.32	0.34	0.34	0.30	0.40	.	0.32	0.35	0.31	0.34	0.29	0.28
7.	0.34	0.35	0.30	0.30	0.30	0.32	.	0.40	0.34	0.31	0.28	0.28
8.	0.31	0.33	0.30	0.31	0.26	0.35	0.40	.	0.39	0.36	0.33	0.31
9.	0.29	0.29	0.35	0.34	0.31	0.31	0.34	0.39	.	0.45	0.29	0.36
10.	0.29	0.29	0.29	0.33	0.30	0.34	0.31	0.36	0.45	.	0.34	0.36
11.	0.22	0.25	0.23	0.22	0.29	0.29	0.28	0.33	0.29	0.34	.	0.39
12.	0.22	0.23	0.26	0.25	0.27	0.28	0.28	0.31	0.36	0.36	0.39	.

*Phi coefficients.

(b) Enumeration II. Item difficulties, biserial correlations with total score, and coefficient alpha reliability for the 21 test items comprising Enumeration II are presented in Table 3. The Pointing items (items 5-8) were the easiest items as a group, and several of the Order items (items 17-20) were among the hardest. The biserials showed considerable variability, with item correlations with total score ranging from .23 to .87. The Counting items, with all biserials above .80, were especially high, and the Order items, for which three of the five biserials were .30 or below, were especially low.

Table 3

Enumeration II
Percent Passing Each Item and Correlations with Total Score (N = 1292-1306)

Item Number and Description of Stimulus Pictures	Percent Passing	Correlation with Total Score
1. 6 three-colored circles in a line	65.9	.84
2. 6 three-colored random circles	54.1	.87
3. 9 three-colored circles in a line	47.9	.87
4. 9 three-colored random circles	32.6	.83
5. 6 two-colored circles in a line	86.5	.60
6. 6 two-colored random circles	73.8	.67
7. 9 two-colored circles in a line	68.8	.69
8. 9 two-colored random circles	40.5	.69
9. 3 birds	54.4	.42
10. 4 circles	44.7	.40
11. 3 cylinders	75.9	.43
12. 5 walnuts	35.4	.43
13. 5 fish	38.8	.34
14. 7 apples	74.4	.41
15. 9 balloons	49.1	.46
16. 7 lollipops	18.1	.23
17. 3 flowers	36.6	.43
18. clothes on clothesline	67.3	.25
19. 3 fish going thru tunnel	35.7	.26
20. 4-car train	28.7	.30
21. 2 turtles	91.5	.32

Alpha = .77

Table 4, which presents the intercorrelations and coefficient alpha reliabilities for the four different groups of items (i.e., Counting, Pointing, Same Number and Order) and the Total Score, shows the Counting, Pointing and Same Number items correlated quite highly with the Total Score (with uncorrected part-whole correlations ranging from .63 to .88), but the correlation of the Order items with the Total Score was quite low (r = .41). The Counting items were the most internally consistent (alpha = .88), followed by the Pointing and Same Number items, which have moderate consistency (.57 and .41, respectively). The internal consistency of the Order items was very low (.11). This, plus the wide difficulty range within this item group (29% to 91% passing an item) and their low correlation with the other subtest scores (.15, .14 and .26 with the Counting, Pointing and Same Number items, respectively) would suggest that the Order items are measuring different abilities than are the other three subtests of Enumeration II. Intercorrelations between the Counting, Pointing and Same Number subscores were also relatively low in magnitude, ranging from .32 to .42.

Table 4

Enumeration II
Subscore Intercorrelations and Reliabilities* (N = 1194-1292)

	(Counting) Items 1-4	(Pointing) Items 5-8	(Same Number) Items 9-16	(Order) Items 17-21	Total Score
Reliability*	.88	.57	.41	.11	.77
Items 1-4		.42	.39	.15	.88
Items 5-8			.32	.14	.63
Items 9-16				.26	.70
Items 17-21					.41

*Coefficient alpha.

Sample Performance

Tables 5 and 6 show the distributions of Total Correct Score for the enumeration items (1-12) and the counting item (item 13) of Enumeration I by sex and by three-month age groups. As shown in Table 5, there was a steady rise in mean total correct as a function of age, with a mean of 4.1 for the youngest group and 7.1 for the oldest. These data support the finding of Potter and Levy (1968) who, with a group of 58 nursery school children ages 2 1/2 to 4, found that accuracy of performance was clearly correlated with age.

Table 5

Enumeration I
Distribution of Total Correct Score* by Age and Sex

Group	N	Mean	SD	Percentiles 10	25	50	75	90
42-44 mo.	82	4.1	3.12	0.1	1.3	3.5	6.2	8.2
45-47 mo.	306	5.1	3.66	0.3	1.9	4.8	7.9	10.4
48-50 mo.	323	5.7	3.53	0.6	2.7	6.0	8.5	10.5
51-53 mo.	367	6.4	3.39	1.6	3.7	6.6	9.0	11.0
54-56 mo.	259	6.6	3.27	2.1	4.1	7.0	9.2	10.9
57-59 mo.	58	7.1	3.15	2.6	4.4	7.4	9.4	11.4
Boys	738	5.6	3.48	0.7	[illegible].8	5.7	8.3	10.4
Girls	657	6.2	3.54	1.0	3.[illegible]	6.4	9.0	11.0
Total	1395	5.9	3.52	0.8	3.0	6.0	8.6	10.7

*Range = 0-12.

Table 5 shows the scores to approximate a normal distribution for the group as a whole. The 50th percentile for the total group coincides almost exactly with the mid-point of six; the 25th and 75th percentiles are located evenly at scores 3.0 and 8.6, respectively. The mean score for girls is slightly higher than that for boys. Data on performance on the counting item (Table 6) also show the percentage passing this item to increase steadily with age. Since the item contained seven circles in a line, the data may be compared to items 1 and 2 (six in a line) and 7 and 8 (nine in a line). Only 29% could pass the counting item, while percentages passing these comparison non-counting items were 75, 69, 47 and 44, respectively. Therefore, for these four-year-old subjects, the request to count a line of circles aloud was a much more difficult task than was that of enumerating all circles by pointing to them.

Table 6

Enumeration I
Distribution of Correct Score on Counting Item (Item 13)*

Group	N	Mean	SD	Percent Response 0	Percent Response 1
42-44 mo.	76	0.13	0.34	86.8	13.2
45-47 mo.	280	0.22	0.41	78.2	21.8
48-50 mo.	307	0.26	0.44	74.3	25.7
51-53 mo.	355	0.36	0.48	64.5	35.5
54-56 mo.	259	0.36	0.48	64.5	35.5
57-59 mo.	59	0.41	0.50	59.3	40.7
Boys	697	0.27	0.44	73.2	26.8
Girls	639	0.32	0.47	67.9	32.1
Total	1336	0.29	0.46	70.7	29.3

*Scoring: 0 = Incorrect; 1 = Correct.

Table 7 presents the distribution of Enumeration II Total Score data by sex and three-month age intervals. This table shows that each successively older age group attained a slightly higher score on the test, and that girls as a group scored somewhat higher than did boys.

Table 7

Enumeration II
Means, Standard Deviations and Percentile Distributions
of Total Correct Score*

Group	N	Mean	SD	Percentiles 10	25	50	75	90
51-53 mo.	75	11.1	5.13	5.7	7.4	10.1	14.5	17.7
54-56 mo.	284	11.7	5.01	5.9	7.6	11.1	16.0	18.6
57-59 mo.	267	12.4	4.83	6.8	8.3	12.1	16.3	19.5
60-62 mo.	330	13.3	4.85	7.2	9.1	13.1	17.1	19.1
63-65 mo.	222	14.4	5.16	7.4	10.0	14.4	18.3	21.5
66-69 mo.	15	14.4	4.81	7.8	11.4	12.9	18.4	21.2
Boys	626	12.2	5.00	6.2	7.8	12.0	16.3	19.1
Girls	567	13.5	5.03	7.3	9.3	12.9	17.6	20.1
Total	1193	12.8	5.06	6.8	8.4	12.4	17.1	19.6

*Range = 0-25.

Age x sex x SES ANOVAS were performed for the longitudinal subjects of the study only (i.e., those subjects for whom data were available in both Years 1 and 2). Since the Enumeration Tasks did not contain the same items in Years 1 and 2, separate analyses of variance were performed on the data from the two years.

When subjects were divided at the median into a "younger" and an "older" group, a significant age difference in favor of the "older" subjects was found in both Year 1 ($F = 30.82$, $df = 1/1057$, $p < .001$) and Year 2 ($F = 35.21$, $df = 1/1091$, $p < .001$). Potter and Levy (1968) also found significant age differences in enumeration ability with their 2 1/2- to 4-year-old subjects, and two other studies of number concepts of preschool children found significant age differences (Rothenberg & Courtney, 1969; Siegel, 1971). The other measure of preoperational number concepts used in the Longitudinal Study, Spontaneous Numerical Correspondence, also found age differences in both Total Score and Configuration Score.

Sex differences were significant for both Enumeration I ($F = 11.93$, $df = 1/1057$, $p < .001$) and for Enumeration II Total Score ($F = 12.75$, $df = 1/1091$, $p < .001$), both favoring girls. Potter and Levy (1968) also found significant sex differences in their four-year-old group in favor of girls.

To test for SES-based differences, the longitudinal subjects of the study were divided into three groups on the basis of their mothers' educational level. ANOVAS were significant beyond .001 in both Years 1 and 2, with the high SES children (children whose mothers had more than 12 years of schooling) obtaining higher Enumeration scores than the middle SES (children with mothers having more than 10 but less than 12 years of schooling) or low SES subjects (those whose mothers had less than 10 years of schooling). The F-ratios were 21.19 for Enumeration I ($df = 2/1057$) and 77.18 for Enumeration II Total Score ($df = 2/1091$). Large SES differences in preoperational number ability have also been found by Rothenberg and Courtney (1969), Siegel (1971) and Almy (1966).

Relationships with Other Measures

Correlations* of Enumeration I total score with other measures in the

---

*These correlations are based on data from the longitudinal subjects only.

Longitudinal Study test battery were generally quite low, ranging from .00 to .34. The two highest correlations were with Form Reproduction total score (r = .34) and Preschool Embedded Figures Test (PEFT) total score (.33), both of which have a distinct perceptual component. Other correlations of .30 or higher included a negative relationship with Seguin log fastest time (-.32) and a correlation of .30 with Preschool Inventory adjusted total score. The "Pointing" items (items 5-8) in the Enumeration II test also correlated most highly with Form Reproduction (.35), Preschool Inventory (.36), Seguin log fastest time (-.31) and with ETS Matched Pictures total score (.30). Thus, in both years, these items appear to have a perceptual component, but in Year 2, general information tests had slightly higher correlations than in Year 1, indicating that the nature of the task may be changing somewhat over time.

Enumeration II total score had a much greater number of correlations above .30 with other tasks than did Enumeration I, with a range of correlations from .01 to .69. As in Year 1, the correlations of the highest magnitude were those with total scores on Preschool Inventory (r = .69) and Form Reproduction (.53). The general information and vocabulary tests, TAMA and Peabody Picture Vocabulary Test (Form A), also correlated highly with Enumeration II (r = .49 and .53, respectively), as did Matched Pictures total score (.50). Other tests with correlations above .40 were Johns Hopkins Perceptual Test total (.45), Children's Auditory Discrimination Inventory total score and nonsense word score (.46 and .42), Seguin Form Board fastest time to correct placement (r = -.46) and Hess and Shipman Eight-Block Sorting Task total score (.40). Peabody Picture Vocabulary Test (Form B), Naming Categories, Motor Inhibition, Spontaneous Numerical Correspondence total score and Massad meaningful words (final sounds) all correlated .38 with Enumeration

II total score. The correlations of Enumeration total score with the other numerical test in the battery, Spontaneous Numerical Correspondence, in both Year 1 (r = .22 with the Spontaneous Numerical Correspondence total correct score, and .20 with Configuration Matching) and Year 2 (r = .38 and .24, respectively) show the relationship between these two tests to be positive but relatively low, compared to other intercorrelations.

Factor analysis of the Year 1 and Year 2 test data showed the highest loading of both Enumeration I and Enumeration II to be with a general information-processing factor ("g") that accounted for most of the common variance among the cognitive-perceptual tasks. Loadings on all other factors were considerably below .30 (see Shipman, 1971, 1972, for detailed presentation of these findings).

Summary

The ETS Enumeration Test, which was included in the Longitudinal Study battery as a measure of early numerical ability, is relatively simple to administer, although care is required in orienting the child to the task during the practice trials. The first year's experience with the task led to the recommendation that the different-colored circles within a circle that appear in some of the items be eliminated in subsequent versions. These circles proved to be confusing to some children, who literally attempted to point to "all" the circles. The ETS adaptation prepared for the Head Start Planned Variation study has eliminated this difficulty.

The total score alpha reliabilities of .85 for Enumeration I and .77 for Enumeration II showed the test to be relatively internally consistent in both years. The "Order" items (items 17-21) of Enumeration II were found as a group to be quite difficult for the subjects, and their low internal consistency and correlation with total score suggest that they be deleted for use with this age sample. Huron Institute (Walker et al., 1973) makes the same recommendation.

The test was sensitive to sex, age and SES differences in both Years 1 and 2. Correlations with other measures were generally low in Year 1 but reached a high of .69 in Year 2. The correlations reflected both perceptual and general information-processing competence. There was no factor analytic evidence for a specific numerical skill underlying Enumeration performance; its relationship to the other numerical task in the battery, Spontaneous Numerical Correspondence, was not strong. However, the Enumeration tasks included various aspects of quantitative functioning, such as counting and itemizing, which were not tapped by Spontaneous Numerical Correspondence, as well as assessing recognition in a non-concrete format of the same spatial order and amount. These qualities mean that Enumeration I and II more closely approximate the format of the paper and pencil group tests for the primary grades than does Spontaneous Numerical Correspondence. Future data collection should provide information concerning the relative contribution of these various components of quantitative competencies.

References

Almy, M. Young children's thinking. New York: Teachers College Press, 1966.

Bissell, J. S. Planned variation in Head Start and Follow Through. Washington, D. C.: U. S. Department of Health, Education, and Welfare, 1972.

Educational Testing Service. Preschool Inventory, Revised Edition. Princeton, N. J.: ETS, 1970.

Madden, R. & Gardner, E. F. Stanford Early School Achievement Test: Level I. New York: Harcourt Brace Jovanovich, 1969.

Moss, M. H. Tests of Basic Experiences. Monterey, Cal.: CTB/McGraw-Hill, 1970.

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

Rothenberg, B. B. & Courtney, R. G. Conservation of number in very young children. Developmental Psychology, 1969, 1, 493-502.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Siegel, L. S. The sequence of development of certain number concepts in preschool children. Developmental Psychology, 1971, 5, 357-361.

Walker, D. R., Bane, M. & Bryk, T. The quality of the Head Start Planned Variation data. Cambridge, Mass.: Huron Institute, 1973, in press.

Washington, E. D. & Teska, J. A. Correlations between the Wide Range Achievement Test, the California Achievement Tests, the Stanford Binet, and the Illinois Test of Psycholinguistic Abilities. Psychological Reports, 1970, 26, 291-294.

Wohlwill, J. A study of the development of the number concept by scalogram analysis. Journal of Genetic Psychology, 1960, 97, 345-377.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

YOUNG CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

A LONGITUDINAL STUDY

Technical Report 5

ETS Matched Pictures Language Comprehension Task I and II

Judith A. Meissner
Virginia C. Shipman
Anne M. Bussis

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

ETS Matched Pictures Language Comprehension Task I and II

Background

It is only within the recent past and with the aid of new linguistic theory that psychology has begun to gain insights into the nature of the language-unfolding process. A most striking feature of this phenomenon is that children apparently learn the underlying grammatical rules and structure of a language, this knowledge enabling them eventually to understand and generate an infinite number of well-formed sentences in their native language. It was Chomsky (1957, 1964, 1965) who first emphasized the importance of language as a system structured by interlinked transformational rules, and the recent work of many developmental psycholinguists (Bloom, 1970; Brown, 1970; McNeill, 1970; Menyuk, 1969) emphasizes grammatical structure (and not vocabulary acquisition) as constituting the heart of language learning.

In order to tap a child's comprehension of certain grammatical contrasts in a structured and systematic way, Berko (1958) developed a test format with sets of two pictures to illustrate the contrast. Using nonsense words, she studied the development of such grammatical features as noun plurals, past tenses and comparative forms of adjectives in four- to six-year-old subjects, finding that the majority could produce most of the correct grammatical endings. A similar form of this test was developed independently in 1957 by Bogoyavlenskiy (in Slobin, 1968) and was used with Russian children. Fraser, Bellugi and Brown (1963) used picture pairs to study the relative difficulty of comprehension, production and imitation of grammatical contrasts in normal three-year-olds, and Lovell and Dixon (1967) replicated the study with a group of six- and seven-year-old retarded children. Both studies found that imitation was an easier task than comprehension, and that comprehension was easier than production. In both studies, the affirmative-negative contrast was the easiest for subjects, and past tenses and direct object-indirect object distinctions were among the most difficult. Parisi (1971) also used this paired-pictures technique with a group of three- to six-year-old Italian children and found the same order of difficulty of grammatical features. The version of matched pictures used in the ETS-Head Start Longitudinal Study was originally developed for a 1967 study of 4- and 5-year-old children in New York City (Bussis, 1968), in which no sex and few age differences were found in ability to distinguish grammatical features. The findings of the order of difficulty of the grammatical contrasts of this study were in accord with that found in all previous studies.

The study of grammatical comprehension is of particular importance in regard to disadvantaged children, since some recent preschool programs (Bereiter & Engelmann, 1966; Engelmann, 1970; Osborn, 1968) have been based on the assumption that "the language of culturally deprived children . . . is not merely an underdeveloped version of Standard English, but is basically a non-logical mode of expressive behavior [Bereiter et al., 1966, p. 113]." These educators claim that disadvantaged preschool children do not use or understand Standard English prepositions, negations or conjunctions. A recent study by Torrey (1969) used a picture pair test to study second-grade Afro-American children's comprehension and production of Standard English plurals, verb endings, possessives and contractions and found that many children could both comprehend and produce Standard English forms which they rarely or never had been heard to use spontaneously. Unfortunately, the study did not include any of the prepositions, negations or conjunctions which Bereiter, Engelmann and Osborn claimed disadvantaged children not to know. Torrey's study gives important evidence that a picture format can be used profitably to study the development of the comprehension of grammatical contrasts in these children, and it is the purpose of the Longitudinal Study to follow the development of grammatical contrasts in a disadvantaged sample over a six-year period of time.

Task Description and Administration

The task materials for Matched Pictures I and II consist of a set of cards, each card containing a pair of black and white pictures. Both pictures in a pair contain similar stimulus elements, but they depict different relationships between the elements. The child's task is to distinguish which relationship a particular word implies and to point to that picture. For example, the child is shown a pair of pictures and told that they are called "Bear is sitting" and "Bear is not sitting," without E indicating which title goes with which picture. The child is then asked to point to the picture called "Bear is not sitting." The task has a counterbalanced design for the position of the "correct" picture (i.e., right or left side of the card) and the sequence in which E names the correct picture title (first or second). There are two practice items at the beginning of the task to ensure that the child understands the response procedure of pointing to the picture asked for.

(a) Matched Pictures I. In the Year 1 version of the task there were 20 picture pairs divided into four syntax types: Future Tense, Past Tense, Negation and Prepositions. The items for these four grammatical types are given on the following page, with the target picture of each pair preceded by an asterisk.

Items of Matched Pictures I

Future Tense (4 items)	Item # in the Test
*The cat will drink. The cat is drinking.	1
The cup is falling. *The cup will fall.	3
Mrs. Mouse is sweeping. *Mrs. Mouse will sweep.	5
The dog is eating. *The dog will eat.	7
Past Tense (4 items)	
*The frog jumped. The frog is jumping.	2
The match is burning. *The match burned.	4
*The dog swam. The dog is swimming.	6
*The cat ran. The cat is running.	8
Negation (6 items)	
*The mouse is not reading. The mouse is reading.	9
*The bear is not sitting. The bear is sitting.	10
The dog with a bone; *the dog with no bone	11
The cake with the candles; *the cake with no candles	12
*The cat is not smiling. The cat is smiling.	13
The basket is empty. *The basket is not empty.	14

Prepositions (6 items)

The procedure was changed slightly here, in that the child was not told the names of both pictures, but was asked only to point to the one named by the examiner. The picture called for in each item is given below (with the contrasting preposition depicted in the paired picture indicated in parentheses).

The cat under the chair (on the chair)	15
The bird in the cage (out of the cage)	16
The rabbit behind the tree (in front of the tree)	17
The ball on the table (under the table)	18
The dog beside the box (in the box)	19
The stick between the monkey's feet (under....feet)	20

Administration time for Matched Pictures I is approximately 6 to 8 minutes.

(b) Matched Pictures II. In the Year 2 version of the Matched Pictures Test the Preposition items were deleted, since the overall proportion of subjects passing these items in Year 1 was very high (over 90% passing for 2 of the 6 items), and six comparative adjectives were substituted, four single comparatives and two coordinate comparatives. In addition, to check on the Year 1 assumption that the present progressive tense was well-established, four of the previously-used verb items which were presented earlier in the test were repeated at the end of the test, with the children at this time asked to point to the present tense picture. These four Present Tense items were scored separately and were not included in the total score. With the addition of these new items, the test took about 10 minutes to administer. The items for each of the new subtests of Matched Pictures II are given below, with the target picture designated by an asterisk.

Future Tense (4 items) Same items and item numbers as Matched Pictures I

Past Tense (4 items) " " " " " " " " "

Negation (6 items) " " " " " " " " "

Single Comparatives (4 items)	Item #. in the Test
Less milk. *More milk.	15
More candy. *Less candy.	16
*Same hats. Different hats.	17
Same size giraffes. *Different size giraffes.	18

Coordinate Comparatives (2 items)	
Less fish but bigger (more fish but smaller)	19
More snakes but shorter (less snakes but longer)	20

Present Tense (4 items)	# in Test	Is Reversal of Item
The cat will drink. *The cat is drinking.	21	1
*Mrs. Mouse is sweeping. Mrs. Mouse will sweep.	22	5
The frog jumped. *The frog is jumping.	23	2
*The cat is running. The cat ran.	24	8

(Present Tense items are not included in Total Score.)

Matched Pictures I and II are relatively simple to administer. The tester must be careful not to point at or look at a picture while describing it, however. The test was generally enjoyed by these subjects.

Scoring

Each item on the test was scored right, wrong, refusal, multiple answer or indeterminate with one point given for each right answer. Therefore, the maximum score possible was 20 for both Matched Pictures I and II.

Score Properties

(a) Matched Pictures I. Table 1 presents item difficulties and biserial correlations of the items with the total score for Matched Pictures I. The items ranged widely in difficulty, the hardest item (item 6: "swam") being passed by only 29.9% of the subjects, and the easiest item (item 15: "under") by 93%. The item difficulties are very similar to those found by Bussis (1968) with her low SES four-year-olds, with the exception of the Negation items, which were considerably easier for Bussis' sample. The biserial correlations with the total score show that the Past Tense items (items 2, 4, 6 and 8) as a group had the lowest biserials (ranging from .20 to .39) and Prepositions (items 15-20) and Negations (items 9-14) had the highest, with biserials ranging from .46 to .58 and .39 to .63, respectively.

Table 2 presents the intercorrelations among subscores and coefficient alpha reliabilities for Matched Pictures I. Subscore intercorrelations were quite low, ranging from .02 for Negation with Past Tense to a high of .40 for Negation with Prepositions. Uncorrected correlations of subscores with total score were of only moderate magnitude, ranging from .46 (Past Tense) to .71 (Negation). The total score's coefficient alpha of .57 showed a moderate amount of internal consistency. Subscore coefficient alphas, which ranged from .12 to .50 were probably depressed because of the small number of items comprising each subscore. In subsequent analyses, only the total score was used.

(b) Matched Pictures II. Table 3 presents the item difficulties and biserial correlations of the total score for Matched Pictures II. Percent correct for individual items ranged from 22.3 (item 2: "jumped") to 92.9

Table 1

Percent Passing Each Item and Biserial Correlations*
for Matched Pictures I Total Score (N = 967)

Item Number	Item Description	Percent Passing	r Biserial
1	will drink	40.5	0.37
2	jumped	33.4	0.31
3	will fall	45.4	0.47
4	burned	47.2	0.39
5	will sweep	46.4	0.44
6	swam	29.9	0.37
7	will eat	48.4	0.47
8	ran	51.9	0.20
9	not reading	71.7	0.63
10	not sitting	81.1	0.54
11	no bone	75.2	0.60
12	no candles	79.0	0.54
13	not smiling	68.9	0.45
14	not empty	70.5	0.39
15	under the chair	93.0	0.58
16	in the cage	90.4	0.52
17	behind the tree	75.3	0.50
18	on the table	78.5	0.46
19	beside the box	61.4	0.52
20	between the monkey's feet	68.7	0.49

*Correlation of each item with the total score.

Table 2

Subscore Intercorrelations and Reliabilities*
for Matched Pictures I (N = 1460)

	Future Tense	Past Tense	Negation	Preposition	Total Score
Reliability*	(.31)	(.12)	(.50)	(.44)	(.57)
Future Tense (4 items)		.24	.19	.12	.61
Past Tense (4 items)			.02	.04	.46
Negation (6 items)				.40	.71
Preposition (6 items)					.66

*Coefficient alpha.

(item 10: "not sitting"), and biserials with total score ranged from .21 (item 8: "ran") to .76 (item 12: "no candles"). Here again, Bussis (1968) found very similar item difficulties for the low SES five-year-olds in her study, although the Negation items were passed by a considerably higher percentage of her subjects than by the present ones. The most difficult items as a group were those of Past Tense (items 2, 4, 6 and 8) and Coordinate Comparatives (items 19 and 20). The Present Tense and Negation

Table 3

Percent Passing Each Item and Biserial Correlations* for Matched Pictures II Total Score (N = 1308)

Item Number	Item Description	Percent Passing	r Biserial
1	will drink	38.1	0.47
2	jumped	22.3	0.34
3	will fall	60.6	0.54
4	burned	54.7	0.44
5	will sweep	58.1	0.60
6	swam	28.3	0.44
7	will eat	57.6	0.63
8	ran	44.3	0.21
9	not reading	87.5	0.63
10	not sitting	92.9	0.70
11	no bone	87.5	0.64
12	no candles	89.9	0.76
13	not smiling	81.0	0.52
14	not empty	75.4	0.37
15	more	85.7	0.56
16	less	66.1	0.57
17	same	66.1	0.58
18	different	71.0	0.52
19	less but bigger	26.8	0.33
20	more but shorter	28.5	0.32
21	is drinking	92.4	0.55
22	is sweeping	88.8	0.40
23	is jumping	84.0	0.26
24	is running	77.4	0.38

*Correlation of each item with the total score.

items were easiest for this sample at this time. These findings are in direct concurrence with all previous syntactic comprehension literature (Berko, 1958; Bussis, 1968; Fraser, Bellugi & Brown, 1963; Lovell & Dixon, 1967; Parisi, 1971).

Table 4 shows that subscore intercorrelations were again low in Year 2, ranging from .01 for Negation with Past Tense to .36 for Negation with Single Comparatives. The uncorrected correlations of the subscores with the total score ranged from .35 (Coordinate Comparatives with Total) to .72 (Future Tense with Total). The Total Score alpha reliability of .66 showed moderately high internal consistency, although the alphas for most of the subscores were much lower, ranging from .15 for Coordinate Comparatives to .55 for Negations.

Table 4

Subscore Intercorrelations and Alpha Reliabilities for Matched Pictures II (N = 1312)

	Future	Past	Score Negation	Single Comps.	Coord. Comps.	Total
Alpha	(.54)	(.31)	(.55)	(.50)	(.15)	(.66)
Future (4 items)		.30	.26	.26	.10	.72
Past (4 items)			.01	.13	.16	.54
Negation (6 items)				.36	.05	.61
Single Comps. (4 items)					.08	.65
Coord. Comps. (2 items)						.35

<u>Sample Performance</u>

The distribution of total correct scores for an N of 1435 subjects on Matched Pictures I was approximately normal, with a mean of 12.78, a median of 12.83 and a standard deviation of 2.94. Table 5 shows the distribution of total score by sex and by three-month age groups. It can be seen in this table that performance on the task tended to improve with each successive age group and that girls obtained slightly higher scores than did boys. An analysis of variance using only the longitudinal subjects (i.e., those subjects for whom data were available in both Years 1 and 2) found marginally significant sex differences ($F = 4.24$, $df = 1/1096$, $p < .05$) favoring girls, and highly significant age* differences ($F = 16.11$, $df = 1/1096$, $p < .001$) favoring the older children and SES differences** ($F = 38.63$, $df = 2/1096$, $p < .001$) favoring the children whose mothers had the most schooling.

Table 5

Distribution of Total Score by Age and Sex for Matched Pictures I

Group	N	Mean	SD	10	25	50	75	90
42-44 mo.	88	12.24	2.86	8.33	10.58	12.23	14.30	15.94
45-47 mo.	301	12.57	3.16	8.34	10.62	12.69	14.83	16.56
48-50 mo.	335	12.50	2.91	8.95	10.52	12.53	14.38	16.35
51-53 mo.	382	12.95	2.96	9.12	10.95	13.06	15.08	16.77
54-56 mo.	270	13.16	2.67	9.60	11.22	13.19	15.09	16.80
57-59 mo.	59	13.51	2.75	9.98	11.59	13.25	15.55	17.32
Boys	753	12.56	3.01	8.76	10.68	12.72	14.81	16.69
Girls	682	12.92	2.86	9.18	10.99	12.97	14.91	16.57
Total	1435	12.78	2.94	8.98	12.82	12.83	14.86	16.64

Note.--Scores ranged from 0 to 20.

*Subjects divided at the median into a "younger" and an "older" group.

**Subjects divided into three groups: those whose mothers had more than 12 years of schooling, less than 12 but more than 10 years of schooling, and less than 10 years.

The distribution of total correct scores for an N of 1312 subjects on Matched Pictures II had a mean of 12.22, and a standard deviation of 3.15. The distribution of this total score is presented in Table 6, grouped by three-month age divisions and by sex. An analysis of variance using only longitudinal subjects showed significant effects for all three variables of age ($F = 6.76$, $df = 1/1129$, $p < .01$), sex ($F = 12.95$, $df = 1/1129$, $p < .001$) and SES ($F = 76.16$, $df = 2/1129$, $p < .001$), with older children, girls and children whose mothers had more schooling obtaining the higher scores. In absolute value, however, these mean differences were very small in both years, all differences except those of SES being approximately a half point of total score. This is perhaps a major reason why the results of other grammatical comprehension studies are somewhat contradictory, Parisi (1971) finding no sex differences in performance and Bussis (1968) finding neither significant sex nor age differences.

Table 6

Distribution of Total Score by Age and Sex for Matched Pictures II

Group	N	Mean	SD	10	25	50	75	90
51-53 mo.	83	11.90	2.97	8.16	10.27	11.71	13.92	15.84
54-56 mo.	313	11.83	3.18	7.69	9.57	11.81	13.92	16.06
57-59 mo.	304	12.14	3.15	8.02	9.96	12.18	14.14	16.28
60-62 mo.	351	12.32	3.17	8.33	10.08	12.23	14.40	16.56
63-65 mo.	245	12.77	3.10	8.65	10.59	12.79	14.86	17.18
66-69 mo.	16	12.12	3.01	8.10	10.50	11.83	14.17	15.90
Boys	697	11.89	3.14	7.75	9.64	11.94	13.96	16.08
Girls	615	12.59	3.13	8.68	10.51	12.61	14.64	16.85
Total	1312	12.22	3.15	8.07	10.06	12.19	14.27	16.48

Note.--Scores ranged from 0 to 20.

The group of 14 items common to Years 1 and 2* was subjected to a repeated-measures analysis of variance, which found a significant increase in score between Years 1 and 2 for these common items ($F = 93.52$, $df = 1/1071$, $p < .001$). Significant sex differences favoring girls ($F = 7.83$, $df = 1/1067$, $p < .01$) and SES differences favoring children whose mothers had more schooling ($F = 61.67$, $df = 2/1067$, $p < .001$) were found across years. As was the case with both Matched Pictures I and II data, mean differences within all variables except SES and total within-year score were very small in absolute value, all being slightly less than a half point of total score.

Relationship with Other Measures

The correlations** of the Matched Pictures I total score with other tests in the Longitudinal Study battery were quite low, ranging from .00 to .36. The only two correlations above .30 were with Peabody Form A total score ($r = .36$) and Preschool Inventory (.35), both tests of general information and vocabulary knowledge. Correlations in the .20's were obtained with Seguin log fastest time, Children's Auditory Discrimination Inventory nonsense words score, Peabody Form B total score, Hess and Shipman Eight-Block Sorting Task total score, ETS Story Sequence I (Part I), Form Reproduction total score, and Johns Hopkins Perceptual Task total score. The factor analytic data (see Shipman, 1971) revealed Matched Pictures I to load most highly on the general intellectual functioning factor, with all other loadings considerably below .30.

Matched Pictures II had a much greater number of correlations above .30 with other tests in the battery than did Matched Pictures I, with the range of correlations from .01 to .59. Here again, the highest correlations were with general information and vocabulary tests: Preschool Inventory total

*Future Tense (4 items), Past Tense (4 items) and Negation (6 items).
**These correlations were based on the longitudinal subjects' data only.

score (.59), TAMA General Knowledge Test (.47), Peabody Form A and B total scores (both .44), and Massad Mimicry Meaningful Words: Final Sounds (.43). Perceptual tests such as Form Reproduction, Johns Hopkins Perceptual Test and Matching Familiar Figures (Errors) also correlated quite highly (.39, .40 and -.41, respectively), as did Enumeration II Counting items (.38), CADI nonsense words (.41) and Story Sequence II (Part I) (.[illegible]). As in Year 1, the highest factor loading was on the general information-processing factor (see Shipman, 1972).

The relatively high correlations with the informational and vocabulary tests reflect environmental influences upon Matched Pictures performance, as do the highly significant SES differences in both years. A perceptual component is also inherent in this task, as shown by its relatively high correlations with perceptual tests. The relatively low correlation of Matched Pictures with the other, more specific language measure, Story Sequence (Part I), is most likely a result of differences in task requirements between these two measures. The Story Sequence task, in which the child is presented with a set of completely different pictures which he must place in the correct sequence to illustrate the story told by the tester, requires a familiarity with time and spatial sequences which is not necessary for success on Matched Pictures.

Summary

The ETS Matched Pictures test was included in the present study to measure children's comprehension of certain grammatical features, such as past and future tenses, negation and prepositions. Since the test required only a pointing response, thereby placing no demands for verbal production on the subjects, it may be considered a more accurate index of verbal comprehension than tasks in this or other test batteries which do involve such demands. In accordance with previous research (Pussis, 1968; Fraser, Bellugi & Brown, 1963; Lovell & Dixon, 1967; Parisi, 1971), and contrary to the claims of some current educators (Bereiter & Engelmann, 1966; Engelmann, 1970; Osborn, 1968), negation and preposition items were found to be quite easy for this predominantly economically disadvantaged sample during this age period, and future and past tense items were somewhat harder. The two coordinate comparative items introduced in Matched Pictures II were found to be extremely difficult for these subjects. Given the range in item difficulties found for this sample during this age period, a revised form which includes all items from Form I and Form II has been developed for use with similar-aged children. It is recommended, however, that in future task development the two-choice format of the test items be altered, since the present format allows a 50% chance of the child's being correct on any item. More preferable would be a multiple-choice format such as that being used in the ETS CIRCUS battery for young children (ETS, 1972) or a four-choice format such as that used in the Northwestern Syntax Screening Test (Lee, 1969).

Both the Year 1 and Year 2 versions of Matched Pictures and the Year 1-Year 2 common items analysis were sensitive to differences in SES, age and sex. The sex differences favoring girls could be a result of different environmental experiences of the sexes in their preschool years or also to a greater willingness of the girls to cooperate with and please the tester, some evidence for which emerged in other test measures. Since the Matched Pictures Test used only standard English grammatical forms, and many of the children in the present sample, especially those in the low

SES classification, may speak a dialect only superficially related to standard English, it is not known to what extent the significant SES effects obtained reflect differences in language usage rather than in syntactic comprehension.

Intercorrelations with other tests in the battery and factor loadings were considerably higher with Matched Pictures II than with Matched Pictures I, giving some evidence of greater cognitive-perceptual integration in Year 2. The intercorrelations suggested general information, vocabulary and perceptual components present in Matched Pictures, although its relationship to more specific verbal skills was less evident. Future data will provide further understanding of the developmental pattern of syntactic comprehension and the relationship of this pattern to family and school determinants. Analysis of the mother's verbalizations in the interaction tasks will provide an index of the restriction of the child's linguistic environment which then can be related to the various measures of linguistic competence used in the study.

References

Bereiter, C. & Engelmann, S. Teaching disadvantaged children in the preschool. Englewood Cliffs, N. J.: Prentice-Hall, 1966.

Bereiter, C., Engelmann, S., Osborn, J. & Reidford, P. A. An academically-oriented preschool for culturally deprived children. In F. M. Hechinger (Ed.), Preschool education today. New York: Doubleday, 1966. Pp. 105-135.

Berko, J. The child's learning of English morphology. Word, 1958, 14, 150-177.

Bloom, L. Language development: Form and function in emerging grammars. Cambridge, Mass.: Massachusetts Institute of Technology Press, 1970.

Brown, R. Psycholinguistics, New York: Free Press, 1970.

Bussis, A. Children's understanding of basic language structure. In R. S. Melton, R. Charlesworth, M. N. Tanaka, B. B. Rothenberg, L. W. Pike, A. M. Bussis, & E. S. Gollin, Cognitive growth in preschool children. Princeton, N. J.: Educational Testing Service, 1968, RM-68-13.

Chomsky, N. Discussion of Miller and Ervin's paper. In U. Bellugi & R. Brown (Eds.), The acquisition of language. Monographs of the Society for Research in Child Development, 1964, 29 (No. 92). Pp. 35-39.

Chomsky, N. Aspects of the theory of syntax. Cambridge, Mass.: Massachusetts Institute of Technology, 1965.

Chomsky, N. Syntactic structures. The Hague: Mouton, 1957.

Educational Testing Service. CIRCUS Test 8: How Words Work. Princeton, N. J.: Educational Testing Service, 1972.

Engelmann, S. How to construct effective language programs for the poverty child. In F. Williams (Ed.) Language and poverty. Chicago: Markham, 1970. Pp. 102-122.

Fraser, C., Bellugi, U. & Brown, R. Control of grammar in imitation, comprehension and production. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 121-135.

Lee, L. L. Northwestern syntax screening test. Evanston, Ill.: Northwestern University, 1969.

Lovell, K. & Dixon, E. M. The growth of the control of grammar in imitation, comprehension and production. Journal of Child Psychology and Psychiatry, 1967, 8, 31-39.

McNeill, D. The acquisition of language: The study of developmental psycholinguistics. New York: Harper, 1970.

Menyuk, P. Sentences children use. Cambridge, Mass.: Massachusetts Institute of Technology Press, 1969.

Osborn, J. Teaching a teaching language to disadvantaged children. Monographs of the Society for Research in Child Development, 1968, 33, 36-48.

Parisi, D. Development of syntactic comprehension in preschool children as a function of socioeconomic level. Developmental Psychology, 1971, 5, 186-189.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Slobin, D. Imitation and grammatical development in children. In N. Endler, L. Boulter & H. Osser (Eds.), Contemporary issues in developmental psychology. New York: Holt, 1968. Pp. 437-443.

Torrey, J. Teaching standard English to speakers of other dialects. Paper presented at The Second International Congress of Applied Linguistics, Cambridge University, 1969.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

Technical Report 6

ETS Story Sequence I and II

Karla S. Goldman
Virginia C. Shipman
Masako N. Tanaka

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

ETS Story Sequence I and II

Background

Early research on language and cognition reported significant correlations between IQ and linguistic parameters such as sentence length, vocabulary, number of adjectives, etc., (McCarthy, 1954). Recently there has been a tendency to look at more process-oriented variables in studying interrelationships between cognitive and linguistic functioning (Blank & Frank, 1971). The ETS Story Sequence task was developed to study the child's understanding of larger units of language, including sentences, short sequences, and sequence relationships.

Language processes have been distinguished in terms of receptive and productive skills. Intelligence tests measuring school readiness, such as the Metropolitan Readiness Test (Hildreth, Griffiths & McGauvran, 1965), are composed almost entirely of receptive language items. But the close relationship of both types of measures with actual school performance suggests the advisability of assessing both receptive and productive language at early ages. There has been little research comparing receptive and productive language in the same child using the same set of stimulus materials. Many of the tests currently being used to measure receptive skills in young children employ a single word stimulus and ask the child to point to the corresponding picture, e.g., Peabody Picture Vocabulary Test (PPVT). Although this method is successful with young children who may be rather reticent in a testing situation, it does not tap productive skills and relies only on a single-word stimulus.

Instead of using a single-word stimulus, Blank and Frank (1971) employed stories, believed to represent a familiar situation and "also demand that an organized structure be imposed on the material if it is to be incorporated [p. 300]." Story retelling thus places more demands on an imitation task than simple recall or parroting. Two modes of story presentation were used with lower-SES children (aged 4 to 6); one group was asked to repeat each sentence of the story as it was told and the other heard the entire story before participating. Each child then retold the story (production) and answered questions about it (semantic performance). Children who were instructed to repeat each sentence had superior recall ($p < .05$) and grasp of story themes ($p < .10$) than the group who did not. The authors suggest this is due to the more active role played by the sentence-repeating subjects. The amount of semantic and syntactic information recalled was highly correlated (.72 and .76) for each type of presentation. For the sentence repeating group, IQ (Slosson Intelligence Test) was significantly correlated with syntactic recall and with semantic recall for the non-repeating group. Repeating sentences seemed to facilitate recall for children with lower MA. For both groups, Ss with lower MA gave significantly more incorrect semantic information.

Many of the studies revealing social class differences in language development have used the PPVT and other similar picture vocabulary tests (Deutsch, 1963; Gray & Klaus, 1965; Jeruchimowicz, Costello & Bagur, 1971; Rieber & Womack, 1968). Using the PPVT, Jeruchimowicz et al. (1971) found that lower-SES but not middle-SES preschool children made significantly more errors in action than object words. Although the lower-SES children had a significantly greater proportion of verb and noun errors than their middle-SES age-mates in receptive language, there were no SES differences in errors of concepts used in their productive language (story retelling). No sex differences were found on the receptive or productive measures.

Other variables which may affect the child's performance on a receptive language measure have been systematically assessed by the Illinois Test of Psycholinguistic Abilities. Weaver (1965) reported that culturally disadvantaged children showed relative strength in visual-motor channels and relative weakness in auditory-vocal channels. Difficulty in auditory discrimination is suggested as another variable (John & Goldstein, 1964; Stern, 1966). With any measure involving oral presentation there is the additional problem in using standard English with children who are more familiar with non-standard English (Baratz, 1969). Recent studies have revealed the existence of a black dialect which is patterned and consistent (Stewart, 1970; Labov, 1970); yet most studies investigating SES differences in language ability have ignored the confounding of SES and race and differences in language usage.

Assessment of productive language has usually involved measuring total output, sentence length, and vocabulary. A number of studies reviewed by Raph (1965) and by Cazden (1966) reported deficits in lower-SES children in these areas. From Labov's work one may conclude that using simpler and less verbose language may not be a deficit at all; rather the middle-class child may just be "wordy." Such differences may reflect sociocultural differences in usage rather than in comprehension. Milgram, Shore & Malasky (1971) analyzed linguistic categories (number of words; number of sentences) and thematic categories (story-relevant sentences; essential themes) used by kindergarten and first grade children retelling an illustrated story which was read aloud to them. Although advantaged Ss surpassed their disadvantaged peers on all four measures, the differences were greater for the linguistic than the thematic categories. The striking difference was in sheer verbal output rather than in retelling the essential features of the story. No sex or race differences were found.

Productive language includes dimensions of oral and written ability but by necessity only the oral dimension is applicable for young children. Yet these different aspects are related to each other. Loban's (1965) longitudinal study of children (K-12) provided considerable evidence of the continued relationship between oral ability and reading and writing abilities. He found that those skilled in one area were skilled in all three areas of language. Bruce (1970), however, found Head Start children and older problem readers to be good at everyday discussion and story telling, but the older ones "read" sentences as a series of individual words rather than understanding the syntax rules that relate words in sentences. Bruce also found Head Start children to feel alien to abstract graphic forms depicting words, but able to understand and feel comfortable with pictures depicting words.

Thus, it would appear from the above studies that simultaneous assessment of receptive and productive skills may provide important clues to understanding language development and that the use of words in the context of a phrase, sentence or story may provide a more realistic measure of the child's functional understanding of language than the use of a single stimulus alone.

Task Description and Administration

The ETS Story Sequence task was designed to assess the young child's understanding and use of language in story sequence using receptive and productive skills. The stimulus materials are black and white cartoon-type drawings on 5 1/2" x 4 1/4" cards. Story Sequence I used in Years 1 and 2 of the study consists of two practice items (3 cards each) and two test items (3 cards and 4 cards). Story Sequence II, used in Years 2 and 3, consists of four items (two use 3 cards each, two use 4 cards each). There is no apparent sequence in any of the picture sets, but rather, the sequence is provided by verbal cues in the stories presented by the tester (I, receptive), or in the story created by the child (II, productive). The stories were written especially for this task in order to avoid the problem of differential familiarity. Because of the difficulty of balancing the distribution of sex, race, and situations in a small number of items, animals were used as characters in the stories.

There are three types of items:

1. Receptive language. The child selects and arranges a card sequence while listening to a story told by the tester. There is no inherent order in the pictured situations and the child is dependent on linguistic cues provided in the story (used in Story Sequence I).
2. Productive language using verbal recall. The tester presents the cards in order as she tells the story and the child is then asked to retell the "same" story (used in Story Sequence II).
3. Productive language using child's story telling. The child chooses the picture cards from an array and tells his own story about them (used in Story Sequence II).

Two practice items are given prior to Story Sequence I to familiarize the child with the idea of physically placing pictures in a left-to-right row, and to give practice in selecting the appropriate sequence from the array of cards. For the first practice item the tester turns pages of "book" made of three cards clipped together telling a simple story of one sentence per card. Then, separating the cards so that "we can see the pictures better," the tester places cards in a left-to-right sequence on a board in front of the child as the story is told. The cards are collected in the same order and the child is then asked to replace them the same way. Errors are corrected. For the second practice item the tester places the same three cards in a random array and asks the child to select each corresponding card as the story is being told. Thus, the first instructional item familiarizes the child with physically ordering the pictures, and the second gives practice in selecting the appropriate card from the array when the tester pauses in the story.

Following practice, the two test items of Story Sequence I are presented. The child is required to choose the appropriate pictures while listening to the story. In both test items there is no replacement of cards in the array as the child makes his choices; thus the array diminishes with each response. The child may, however, decide to reuse a card he has already chosen rather than one of the cards remaining in the array. The decision to avoid replacement of cards was based on pretesting experience which suggested that such a procedure would be confusing to the child. After the child has finished with each item, the number on the back of each card in his sequence is recorded. This number identifies the correct position of each card in his sequence (e.g., a record of 1, 2, 4, 3 means that S chose the correct cards for the first two statements only).

Story Sequence II has two parts with two items in each. The first part is story recall. The child is asked to listen carefully to a story because he will be asked to repeat the same story afterward. As the tester tells the story, a card accompanying each statement is placed in order in front of the child, who is then asked to repeat or retell the story. The second part is story-telling, tapping the child's ability to make up a story using sets of picture cards. Each set of cards is placed in front of the child who is asked to use them to make up a story. The instructions encourage the child to choose as many of the pictures as desired and to arrange them in any order. The child's story is taped.

Average administration time is 15 minutes. During training the tester must learn to use a tape recorder and must become familiar and comfortable with the stories. Since many young children require considerable encouragement to tell a story, the tester's ability to establish rapport and a relaxed setting is critical. Testers reported that most children greatly enjoyed this task.

Scoring

Story Sequence I (receptive language), used alone in Year 1 and in conjunction with II in Year 2, is composed of two items, each with its own story: Tommy Kitten (3 cards) and Timothy Mouse (4 cards). Each correctly selected card is given one point so the possible score range for the first item is 0-3, for the second, 0-4, and for the combined score, 0-7.

Story Sequence II, used in conjunction with Part I in the second year, is composed of four items: 1. Susie Bear (3 cards); 2. Rabbit (3 cards); 3. Mr. Turtle (4 cards); and 4. Mrs. Turtle (4 cards). For each of these four items a card order score (the number of cards placed in the correct order) and a fluency score (total number of words used) were computed. For items 1 and 3 (story recall) a language order score was computed along with eight additional scores*, which are not reported here since they proved to be redundant and/or unreliable. Language order was scored on a scale of 0-5, that is, from no time sequencing phrase (0) to the exact sequence phrase used for each card (5).

All scoring was done at the ETS Princeton Office. Although Part I presented only routine scoring problems, Part II required considerable repeated revision of the coding manual and intensive training of scorers.

Score Properties

For Story Sequence I total score reliability (coefficient alpha) in Year 1 was .50; in Year 2, .37. The correlation between Year 1 and Year 2 scores was only .28, in part due to the increased restriction in range of scores as a ceiling effect became apparent.

Reliability estimates obtained in Year 2 for Story Sequence II on items 1 and 3 (story recall) were .63 for card order, .65 for fluency and .67 for language order; on items 2 and 4 (story telling), coefficients of .72 for card order and .70 for fluency were obtained.

Sample Performance

Table 1 reports means and standard deviations by three-month age intervals and sex. The composite mean score for Story Sequence I was 4.3 (SD = 2.25) in Year 1 and 5.5 (SD = 1.86) in Year 2. The potential range of scores, 0-7, was found at each age level in Year 1, but the lower limit was not found in all age groups in Year 2. Except for the oldest age group, performance showed a regular trend with age in Year 1; this

*Noun number, noun quality, verb number, verb quality, depicted verbs, non-depicted verbs, content words, and non-content words.

Table 1

Means and Standard Deviations by Age and Sex: Part I, items 1 and 2*

Group	N	Mean	SD
Year 1			
42-44 mo.	89	3.9	2.41
45-47 mo.	315	3.9	2.19
48-50 mo.	331	4.1	2.26
51-53 mo.	383	4.5	2.25
54-56 mo.	270	4.7	2.19
57-59 mo.	60	4.6	2.29
Boys	770	4.1	2.28
Girls	678	4.4	2.20
Total	1448	4.3	2.25
Year 2			
51-53 mo.	64	5.3	2.00
54-56 mo.	209	5.2	1.87
57-59 mo.	191	5.4	2.02
60-62 mo.	239	5.8	1.72
63-65 mo.	151	5.6	1.74
66-69 mo.	8	5.6	1.92
Boys	450	5.3	1.94
Girls	412	5.7	1.76
Total	862	5.5	1.86

*Range = 0-7.

trend was less pronounced, however, in Year 2 as a ceiling effect became apparent. The composite score for girls (4.4 and 5.7) was only slightly higher than for boys (4.1, 5.3).

The item analyses in Table 2 show similar trends, with girls performing better than boys and scores generally improving with age. There was some difference in the proportion of children who had perfect score for item 1 and for item 2, (41.7% and 49.7% in Year 1 and 67.4% and 68.5% in Year 2, respectively). Perfect scores, as was expected, increased in the older age groups. If the six age groups in Year 1 are divided at the median (older vs. younger), each group shows a gain in the percentage of perfect scores from item 1 to item 2. That is, the younger group goes from 36.8% to 44.3% and the older group goes from 48.5% to 57.0%. This suggests that some children are learning from item to item, and perhaps, for this age range, more practice items prior to administration would be in order. This may also mean that even though item 2 has one more card (and sentence) than item 1, it is an easier item.

Further examination of Table 2 shows that children did make use of the option to reuse a card already chosen. A few children in Year 1 (2.8%) and in Year 2 (2.3%) did get a score of "2" on item 1 showing that this option was used. The same was true for item 2, with 2.2% in Year 1 and 1.9% in Year 2 scoring "3" out of a possible "4". There were very few refusals and indeterminate responses on these two items (Table 2, category 7). As might be expected, such responses diminished with age.

Story Sequence II, given only in Year 2, assessed productive language. The means and standard deviations by three-month age intervals and sex for these items are given in Table 3.

Table [illegible]

Means, Standard Deviations and Percent Response by Age and Sex

Group	N	Mean	SD	N	0	1	2	3	4	7*
					Percentage Response					
Year 1										
Item 1: Tommy Kitten (range 0-3)										
42-44 mo.	89	1.5	1.19	89	22.5	39.3	3.4	34.8		0.0
45-47 mo.	312	1.6	1.16	318	17.3	39.0	4.4	37.4		1.9
48-50 mo.	326	1.6	1.22	336	21.7	34.8	2.4	38.1		3.0
51-53 mo.	380	1.8	1.20	386	17.1	34.5	2.1	44.8		1.6
54-56 mo.	268	1.8	1.24	270	19.6	28.9	2.6	48.1		0.4
57-59 mo.	60	1.7	1.29	60	25.0	26.7	1.7	46.7		0.0
Boys	759	1.6	1.22	777	21.4	33.5	3.1	39.8		2.3
Girls	676	1.7	1.19	682	17.0	35.6	2.5	44.0		0.7
Total	1435	1.7	1.21	1459	19.3	34.5	2.8	41.7		1.6
Item 2: Timothy House (range 0-4)										
42-44 mo.	87	2.4	1.62	89	16.9	19.1	14.6	1.1	46.1	2.2
45-47 mo.	309	2.3	1.52	315	14.6	21.0	21.9	2.2	38.4	1.9
48-50 mo.	324	2.6	1.47	331	8.8	20.2	19.6	1.8	47.4	2.1
51-53 mo.	379	2.8	1.47	383	8.9	16.7	17.2	1.8	54.3	1.0
54-56 mo.	268	2.9	1.43	270	8.9	12.2	15.9	3.7	58.5	0.4
57-59 mo.	60	2.9	1.44	60	8.3	13.3	18.3	1.7	58.3	0.0
Boys	752	2.6	1.50	770	10.9	17.7	19.9	2.6	46.6	2.2
Girls	675	2.7	1.50	678	10.2	17.6	16.8	1.8	53.2	0.4
Total	1427	2.6	1.50	1448	10.6	17.6	18.4	2.2	49.7	1.4
Year 2										
Item 1: Tommy Kitten (range 0-3)										
51-53 mo.	63	2.3	1.06	65	7.7	20.0	1.5	67.7		1.5
54-56 mo.	209	2.2	1.11	211	9.0	27.0	1.9	61.1		0.5
57-59 mo.	191	2.3	1.06	192	6.3	27.1	1.6	64.6		0.0
60-62 mo.	239	2.5	0.96	243	6.2	14.8	2.5	74.9		0.0
63-65 mo.	151	2.3	1.01	152	5.3	23.7	3.3	67.1		0.0
66-69 mo.	8	2.9	0.35	9	0.0	0.0	11.1	77.8		0.0
Boys	449	2.3	1.05	455	7.0	23.5	2.4	65.7		0.4
Girls	412	2.4	1.02	417	6.5	20.9	2.2	69.3		0.0
Total	861	2.3	1.04	872	6.8	22.2	2.3	67.4		0.2
Item 2: Timothy Mouse (range 0-4)										
51-53 mo.	64	3.1	1.40	65	6.2	15.4	7.7	3.1	66.2	1.5
54-56 mo.	207	3.1	1.31	211	5.7	9.5	17.5	2.8	62.6	0.9
57-59 mo.	191	3.2	1.38	192	9.9	5.2	12.5	2.1	69.8	0.0
60-62 mo.	239	3.3	1.22	243	4.9	5.8	15.6	1.6	70.4	0.0
63-65 mo.	151	3.3	1.24	152	5.9	5.3	13.8	0.7	73.7	0.0
66-69 mo.	8	2.8	1.75	9	11.1	22.2	0.0	0.0	55.6	0.0
Boys	447	3.1	1.37	455	8.4	7.0	16.9	2.2	63.7	0.7
Girls	413	3.3	1.22	417	4.6	7.7	11.5	1.7	73.6	0.0
Total	860	3.2	1.30	872	6.5	7.3	14.3	1.9	68.5	0.3

fusals and indeterminate responses.

Table 3

Means and Standard Deviations by Age and Sex:
Story Sequence II (Year 2, Part II)

Group	Item 1 N	Mean	SD	Group	Item 3 N	Mean	SD
Card Order							
51-53 mo.	58	3.0	2.19	51-53 mo.	63	3.8	2.06
54-56 mo.	188	2.8	2.25	54-56 mo.	200	3.2	2.17
57-59 mo.	178	2.8	2.21	57-59 mo.	183	3.6	2.16
60-62 mo.	227	3.4	2.14	60-62 mo.	233	3.9	1.89
63-65 mo.	140	3.7	2.04	63-65 mo.	145	4.0	1.83
66-69 mo.	8	3.1	2.59	66-69 mo.	9	3.1	2.42
Boys	407	2.9	2.22	Boys	428	3.5	2.08
Girls	392	3.4	2.14	Girls	405	3.8	2.00
Total	799	3.1	2.20	Total	833	3.7	2.04
Fluency							
51-53 mo.	57	4.5	1.97	51-53 mo.	61	4.4	1.71
54-56 mo.	175	5.1	2.05	54-56 mo.	185	4.7	1.70
57-59 mo.	161	5.2	1.91	57-59 mo.	167	4.7	1.55
60-62 mo.	219	5.5	2.03	60-62 mo.	228	4.9	1.62
63-65 mo.	137	5.4	1.96	63-65 mo.	145	5.1	1.81
66-69 mo.	7	4.6	2.07	66-69 mo.	8	5.1	1.46
Boys	381	4.9	1.95	Boys	408	4.6	1.75
Girls	375	5.5	2.02	Girls	386	5.0	1.56
Total	756	5.2	2.01	Total	794	4.8	1.67
Language Order							
51-53 mo.	57	.9	1.05	51-53 mo.	61	1.1	1.04
54-56 mo.	175	1.1	1.27	54-56 mo.	185	1.1	1.12
57-59 mo.	162	1.1	1.26	57-59 mo.	167	1.2	1.09
60-62 mo.	219	1.3	1.34	60-62 mo.	227	1.3	1.17
63-65 mo.	137	1.5	1.37	63-65 mo.	145	1.3	1.13
66-69 mo.	7	1.3	1.80	66-69 mo.	8	1.6	1.30
Boys	382	1.1	1.26	Boys	407	1.2	1.13
Girls	375	1.4	1.34	Girls	386	1.3	1.12
Total	757	1.2	1.30	Total	793	1.2	1.13

Table 3 (Continued)

Means and Standard Deviations by Age and Sex:
Story Sequence II (Year 2, Part II)

Group	Item 2 N	Mean	SD	Group	Item 4 N	Mean	SD
Card Order							
51-53 mo.	62	3.4	2.14	51-53 mo.	62	3.6	2.07
54-56 mo.	197	3.4	2.22	54-56 mo.	190	3.4	2.17
57-59 mo.	167	3.5	2.18	57-59 mo.	165	3.5	2.15
60-62 mo.	222	4.0	1.93	60-62 mo.	211	3.9	1.98
63-65 mo.	139	4.0	1.90	63-65 mo.	141	4.0	1.96
66-69 mo.	9	2.9	2.52	66-69 mo.	9	2.2	2.64
Boys	412	3.5	2.18	Boys	406	3.5	2.19
Girls	384	3.9	1.97	Girls	372	3.9	1.94
Total	796	3.7	2.09	Total	778	3.7	2.09
Fluency							
51-53 mo.	51	28.8	21.66	51-53 mo.	52	45.6	30.27
54-56 mo.	169	25.7	19.30	54-56 mo.	169	38.8	22.36
57-59 mo.	147	27.6	23.28	57-59 mo.	147	39.2	25.83
60-62 mo.	203	28.7	24.88	60-62 mo.	204	37.6	23.79
63-65 mo.	122	27.0	22.18	63-65 mo.	121	40.8	38.87
66-69 mo.	7	19.1	11.20	66-69 mo.	7	31.1	18.98
Boys	354	24.4	21.74	Boys	354	34.9	21.75
Girls	345	30.4	22.80	Girls	346	43.8	31.81
Total	699	27.3	22.86	Total	700	39.3	27.54

Item 1 (using 3 cards) and item 3 (using 4 cards) both assess the child's ability to retell a story told by the tester and also to reproduce the original card order. The means generally increased with age and girls generally performed better than boys.

Items 2 (3 cards) and 4 (4 cards) also assess productive language, but here the child is asked to make up his own story. Again the means generally increased with age with the exception of the oldest age group. Given the extremely small cell size, and the fact that these are all children from one location, these particular data are not interpretable. Girls scored higher than boys, especially on fluency.

For Story Sequence I, given in both years of the study, a repeated-measures analysis of variance (age x sex x SES*) was performed on the longitudinal sample (those children tested in both Years 1 and 2). Summing the scores across years, sex differences favoring girls ($F = 7.33$, $df = 1/656$, $p < .01$), and SES differences ($F = 11.98$, $df = 2/656$, $p < .001$) favoring children whose mothers had more schooling were found.

Story Sequence II, given only in Year 2, was subjected to a separate age x sex x SES* ANOVA for the longitudinal sample. Significance levels for between-group differences for these variables are reported in Table 4. For all subscores a significant sex difference was found favoring girls. For language order (items 1+3) there was also a significant sex x SES interaction ($F = 4.40$, $df = 2/629$, $p < .02$), favoring girls in the middle SES group and boys in the lower and higher groups. Significant SES differences were found, favoring children whose mothers had more schooling, for three of the four subscores reported in Table 4, and a significant age trend (using a median-age split) was found for card order in items 1 and 3 ($F = 7.64$, $df = 1/633$, $p < .01$) and items 2 and 4 ($F = 6.27$, $df = 1/597$, $p < .02$).

Table 4

Analysis of Variance: Between-Group Differences on Story Sequence II

	Sex	Age	SES*
Items 1 + 3 (story recall)			
Card Order	$F = 8.38$, $p < .004$	$F = 7.64$, $p < .01$	$F = 6.35$, $p < .002$
Language Order	$F = 7.95$, $p < .005$	ns	$F = 11.94$, $p < .001$
Items 2 + 4 (story telling)			
Card Order	$F = 6.68$, $p < .01$	$F = 6.27$, $p < .02$	ns
Fluency	$F = 20.79$, $p < .001$	ns	$F = 4.02$, $p < .02$

*Mother's education (below 10th grade, 10th-12th grade, above 12th grade) was used as an index of SES.

These findings are in general agreement with those of Tanaka (1968), using this measure with preschool and kindergarten children in New York City. In that study of 4- and 5-year-olds, a significant SES difference ($p < .02$) was reported. Further, there was also an interaction of SES with age and sex ($p < .04$). For the middle-SES Ss, girls were superior to boys at both age levels, whereas low-SES boys performed better in the older group and there was no sex difference in the younger group.

Relationship with Other Measures

Among the main findings of a principal component factor analysis of the child test data was clear evidence for a general information-processing factor ("g") that accounted for most of the common variance among the cognitive tasks. In Year 1 for the longitudinal sample, eight out of the twenty-three tasks analyzed had varimax loadings of .50 or higher; Story Sequence I had a loading of .46. In Year 2, eleven of the tasks had loadings of .50 or higher; Story Sequence I had a loading of .48. Further analyses were conducted for the Year 1 total sample data in an effort to identify clusters of measures within this "g" factor (i.e., verbal, quantitative or perceptual items) and to delineate receptive and productive language skills within the subset of verbal measures. There was no evidence of such clusterings. It appears that at this age level for this population the child's performance on Part I of Story Sequence is more closely related to his general ability to process information than to a specific language component.

In the Year 2 factor extension matrix, card order and language order scores for items 1 and 3 (Story Sequence, Part II) correlated .44 and .50, respectively, with factor 1 in the 13-factor Varimax solution for the longitudinal sample. Card order and fluency scores for items 2 and 4 correlated .16 and .08, respectively, with factor 1. These data suggest that the ability to retell a story is closer to the cognitive information-processing skills tapped by factor 1 than is the ability to produce one's own story.

In Year 1, the highest correlations* between Story Sequence I and other tasks were moderate, starting with the Preschool Inventory (Caldwell) (r = .39) and followed by the two Peabody Picture Vocabulary Tests (PPVT): Form B - productive version (r = .36) and Form A - receptive version (r = .35). In Year 2, the highest correlation was again with the Preschool Inventory (r = .43), but correlations with TAMA (r = .36) and with Massad Mimicry, real words (r = .36) slightly surpassed PPVT, Form B (r = .35) and PPVT, Form A (r = .33).

Correlations obtained for Story Sequence I with the two perceptual picture-decoding tasks were similar within year, but were lower in Year 2: Year 1, Matching Familiar Figures error score (r = -.33), and Johns Hopkins Perceptual Test (r = .35); Year 2, Matching Familiar Figures error score (r = -.28) and Johns Hopkins Perceptual Test (r = .27). In Year 1 at least this pattern of correlations, with its similarity to the PPVT, suggests that performance on Story Sequence I was affected both by the child's single word vocabulary and his ability to "read" the perceptual information in the drawings used in the task. This inference may explain the low correlation between the Story Sequence I and ETS Matched Pictures scores (r = .22 in Year 1, r = .33 in Year 2). In view of the apparently similar nature of the tasks (i.e., the child selects a drawing in response to a sentence), this correlation was surprisingly low. However, a closer look at the stimulus drawings in the ETS Matched Pictures task shows that

---

*These correlations are based on data from the longitudinal Ss only; correlations for the total sample tested in Year 1 were slightly higher (.06 to .10).

the same elements (e.g., objects, animals, etc.) are re ated for the response choices, and the child is being asked to recognize the relationship between the elements (e.g., "dog beside the box" vs. "dog inside the box"). This is in contrast to the Story Sequence drawings which define the elements being asked of the child (e.g., Timothy Mouse "looking under the bed" vs. "sitting down in his chair"). In Year 1, this difference between the two tasks in regard to the use of perceptual cues is further suggested by the correlations of -.24 and .23 between ETS Matched Pictures and Matching Familiar Figures (error score) and the Johns Hopkins Perceptual Test, as compared to the Story Sequence correlations (r = -.33 and .35, respectively). In Year 2, however, ETS Matched Pictures correlated -.41 with Matching Familiar Figures (error score) and .40 with the Johns Hopkins Perceptual Test, and the correlations with Story Sequence dropped (r = -.28 and .27, respectively) though the correlation between ETS Matched Pictures and Story Sequence rose from .22 to .33. This may be evidence for verbal-perceptual differentiation.

Story Sequence II, given only in Year 2, also had highest correlations with measures that have high loadings on factor 1. This was most marked for items 1 and 3 (story recall). For card order these items correlated highest with the Preschool Inventory (r = .39), Peabody Picture Vocabulary Test Forms A and B (r = .32, .29), and TAMA (r = .31). For language order the correlations were similar, though the highest correlation was .44 with Massad Mimicry (real words). The Preschool Inventory correlated .44, PPVT Forms A and B correlated .40 and .35, respectively, TAMA correlated .39, and Children's Auditory Discrimination Inventory (nonsense words) correlated .34.

Correlations for items 2 and 4 (story telling) of Story Sequence II were all quite low. For card order and fluency the highest correlations were with Massad Mimicry (nonsense words): r = .21 and .18, respectively. This suggests that items 1 and 3 are measuring different productive language skills than are items 2 and 4. Items 2 and 4 make demands on the child's creative ability to make up his own story, whereas items 1 and 3 require only that the child retell a story just told to him; however, items 1 and 3 make greater demands on memory skills.

Summary

The ETS Story Sequence Task proved to be appropriate for this age range and was greatly enjoyed by most children. It should be noted, however, that the productive items were the most difficult and not recommended for use with similar populations below the age range presently sampled (4 1/2 - 5 1/2).

From analyses performed thus far, it appears that the child's sex was the variable most consistently affecting performance on this task. SES, as indexed by mother's education, and age also showed a significant effect for several of the scores. With the exception of scores for the child's production of his own stories, scores correlated highest with tasks tapping general information-processing skills (factor 1). Performance may also be related to creativity and memory ability. Measures and analyses in future years of the study may help clarify these relationships. As with other language measures, particular attention will be paid to investigating the child's verbal environment as indicated in parent interview and structured mother-child interaction sessions. Additional understanding is also expected when data on school achievement, particularly reading, and preschool attendance are analyzed.

ETS Story Sequence was successful in avoiding the single-word stimulus approach of so many standard language tests which usually assess only receptive language, and then to a limited extent. Although these more traditional types of measures are easier to administer and score, greater understanding of the complex nature of verbal behavior may be expected from a measure such as the present one.

References

Baratz, J. C. A bi-dialectal task for determining language proficiency in economically disadvantaged Negro children. Child Development, 1969, 40, 889-902.

Blank, M. & Frank, S. Story recall in kindergarten children: Effect of method of presentation on psycholinguistic performance. Child Development, 1971, 42, 299-312.

Bruce, D. Representational and syntactic competence of problem readers. Child Development, 1970, 41, 3, 881-887.

Cazden, C. B. Subcultural differences in child language. Merrill-Palmer Quarterly, 1966, 12, 185-219.

Deutsch, M. The disadvantaged child and the learning process. In A. H. Passow (Ed.), Education in depressed areas. New York: Bureau of Pub., Teachers College, Columbia University, 1963. Pp. 163-179.

Gray, Susan W. & Klaus, R. An experimental preschool program for culturally deprived children. Child Development, 1965, 36, 887-898.

Hildreth, G. H., Griffiths, N. L. & McGauvran, M. E. Metropolitan Readiness Tests, Form A. Harcourt, Brace, and World, Inc., 1965.

Jeruchimowicz, R., Costello, J. & Bagur, S. Knowledge of action and object words: A comparison of lower and middle-class Negro preschoolers. Child Development, 1971, 42, 455-464.

John, Vera P. & Goldstein, L. The social context of language acquisition. Merrill-Palmer Quarterly, 1964, 100, 265-275.

Kirk, S. A., McCarthy, J. J. & Kirk, W. D. Illinois Test of Psycholinguistic Abilities. Revised Edition, Urbana, Illinois, University of Illinois, 1968.

Labov, W. The logic of nonstandard Negro English. In R. D. Abrahams and R. C. Troike (Eds.), Language and cultural diversity in American education. Englewood Cliffs, N. J.: Prentice-Hall, 1972. Pp. 225-261.

Loban, W. Language proficiency and school training. In J. E. Krumboltz (Ed.), Learning and the educational process. Chicago, Illinois: Rand McNally, 1965. Pp. 113-131.

McCarthy, D. Language development in children. In L. Carmichael (Ed.), Manual of child psychology. (2nd ed.) New York: Wiley, 1954. Pp. 492-630.

Milgram, N. A., Shore, M. F. & Malasky, C. Linguistic and thematic variables in recall of a story by disadvantaged children. Child Development, 1971, 42, 637-640.

Raph, J. E. Language development in socially disadvantaged children. Review of Educational Research, 1965, 35, 389-400.

Rieber, M. & Womack, M. The intelligence of preschool children as related to ethnic and demographic variables. Exceptional Children, 1968, 34, 609-616.

Stearns, K., Hodges, W. & Spicker, H. Interim report: A diagnostically based language curriculum for psycho-socially deprived preschool children. Presented at a Symposium on "Early First Language Instruction for the Culturally Deficient Child," American Educational Research Association meetings, Chicago, 1966.

Stern, Carolyn. Language competencies of young children. Young Children, 1966, 22(1), 44-50.

Stewart, W. A. Toward a history of American Negro dialect. In F. Williams (Ed.), Language and poverty: Perspectives on a theme. Chicago: Markham Publishing Co., 1970. Pp. 351-379.

Tanaka, M. Time sequence task. In Melton, R. S., Charlesworth, R., Tanaka, M. N., Rothenberg, B. B., Pike, L. W., Bussis, A. M. and Gollin, E. S., Cognitive growth in preschool children, Princeton, N. J.: Educational Testing Service, 1968, RM-68-13.

Weaver, S. J. Interim report: Psycholinguistic abilities of culturally deprived children. Unpublished manuscript. George Peabody College for Teachers, 1965.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

Technical Report 7

Fixation Time

Virginia C. Shipman
Karla S. Goldman

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Fixation Time

Background

It has been demonstrated that at least in the early years of life attention is an index of early cognitive functioning (Kagan & Lewis, 1965; Lewis, 1971, 1972; Lewis, Goldberg & Campbell, 1970). Individual differences in attention may also have direct effects on learning. Thus, in the preschool child, attention may relate to later as well as current cognitive functioning. It is also possible that attention is noncognitively determined by the intentions and desires of the child (Messick, 1968).

One aspect of the study of attention is the investigation of habituation which may be measured by the technique of repetitive stimulation (Lewis & Goldberg, 1969; Lewis, Goldberg & Campbell, 1970). The same stimulus is presented on repeated trials, followed by a novel stimulus at the end of the series. Rate of response decrement over repetitious trials provides an index of habituation. Response recovery to the final novel stimulus provides assurance that the preceding decrement is not a function of general fatigue.

Visual habituation is the most easily measured form of habituation, and probably the most extensively explored (Fantz, 1964; Kagan & Lewis, 1965; Lewis, 1971; Lewis, Goldberg & Campbell, 1970). The most direct measure of visual habituation is fixation, or looking-time. A review of much of the research on fixation in infancy is offered by Kagan (1971) who claims that change (physical contrast or movement), discrepancy from schema, and activation of hypotheses can all be indexed by fixation measures during the first two years of life.

It is assumed that differences in rate of habituation have important effects on children's learning. The child who cannot concentrate or grows bored quickly cannot obtain as much information from his environment as the child who can; while the child who shifts attention rapidly because he has absorbed the available information (the "quick" child) has an advantage in assimilating information. In a recent study of habituation in disadvantaged black children (ages 48-73 months), Hamrick, Hamrick, and Kilpatrick (1971) found Ss "unable to maintain a state of habituation" to a novel auditory stimulus over a period of time. The authors interpret this result in terms of adaptation, suggesting that it may be "maladaptive for children reared in a socially deprived environment to maintain a prolonged decrease in vigilance (extended habituation) [p. 6]."

Task Description and Administration

The fixation task used in this study measures the amount of time a child fixates or looks at a given picture (S_1) as it is repeated over six trials and then followed by a novel picture (S_2) on the seventh.

Two series of slides were used. The first consisted of a redundant nonsocial visual stimulus: six trials of a slide showing twenty chromatic straight lines and a seventh of chromatic curved lines. The second series, a social array, consisted of a chromatic schematic representation of a family shown for six trials, and a seventh presentation of the same schematic without color. Each slide was shown for 30 seconds and followed by a black slide for 30 seconds before the next presentation. The slides were shown automatically with a projector unit which resembled a large television. The child was seated approximately three feet in front of the screen and the observer was seated next to the projector, hidden from view by a peg-board partition through which she watched the child. Fixation time was recorded by a stopwatch as the amount of time (in seconds) the child looked at each picture before turning away the first time.* If a child did not look at the picture within ten seconds of presentation, he was reminded to do so; if he was already looking at the screen when the slide appeared, timing began immediately. A short break was allowed between the two series but not between slide (trial) presentations.

Although this task has previously been used in laboratory settings under highly controlled conditions, these ideal conditions did not exist for the field operations of the present study. The differential lighting conditions affected the visibility of the slides and of the observer and the darkened condition of the room was frightening to some children. Additional difficulties were caused by external noise which distracted the child, and by equipment failures such as jamming of slides and lack of precise machine timing of slide intervals. In Year 1, data from one of the four study sites were discarded due to continual equipment problems.

Although not insurmountable, problems were also created by the simultaneous demands upon the observer. While watching the child's eye movement she also had to be timing with a stopwatch, reading these times, and recording them to a tenth of a second.

Scoring

By presenting slides in two series (social and nonsocial), each consisting of six repeated stimuli and then one novel stimulus, it was

*In past research, interobserver reliability for determining whether the child was looking at the screen varied between .60 and .99 (ETS, 1968).

possible to obtain separate reliability estimates and to examine stimulus content differences. Three basic measures of attention were obtained: response decrement (habituation), response recovery (stimulus differentiation) and mean initial viewing time (amount of attention). Response decrement, or habituation, was measured by comparing the amount of fixating time to a repeated stimulus over successive trials. It was calculated as 1/2 (trial 1 plus trial 2 minus trial 4 minus trial 5). This procedure uses four points on the curve to determine decrement and avoids using the same points as in the recovery score. Stimulus differentiation was measured by response recovery when a novel stimulus was presented after the repeated stimuli. This score was calculated as the difference between viewing time in trial 7 and trial 6. Amount of attention was the mean initial viewing time over all repeated trials (i.e., trials 1 to 6) for each series.

Score Properties

Intercorrelations for recovery, response decrement, and mean initial viewing time scores for the nonsocial and social series are presented in Table 1. The highest correlations are consistently those involving mean initial viewing time. In Year 1 mean initial viewing time on the social and nonsocial series correlated .89 and .86, respectively, with mean initial viewing time for the series combined. (These are part-whole correlations which have not been corrected for overlap.) Mean initial viewing time for each series correlated .52. In Year 2, similar correlations were found: .88, .83, and .45, respectively. The recovery scores correlated only .19 across series in Year 1 and .09 in Year 2. Similarly, the response decrement scores correlated .13 across series in Year 1 and .05 in Year 2.

Table 1

Intercorrelations for Recovery, Response Decrement, and Mean Initial Viewing Time Scores

	1	2	3	4	5	6	7
Within Year 1 (N = 941-1024)							
Nonsocial:							
1. recovery	--	.11	-.06	.19	.18	.17	.07
2. response decrement	.11	--	.12	.08	.13	.08	.11
3. mean initial viewing time	-.06	.12	--	.09	.08	.52	.86
Social:							
4. recovery	.19	.08	.09	--	.19	-.04	.03
5. response decrement	.18	.13	.08	.19	--	.07	.09
6. mean initial viewing time	.17	.08	.52	-.04	.07	--	.89
Series combined:							
7. mean initial viewing time	.07	.11	.86	.03	.09	.89	--
Within Year 2 (N = 674-765)							
Nonsocial:							
1. recovery	--	.06	-.15	.09	.09	.05	-.04
2. response decrement	.06	--	.07	.07	.05	-.10	-.02
3. mean initial viewing time	-.15	.07	--	-.01	-.05	.45	.83
Social:							
4. recovery	.09	.07	-.01	--	.12	-.23	-.16
5. response decrement	.09	.05	-.05	.12	--	-.11	-.10
6. mean initial viewing time	.05	-.10	.45	-.23	-.11	--	.88
Series combined:							
7. mean initial viewing time	-.04	-.02	.83	-.16	-.10	.88	--
Year 1 x Year 2 (N = 548-617)							
Nonsocial:							
1. recovery	-.06	-.01	.00	-.02	.02	-.03	-.02
2. response decrement	.03	.02	.04	.00	.04	-.02	.00
3. mean initial viewing time	-.01	.12	.25	.03	.00	.14	.20
Social:							
4. recovery	-.07	.00	-.02	.04	.01	-.06	-.05
5. response decrement	-.10	-.09	-.01	-.04	-.05	.00	.00
6. mean initial viewing time	.06	.12	.19	.12	.02	.15	.19
Series combined:							
7. mean initial viewing time	.02	.12	.25	.08	.01	.16	.22

Across years the pattern remains the same with mean initial viewing time demonstrating the highest correlations, but the differences are not as striking. Mean initial viewing time for the nonsocial series had the highest correlation across years (r = .25), followed by mean initial viewing time for the series combined (r = .22), and mean initial viewing time for the social series (r = .15). Recovery and response decrement scores had approximately zero correlations across years. Given these patterns of correlations, mean initial viewing time over both series was the only score used in subsequent structural analyses.

Sample Performance

Data in Tables 2-5 are based on the total sample tested in each year. Data in Table 2 show that, as predicted, initial looking time decreased as the repeated stimulus was shown over trials 1 through 6. Then as the novel stimulus was shown on trial 7, initial looking time more than doubled what it was on trial 6. It is also clear from these data that social stimuli were attended to longer than nonsocial stimuli, even though they were presented as the second series of slides.

Mean initial viewing time for the repeated stimulus over trials 1 through 6 and for the two series combined is presented by sex and three-month age intervals in Table 3. No clear age trends appear in either year for either series. Sex differences did not appear in Year 1 and in Year 2 only for the social stimuli, with girls looking longer than boys. Mean recovery scores (Table 4) also show no clear age trends except for social stimuli in Year 2 when scores decreased with age. Although boys had slightly higher recovery scores in Year 1, in Year 2 girls had slightly higher scores than boys. Similarly, the mean response decrement scores (habituation) presented in Table 5 reveal no clear age trends. Girls had slightly higher decrement scores for the nonsocial stimulus in Year 1 and higher scores for the social stimulus in Year 2. For the social stimulus in Year 1 and the nonsocial stimulus in Year 2, there were no sex differences. Nonetheless, it is clear that the social stimuli consistently evoked greater habituation.

Table 2

Mean Initial Viewing Time for Total Group by Trial*
in Year 1 and Year 2

	Nonsocial N	Nonsocial Mean	Nonsocial SD	Social N	Social Mean	Social SD
Year 1						
Trial 1	1212	12.3	8.51	1220	19.6	9.52
Trial 2	1218	9.9	8.03	1212	14.3	9.37
Trial 3	1215	8.3	7.52	1210	11.4	9.21
Trial 4	1214	7.8	7.42	1214	10.1	9.01
Trial 5	1207	7.6	7.21	1214	9.0	.32
Trial 6	1195	7.4	7.61	1206	8.7	8.31
Trial 7	1208	16.8	9.78	1207	18.9	9.44
Year 2						
Trial 1	863	11.5	8.64	864	22.6	8.58
Trial 2	874	10.4	7.87	858	17.4	9.82
Trial 3	869	9.7	8.22	858	13.8	9.98
Trial 4	869	9.3	8.46	849	11.9	9.63
Trial 5	872	9.0	8.36	852	11.1	9.47
Trial 6	844	8.6	8.10	843	10.4	9.18
Trial 7	867	20.7	9.26	839	22.3	8.65

*in seconds.

Table 3

Mean Initial Viewing Time* (Trials 1 to 6) by Sex and Age
for Year 1 and Year 2

	Nonsocial			Social			Combined Series		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Year 1									
42-44 mo.	89	53.0	31.42	88	73.2	34.44	94	63.2	29.84
45-47 mo.	271	52.0	27.45	269	75.6	34.02	281	63.7	27.64
48-50 mo.	269	54.0	32.06	277	70.8	34.54	286	62.5	30.38
51-53 mo.	284	56.4	30.98	295	75.7	34.76	304	66.4	30.19
54-56 mo.	220	51.6	31.11	223	68.6	34.61	227	60.4	28.49
57-59 mo.	44	50.7	27.47	43	66.6	33.01	44	58.1	26.51
Boys	633	53.5	31.36	640	72.7	34.58	667	63.4	30.00
Girls	544	53.3	29.24	555	72.7	34.44	569	63.1	28.35
Total	1177	53.4	30.39	1195	72.7	34.50	1236	63.3	29.24
Year 2									
51-53 mo.	70	55.3	29.37	71	81.9	32.59	75	67.7	25.64
54-56 mo.	192	59.9	27.86	183	87.1	32.23	199	73.3	25.95
57-59 mo.	192	55.7	28.77	195	89.4	33.08	201	72.1	27.65
60-62 mo.	226	59.0	26.86	221	88.9	34.36	239	73.4	27.49
63-65 mo.	139	58.1	29.26	135	87.4	34.77	149	71.7	29.64
66-69 mo.	13	61.6	26.03	13	90.6	36.97	13	76.1	28.36
Boys	430	57.3	28.08	427	85.1	33.81	459	70.4	27.42
Girls	402	58.8	28.18	391	90.7	32.97	417	74.5	27.25
Total	832	58.0	28.12	818	87.8	33.51	876	72.3	27.40

*in seconds.

Table 4

Mean Response Recovery (Trial 7-Trial 6)* by Sex and Age in Year 1 and Year 2

	Nonsocial			Social		
	N	Mean	SD	N	Mean	SD
Year 1						
42-44 mo.	78	8.3	10.60	81	10.1	10.48
45-47 mo.	272	9.6	10.78	270	11.1	10.56
48-50 mo.	288	9.0	10.53	290	9.6	10.61
51-53 mo.	295	9.7	10.76	302	10.2	10.81
54-56 mo.	218	9.7	10.15	213	9.9	11.07
57-59 mo.	35	8.8	8.38	36	11.5	10.35
Boys	631	9.6	10.62	633	10.5	10.89
Girls	555	9.1	10.39	559	10.1	10.50
Total	1186	9.4	10.51	1192	10.3	10.70
Year 2						
51-53 mo.	66	11.8	10.30	67	13.9	11.34
54-56 mo.	194	12.7	10.62	192	12.5	10.31
57-59 mo.	197	12.4	11.45	195	11.2	11.59
60-62 mo.	232	12.1	10.87	232	11.7	11.24
63-65 mo.	137	11.5	11.81	134	11.4	11.37
66-69 mo.	10	7.8	12.74	10	7.7	13.32
Boys	429	11.7	11.38	427	11.7	11.08
Girls	407	12.5	10.73	403	11.9	11.28
Total	836	12.1	11.07	830	11.8	11.17

*in seconds.

Table 5

Mean Response Decrement* by Sex and Age in Year 1 and Year 2

	Nonsocial			Social		
	N	Mean	SD	N	Mean	SD
Year 1						
42-44 mo.	92	2.3	6.76	90	7.8	7.94
45-47 mo.	275	3.9	6.66	274	7.7	8.30
48-50 mo.	277	3.8	6.87	278	7.9	8.19
51-53 mo.	294	3.2	7.26	299	7.1	9.43
54-56 mo.	223	3.2	6.85	225	7.1	7.45
57-59 mo.	46	2.5	6.00	44	5.6	7.10
Boys	651	3.1	7.13	651	7.4	8.39
Girls	556	3.8	6.58	559	7.4	8.32
Total	1207	3.4	6.88	1210	7.4	8.36
Year 2						
51-53 mo.	72	3.3	7.94	72	7.2	7.62
54-56 mo.	197	1.7	7.75	188	7.8	8.31
57-59 mo.	197	2.0	7.64	196	9.1	8.96
60-62 mo.	237	1.9	8.34	227	8.9	9.11
63-65 mo.	147	1.5	7.71	141	8.9	9.66
66-69 mo.	13	-1.7	6.82	13	7.4	7.58
Boys	449	1.9	7.73	438	7.9	8.90
Girls	414	1.8	8.07	399	9.2	8.76
Total	863	1.9	7.89	837	8.5	8.86

*in seconds.

These data are based on the total sample tested in Year 1 and the total sample tested in Year 2 and are thus overlapping, not identical, samples. In an age x sex x SES analysis of variance on the longitudinal sample (those children tested in both Year 1 and Year 2) separately by year, age differences (using a median age split) were not significant for the nonsocial series for any measure. For the social series, age was marginally significant ($F = 4.61$, $df = 1/664$, $p < .04$) for recovery scores in Year 2, with older Ss looking less.

A repeated-measures ANOVA (age x sex x SES) was also performed on the longitudinal sample using scores for mean initial viewing time (trials 1 through 6, series combined), recovery, and detrement scores for each series separately. The only variable to reach statistical significance was SES (indexed by mother's education: below 10th grade, 10th-12th grade, above 12th grade) for recovery scores on the nonsocial series ($F = 3.58$, $df = 2/559$, $p < .03$), with recovery scores increasing with SES level. The repeated-measures ANOVA did demonstrate significant increases in scores from Year 1 to Year 2. For the nonsocial series the recovery score showed a highly significant year effect ($F = 24.17$, $df = 1/563$, $p < .001$). For the social series, significant year effects were obtained for both recovery ($F = 6.96$, $df = 1/565$, $p < .01$) and response decrement scores ($F = 8.34$, $df = 1/552$, $p < .004$). For the series combined, mean initial viewing time also was significant ($F = 6.56$, $df = 1/592$, $p < .01$). No significant sex differences were obtained.

Relationship with Other Measures

Correlations* with other measures in the study were low. The range of

*These correlations are based on data from longitudinal subjects only. Correlations for the total Year 1 sample were essentially identical.

correlations for mean initial viewing time (which was the score used in structural analyses) in Year 1 was .00 to .24, with highest correlations with the Peabody Picture Vocabulary Test, Form B (r = .24), and the Matching Familiar Figures Test, errors (r = .20); in Year 2 correlations ranged from .00 to .12. Mean initial viewing time defined task-specific factors when factor analyses were performed on Year 1 data both for the total sample and for longitudinal subjects only. Factor analyses of the Year 2 longitudinal sample data showed fixation to load highly on a factor defined by itself and rating of the child's cooperation in the Hess and Shipman Eight-Block Sorting Task (13-factor Varimax solution; see Shipman, 1971 and 1972, for detailed discussion of these structural analyses). As suggested by Shipman (1972) this may indicate the generalization of personal-social behaviors (controlling mechanisms, compliance, desire to please) across tasks. However, the lack of correlation with other measures and the low correlations across subscores within the task make the findings difficult to interpret.

Summary

In spite of field operation difficulties, results obtained for the Fixation Task are consistent with previous research findings. There was response decrement over repeated stimuli and recovery with a changed stimulus, and these appeared to be influenced by SES, measured by amount of mothers' formal education. However, the low correlations across stimuli, among subscores, and with other measures provide little understanding of these data at this time.

As measured by the Year 1-2 test battery, attention does not appear to be directly related to cognitive ability in this population at this age. It is plausible that attention, as assessed here, is more a motivational construct within the domain of personality. Although Lewis and his associates (Lewis, 1971; Lewis et al., 1970) found attention to be an index of early cognitive functioning, attentional variables are among those in this study that cut across relatively arbitrary distinctions between cognitive and personal-social functioning and may be "non-cognitively" determined by the child's intentions and desires. Shipman (1972) also suggests that differences in viewing time may reflect differences in the desire to follow instructions and please the examiner. Since the child is explicitly instructed to look, his behavior may be controlled as much by the instructional set as by rate of information processing. As mentioned earlier, mean initial viewing time did load on a factor defined primarily by the child's cooperation during the Eight-Block Sorting Task.

These are only tentative conclusions based on less-than-ideal testing conditions, including equipment failures, competing stimuli, differential lighting conditions, the child's fear of the dark, and problems of compliance to social expectations. Future analyses will reveal the extent to which performance on this task during this age period is predictive of other indices of cognitive, affective, and social functioning in subsequent years.

References

Educational Testing Service. Disadvantaged children and their first school experiences: Theoretical considerations and measurement strategies. Princeton, N. J.: December 1968, PR-68-4 (2 vols.). Prepared under Contract OEO 4206 and Grant CG-8256, Office of Economic Opportunity. (ED 037 486)

Fantz, R. L. Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. Science, 1964, 146, 668-670.

Hamrick, D. D., Hamrick, O. M., Jr. & Kilpatrick, D. G. Habituation of the orienting response in socially deprived Negro children. Paper presented at the meeting of the Society for Psychophysiological Research, St. Louis, Missouri, October 1971.

Kagan, J. Change and continuity in infancy. New York: Wiley, 1971.

Kagan, J. & Lewis, M. Studies of attention in the human infant. Merrill-Palmer Quarterly, 1965, 11, 95-127.

Lewis, M. Individual differences in the measurement of cognitive development. In J. Hellmuth (Ed.), Exceptional infant, Vol. 2. New York: Brunner/Mazel, Inc., 1971, 172-211.

Lewis, M. The cardiac response during infancy. In R. F. Thompson and M. M. Patterson (Eds.), Methods in physiological psychology, Vol. 1. Recording of bioelectric activity. New York: Academy Press, 1972.

Lewis, M. & Goldberg, S. The acquisition and violation of expectancy: An experimental paradigm. Journal of Experimental Child Psychology, 1969, 7, 70-80.

Lewis, M., Goldberg, S. & Campbell, H. A developmental study of learning within the first three years of life: Response decrement to a redundant signal. Monographs of The Society for Research in Child Development, 1970, 34 (9, Serial No. 133).

Messick, S. Children's cognitive and perceptual development. In Disadvantaged children and their first school experiences: Theoretical considerations and measurement strategies, Volume 1. Princeton, N. J.: Educational Testing Service, December 1968, PR-68-4. Prepared under OEO Contract 4206 and Grant CG-8256, Office of Economic Opportunity (ED 043 397).

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

Technical Report 8

Form Reproduction

David R. Lindstrom
Virginia C. Shipman

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Form Reproduction

The Form Reproduction task was included in the Longitudinal Study test battery to 1) examine visual-motor performance over time; 2) study processes associated with visual analysis and synthesis and visual-motor integration; and 3) examine its predictive validity for later academic/skill behaviors, e.g., writing.

Maccoby (1968) has suggested that the processes involved in copying forms are perceptual, conceptual, discriminatory, and motor in nature and that the ability is "a very slowly acquired skill, and ...there is a highly predictable order in which forms become reproducible [p. 163].". This order is based on a developmental progression from holistic perceptions and reproductive attempts to (1) fractionation of stimulus elements, (2) maintenance of constancy of part-whole relations, and (3) conceptual organization of these into functional relationships.

The relationship between age and success in form reproduction has been reported in several studies (Beery, 1967a, 1967b; Birch & Lefford, 1967; Gordon & Hyman, 1970). Typically, higher scores are earned by older Ss, with older Ss being more competent in copying complex forms.

Sex differences in ability to reproduce forms were reported by Denton (1968) with kindergarten-aged females performing better than males. No sex differences, however, were reported by Connor (1967) for second grade Ss on the Bender-Gestalt or by Pascale (1970) using the Developmental Test of Visual Motor Integration (VMI) with kindergarten Ss (mean age of 60.1 months).

Much research on visual-motor skill has emphasized its relation to academic achievement and learning difficulties. Owen, Adams, Forrest, Stolz and Fisher (1971) found poorer performance on the Bender-Gestalt for Ss with learning disabilities compared with non-impaired controls. Brown (1970), Connor (1967), Lackmann (1960), Nielson and Ringe (1969), and O'Donnell and Eisenson (1969) reported poorer performance on tests of visual-motor ability for Ss diagnosed as having reading difficulties in comparison with Ss not having difficulties. Beery (1967a), Wechsler (1967), and Egeland, DiNello and Carr (1970) reported that scores on visual-motor tasks correlated with MA, CA, IQ and various language and achievement skills.

Several studies have demonstrated the effectiveness of visual and motor skill experience and training in improving visual-motor performance. Campbell (1971), Lipton (1969), Maccoby (1968) and Pascale (1970) reported that Ss' performance on tests of perception, drawing, writing and reading improved after participation in programs emphasizing perceptual, visual discrimination, motor coordination, and visual-motor integration skills. Roberts (1970) reported that Ss who had attended kindergarten performed better on tests of intersensory integrative functioning (in first grade) when compared with a group not having had kindergarten experience.

In summary, successful reproduction of geometric forms is a complex process involving the development, integration and coordination of several sense modalities. The studies reported above indicate that for preschool and primary grade Ss it is positively correlated with age, scholastic achievement, and intelligence test scores and is related to experiential factors.

Task Description and Administration

Form Reproduction in Year 1 data collection consisted of six geometric forms. Four of these items were included in the Preschool Inventory which was administered as part of the overall test battery--vertical line, circle, square and triangle. The oblique cross and right oblique line items were added from the VMI (Beery, 1967a). For Year 2 data collection, Form Reproduction consisted of nine forms, the above six forms plus the inverted T, three vertical lines and the adjoining square and circle from the WPPSI. The tester asked the child to copy the four Preschool Inventory forms during administration of this test. Upon completion the child was asked to copy the remaining forms on a supplemental sheet. It should be noted that the first four items are included in the VMI, although their ordering is somewhat different. Test materials consist of the two answer sheets and a kindergarten pencil for copying.

Administering the Form Reproduction task is fairly simple. For the items included in the Preschool Inventory, testers must be sure to insert a sheet of black construction paper under the sheet presented to S to prevent him from copying the forms printed on the reverse side. For all items, testers must be sure the pages are presented to the child in the proper orientation. Standard encouragement probes are used for refusals and "don't knows." Only one trial is allowed for each item unless the child spontaneously rejects his figure and draws another; erasures are permitted but only at the child's instigation. Drawings should be carefully numbered in case the child does not draw in the areas indicated.

Scoring

The circle, square, inverted T, and parallel lines were scored 0, 1, 2 and the square and circle 0, 1, 2, 3 using Wechsler's (1967) criteria; the remaining forms were scored 0, 1 using Beery's (1967a) criteria. Total scores were obtained by summing across items for a maximum total score of 8 for Year 1 and 15 for Year 2. All tests were double-scored at the ETS Princeton office and checked by senior research staff who resolved any scoring discrepancies. Inter-scorer agreement was high.

Score Properties

Table 1 presents inter-item and total score correlations for the three-site longitudinal sample (Ss tested in both years).* Inter-item correlations were low for both years' data. The highest relationship in both years was between the square and the triangle (.39 in Year 1 and .49 in Year 2), which both require the integration of vertical and horizontal elements for successful reproduction. Item correlations with total score were moderate in both years. (These item-scale correlations are part-whole correlations and have not been corrected for overlap.) The rank ordering of correlations between items included in both years' testing was only moderate (rho = .42). Reliability (coefficient alpha) for total score in Year 1 was .61, and .71 in Year 2. Although these values are lower than those reported by Beery (1967a) and Wechsler (1967), this test was also shorter than the VMI and WPPSI. The correlation between total score for both years was .52; the correlation based on identical items only was .49.

Sample Performance

Mean total scores are reported in Tables 2 and 3 by sex and three-month age intervals for Year 1 and Year 2 data. An increase in total score with age is evident in both years. Also, in Year 1 increased variability was associated with higher scores. In Year 2, the variances were generally higher than in Year 1 which may have been due to the increase in the length of the test and a wider range of possible scores.

For longitudinal Ss (i.e., those tested in both years) analysis of variance for age was performed for each year separately using a median-split. There was a highly significant difference both in Year 1 ($F = 66.72$, $df = 1/1109$,

*The supplemental items were not administered to the Lee County sample in Year 2.

Table 1

Inter-item and Total Score Correlations for Years 1 and 2 for the Three-Site Longitudinal Sample ($\underline{N}$ = 646-790)

	Line	Circle	Square	Triangle	Right Oblique	Cross	Inverted T	Three Vertical Lines	Square and Circle	Total Year 2
Line	.08	.18	.07	.07	.12	.02	.08	.13	.07	.28
Circle	.28	.16	.28	.24	.12	.09	.26	.18	.21	.54
Square	.18	.28	.36	.49	.29	.23	.37	.30	.43	.70
Triangle	.15	.24	.39	.24	.28	.28	.31	.29	.40	.64
Right Oblique	.19	.24	.25	.17	.23	.27	.24	.19	.25	.51
Cross	.06	.16	.14	.09	.14	.13	.18	.23	.19	.45
Inverted T	--	--	--	--	--	--	--	.31	.28	.62
Three Vertical Lines	--	--	--	--	--	--	--	--	.18	.59
Square and Circle	--	--	--	--	--	--	--	--	--	.6[illegible]
Total-Year 1	.58	.75	.59	.48	.58	.38	--	--	--	.52

Note.--Values to the left of and below the diagonal represent Year 1 data.
Values to the right of and above the diagonal represent Year 2 data.
Values in the diagonal reflect Year 1 - Year 2 correlations between items.
For $\underline{N}$ = 500, $\underline{r}_{.001}$ = .148.

Table 2

Mean Total Score* by Age and Sex for Year 1 (four-site sample)

Group	N	Mean	SD
42-44 mo.	88	1.49	1.24
45-47 mo.	318	1.65	1.34
48-50 mo.	346	1.82	1.41
51-53 mo.	384	2.40	1.65
54-56 mo.	271	2.53	1.72
57-59 mo.	61	2.84	1.88
Boys	783	1.93	1.56
Girls	685	2.27	1.59
Total	1468	2.09	1.58

*range = 0-8.

Table 3

Mean Total Score* by Age and Sex for Year 2 (three-site sample)

Group	N	Mean	SD
51-53 mo.	73	4.10	2.55
54-56 mo.	210	4.41	2.45
57-59 mo.	207	4.61	2.73
60-62 mo.	239	5.26	2.79
63-65 mo.	141	5.98	2.78
66-69 mo.	9	6.33	2.60
Boys	460	4.54	2.71
Girls	419	5.36	2.70
Total	879	4.93	2.74

*range = 0-15.

$p < .001$) and in Year 2 ($F = 27.43$, $df = 1/714$, $p < .001$) favoring the older group.

Repeated-measures analysis of variance (sex x age x SES) was performed for the longitudinal sample using scores based on common items only. This analysis indicated a highly significant sex difference ($F = 19.03$, $df = 1/689$, $p < .001$) with girls obtaining higher scores. This analysis also indicated a significant age effect across years ($F = 444.20$, $df = 1/693$, $p < .001$) favoring the older group. SES differences were examined using a three-way split for mother's education--above 12 years, 10-12 years, below 10 years. A highly significant difference overall was found ($F = 29.44$, $df = 2/689$, $p < .001$), with mean scores increasing as mother's education increased.

Analysis of variance was also performed using the total score from the Year 2 administration of the test. Again, significant sex differences ($F = 12.66$, $df = 1/714$, $p < .001$) and SES differences were obtained ($F = 38.91$, $df = 2/714$, $p < .001$), with girls and children whose mothers had attended school longer obtaining higher scores.

Maccoby (1968) has suggested that successful performance in reproducing forms depends on an interaction between age and stimulus complexity (e.g., single strokes, lines involving directionality, forms involving integration of parts). Beery (1967a) has constructed the VMI in an age-sequential pattern in which stimuli progress from single, elementary shapes to complex integrated forms. Inspection of percent passing the Form Reproduction items indicated performance increased for common items during this age period and the vertical line and circle were the easiest for this sample in both years. Data for percent of Ss receiving full and partial credit for each item are presented in Table 4. In Year 1 ordering of forms from least to most difficult for this sample would be: vertical line, circle, right oblique line, square, cross and triangle. For Year 2, the ordering would be: vertical line, circle, inverted T, three vertical lines, right oblique line, square, triangle, circle and square, and cross. The most difficult items were those requiring complex integration of elements and differentiation of part-whole relationships.

Table 4

Percent of Subjects Receiving Full and Partial Credit for Form Reproduction Items

Item	Year 1 (N = 1470)	Year 2 (N = 905)
Vertical Line	69.05	85.88
Circle	60.27	77.28
Square	13.75	29.84
Triangle	7.07	25.98
Right Oblique Line	30.54	40.23
Cross	11.83	17.38
Inverted T	*	65.78
Three Vertical Lines	*	41.65
Square and Circle	*	21.58

*Item not administered in Year 1.

Relationship with Other Measures

Factor analyses of child test data for Years 1 and 2 indicated that the Form Reproduction score had high loadings on a "g" or general information-processing skills factor (see Shipman, 1971, 1972 for a detailed description of these analyses). This factor was defined by scores from verbal, general information, classification, and discrimination measures.

Correlations among Form Reproduction and purported measures of achievement and language skills (Preschool Inventory, Peabody Picture Vocabulary Test receptive and productive vocabulary), form discrimination (Johns Hopkins Perceptual Test, Matching Familiar Figures), form analysis (Preschool

Embedded Figures Test, WPPSI Picture Completion) and eye-hand coordination (Seguin Form Board) are reported in Table 5. Moderate correlations between Form Reproduction scores and these variables are apparent. Although analytic perception did not emerge as a separate factor, there were moderate correlations between Form Reproduction and purported analytic perception measures (e.g., Johns Hopkins Perceptual Test, Preschool Embedded Figures). The loadings for Form Reproduction on the "g" factor and its relationship with other measures indicate that comprehension and verbal skills in addition to perceptual-motor coordination contributed to this score.

Table 5

Inter-Correlations Among Form Reproduction and Selected Scores for Years 1 and 2 for the Three-Site Longitudinal Sample

		1	2	3	4	5	6	7
Form Reproduction	1	--	.54	.46	.41	.39	-.46	.40
Preschool Inventory	2	.51	--	.66	.63	.48	-.54	.36
PPVT A	3	.40	.58	--	.73	.37	-.50	.29
PPVT B	4	.33	.50	.69	--	.33	-.50	.27
Johns Hopkins	5	.32	.32	.32	.25	--	-.42	.31
MFF (errors)	6	-.39	-.42	-.43	-.32	-.43	--	-.30
PEFT	7	.32	.32	.31	.13	.25	-.23	--
*Picture Completion	8	.45	.59	.56	.56	.30	-.40	.29

Note.--Values to the right of the diagonal are for Year 2 (N = 719-796). Values to the left of the diagonal are for Year 1 (N = 433-781). All correlations are significant beyond the .001 level of confidence (r.001 = .148 for N = 500).

*Since this task was administered in Year 1 only, correlations are based on data for the total Year 1 sample (N = 942-1386).

Summary

Several aspects of the Year 1 and 2 data support the use of this measure of form reproduction to tap an ability related to young children's cognitive development. Despite restricted ranges, performance increased with age and varied as a function of the complexity of the stimulus items. SES differences were consistent with previous research in similar aged samples and indicate that experiential factors contribute to this difference. Sex differences favoring females, although statistically significant, in absolute terms were small.

The relationships among form reproduction scores and achievement measures, although moderate, were in the expected direction and consistent with past research. Form reproduction loaded onto a general "information-processing" factor with cognitive, perceptual and language variables, indicating other variables associated with performance on this task. The reliability coefficients obtained indicated moderate intra-individual consistency in performance; stability of performance across years was moderately high for this age sample. Future analyses will focus on the sequential development of form reproduction skill, examining the pattern of correlates with academic and other cognitive-perceptual skills within and across age periods and its relationship to experiential variables in the home and school.

References

Beery, K. E. Visual-motor integration (Monograph). Chicago: Follett Educational Corporation, 1967a.

Beery, K. E. Developmental test of visual-motor integration: Administration and scoring manual. Chicago: Follett Education Corpora ion, 1967b.

Birch, H. G. & Lefford, A. Visual differentiation, intersensory integration, and voluntary motor control. Monographs of the Society for Research in Child Development, 1967, 32 (2, Serial No. 110).

Brown, L. S. Good and retarded readers compared according to performance on visual-motor, simulated language learning and memory tasks. (Doctoral dissertation, University of Northern Colorado) Ann Arbor, Michigan: University Microfilms, 1970. No. 71-14, 519. (Abstract)

Campbell, J. H. The effects of different types of training of visual discrimination, auditory discrimination, and visual-motor coordination on reading readiness test scores of kindergarten children. (Doctoral dissertation, The Florida State University) Ann Arbor, Michigan: University Microfilms, 1971. No. 71-6977. (Abstract)

Connor, J. P. The relationship of Bender-Visual-Motor Gestalt Test performance to differential reading performance of second grade children. (Doctoral dissertation, Kent State University) Ann Arbor, Michigan: University Microfilms, 1967. No. 67-9416. (Abstract)

Denton, C. J. Visual-motor skills and body strength of disadvantaged children. (Doctoral dissertation, the University of Oklahoma) Ann Arbor, Michigan: University Microfilms, 1968. No. 68-13,557. (Abstract)

Egeland, B., DiNello, M. & Carr, D. The relationship of intelligence, visual-motor, psycholinguistic and reading readiness skills with achievement. Educational and Psychological Measurement, 1970, 30, 451-458.

Gordon, G. & Hyman, I. The measurement of perceptual-motor abilities of Head Start children. Princeton, N. J.: Educational Testing Service, 1970, RB-70-11.

Lackmann, F. M. Perceptual-motor development in children retarded in reading ability. Journal of Consulting Psychology, 1960, 24, 427-431.

Lipton, E. D. The effect of a physical education program to develop directionality of movement on perceptual-motor development, visual perception, and reading readiness of first grade children. (Doctoral dissertation, New York University) Ann Arbor, Michigan: University Microfilms, 1969. No. 69-21,216. (Abstract)

Maccoby, E. E. What copying requires. Ontario Journal of Educational Research, 1968, 10, 163-170.

Nielson, H. H. & Ringe, K. Visuo-perceptive and visuo-motor performance of children with reading disabilities. Scandinavian Journal of Psychology, 1969, 10, 225-231.

O'Donnell, P. A. & Eisenson, J. Delacato training for reading achievement and visual-motor integration. Journal of Learning Disabilities, 1969, 2, 441-447.

Owen, F. W., Adams, P. A., Forrest, T., Stolz, L. M. & Fisher, S. Learning disorders in children: Sibling studies. Monographs of the Society for Research in Child Development, 1971, 36 (4, Serial No. 144).

Pascale, M. A. The effect of a visual-motor integration training program on beginning writing skills of kindergarten children. (Doctoral dissertation, Columbia University) Ann Arbor, Michigan: University Microfilms, 1970. No. 70-12,566. (Abstract)

Roberts, A. J. The relationship between kindergarten experience and fine-muscle eye-hand coordination abilities of first grade children. (Doctoral dissertation, Oregon State University) Ann Arbor, Michigan: University Microfilms, 1970. No. 70-22,772. (Abstract)

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education and Welfare.

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

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

YOUNG CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

A LONGITUDINAL STUDY

Technical Report 9

Johns Hopkins Perceptual Test

Lynn E. Gilbert
Virginia C. Shipman

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Johns Hopkins Perceptual Test

Purpose

The Johns Hopkins Perceptual Test was developed in 1966 by L. A. Rosenberg, A. M. Rosenberg, and M. Stroud, as a brief measure of intelligence in young children. It was recognized that available measures of intellectual function had serious limitations with children who had functional or organically determined speech defects, limited verbal and experiential repertoires, or motor handicaps, as well as with very young or retarded children. The aim, therefore, was to develop a diagnostic instrument for the evaluation of such children. In preliminary work with this test using 340 children ranging from 3 to 6 years of age, Rosenberg (1966) obtained correlations of .62 and .45 with the Peabody Picture Vocabulary Test for middle-class and lower-class children, respectively, and correlations of .80 and .66 with the Columbia Mental Maturity scale. The perceptual nature of the task, however, was a major factor in its inclusion in the Longitudinal Study battery.

A number of studies with preschool children have indicated that they can discriminate forms (Brown & Goldstein, 1967; Gaines, 1969; Wohlwill & Wiener, 1964), with the number of errors decreasing as age increases (Cronin, 1967; Robinson & Higgins, 1967). Taylor and Wales (1970) delineated three stages in form discrimination among their sample of 47 three- and four-year-olds using a match-to-standard task: first, a period characterized by a high number of errors, with responses made largely in terms of positional effect (i.e., proximity of a comparison figure to the standard); second, a period of multiple responses, with choices made on the basis of similarity to the stimulus figure but without sufficient differentiation to produce the correct match; and third, a correct and unique choice made.

Much research has focused on the exact nature of the discrimination to be made. with variations among forms occurring in overall shape (configuration, number of sides), orientation (up-down, right-left reversals), and detail (open or closed lines, linear or curvilinear). Special emphasis is given those discriminations which appear to be important in the development of reading skills. Rudel and Teuber (1963) and Enterline (1970) both found left-right reversals to be more difficult than up-down ones. Gibson, Gibson, Pick and Osser (1962) used twelve letter-like forms and various transformations of them as comparisons. Children of ages 4 to 8 were to select all the ones like the standard from a row of thirteen alternatives, and errors were classified according to the type of transformation erroneously identified with the standard. Most difficult, for younger as well as older Ss, was the "perspective" transformation (e.g., a 45 degree slant of the figure); easiest to discriminate were "topological" changes (e.g., a broken or closed line). Even the youngest children, though, made very few errors of omission--that is, failed to select the true match. In a study with 36 nursery school children, ages 3-1 to 5-1, Braine (1965) found the effect of orientation on form recognition to change with age. He interpreted this as a change in the part used by the child as the starting point for scanning, the younger children looking first at a "focal point" of the figure and the older children looking first at the top of the figure.

Attention has also been focused on methodology. Cronin (1967) presented mirror-image triangles to kindergarteners and first-graders using three different formats: 1) a dyad, in which Ss were asked to make a judgment of "same" or "different:" 2) oddity problem, in which Ss were asked "which one is different" out of three figures; 3) match-to-standard, in which Ss were asked which of two triangles is the same as a third. Methods two and three did not differ significantly, while the dyad proved to be the most difficult. Furthermore, the only children who failed to show improvement during the task were those given the dyad without corrective feedback.

Robinson and Higgins (1967) asked children to judge pairs of triangular forms as "same" or "different" and then had them show via gestures in what direction the figures pointed. They found that children could discriminate the forms although they called them the "same," and concluded that inadequacies may be related to testing procedures rather than to perceptual lack. Jeffrey (1958) found that 4-year-olds who learned to press buttons oriented in the same direction as stick figures were subsequently able to label the figures more successfully.

Ricciuti (1963) conducted a study to determine what stimulus components the child considers when making a similarity judgment. He presented his subjects, ages 2-11 to 8-3, with four geometric figures and a standard (no two identical) and asked which one was most like the standard and why. Overall, about 20% of the judgments were based on stimulus detail rather than on gross form.

Task Description

The Johns Hopkins Perceptual Test requires the child to choose form identical to a standard. It consists of 3 practice and 30 test items, all involving black geometric figures printed on white cards. There is one booklet for stimulus cards and one for response cards.

The child is presented with a stimulus form and asked to point to the one just like it among several alternatives. The alternative figures are all on one card, while each stimulus figure is printed separately. The pages of the stimulus booklet are thus turned after each item, and the pages of the response booklet are turned after each alternative figure on the page has been presented once as the stimulus.

Item difficulty is varied by differences in number of angles in the figure and in number of alternatives given the child (either 2, 3, or 5). In any given set of alternative figures, overall form remains constant; differences may occur in size, orientation, and/or degree of angularity.

The task is not difficult to administer and takes about 10 minutes with three- to five-year-old children. Pointing to each alternative when introducing a new response card, E included special instructions ("Look at this one, this one, and this one.") to focus the child's attention on all the figures of a given item. This procedure, as well as the inclusion of practice items with corrective feedback, was introduced in order to increase the likelihood that errors were due to perceptual abilities and not to stylistic factors or irrelevant elements in the testing situation, a possibility raised by Robinson and Higgins (1967).

Scoring

Items were scored as correct, incorrect, refused, or indeterminate (e.g., multiple answers). The total score is the number of correct matches made (maximum is 30). For the Longitudinal Study two "subset" scores were also computed. Gordon (1969) had distinguished between items in which the child responds to the figure as a whole and makes a "global" comparison, and items involving more complex figures in which the child compares them in terms of subtle differences in component parts. The former type of discrimination was hypothesized to constitute a "form perception" subset of the test, whereas the latter type would constitute an "analysis" subset.

Gordon distinguished 16 "form perception" and 14 "analysis" items.

Score Properties

Item analysis did not support the use of separate perception and analysis subscores. Item-intercorrelations in general were moderate to low and were as high across the two item types as they were within each type. The confounding of item type with order of presentation (9 of the 14 "analysis" items were in the second half of the test) and difficulty level (all analysis items had the maximum number of alternative responses) makes it difficult to tease out process differences among items. Given the above, only the total correct score was used.

Table 1 presents the biserial correlations and percent passing each item for Years 1 and 2. For the total sample in Year 1 the alpha coefficient of reliability for total score for a *N* of 1410 was .76; *r*-biserials ranged from .27 to .72. Reliability coefficients were highly similar across groups differentiated by age, sex, and SES. In Year 2, with *N* = 1317, coefficient alpha was .73, with *r*-biserials between .30 and .62.

The correlation between Year 1 and Year 2 scores was .27.

Sample Performance

Tables 2 and 3 present the means, standard deviations, and percentile distributions for total score by age and sex for Year 1 and Year 2. The task proved to be of moderate difficulty for most of the children in this sample, and scores were relatively well distributed throughout the possible range. These scores are similar to those obtained by Rosenberg (1972) with a sample of mixed SES. (For his subsample of 158 children ages 4-0 to 4-11, a score of 19 fell at the 50th percentile and a score of 22 to 24 at the 90th. For 110 children ages 5-0 to 5-5, the 50th percentile score was 21 and the 90th was 25.)

Table 1

Item Data: Percent Passing and Biserial Correlations in Year 1 and Year 2

Item	Year 1 (N = 1410) Percent Correct	R Biserial	Year 2 (N = 1317) Percent Correct	R Biserial
1	69.3	.49	83.6	.55
2	71.1	.51	84.1	.59
3	63.7	.59	83.1	.62
4	87.4	.70	93.2	.50
5	74.3	.58	87.0	.59
6	65.1	.64	77.4	.59
7	76.2	.48	77.7	.42
8	72.6	.57	78.8	.50
9	90.0	.72	94.4	.57
10	66.1	.57	79.2	.59
11	59.3	.49	76.2	.52
12	45.6	.44	55.0	.47
13	34.2	.35	41.6	.46
14	49.9	.37	54.5	.46
15	86.5	.68	91.7	.45
16	43.5	.47	49.3	.47
17	55.6	.46	62.2	.50
18	96.5	.52	97.1	.44
19	59.9	.56	70.2	.60
20	49.6	.49	63.2	.49
21	39.9	.31	51.7	.37
22	53.6	.49	57.9	.50
23	54.3	.58	60.7	.58
24	51.7	.41	53.7	.39
25	31.5	.36	38.6	.39
26	29.9	.55	44.8	.45
27	33.3	.27	33.7	.30
28	23.0	.35	28.8	.33
29	32.6	.29	32.6	.32
30	26.5	.30	27.3	.32

Table 2

Distributions of Total Score* by Age and Sex, Year 1

Group	N	Mean	SD	Percentiles 10	25	50	75	90
42-44 mo.	80	16.5	5.23	9.5	12.8	16.7	20.2	23.0
45-47 mo.	295	15.7	4.82	8.9	12.6	15.8	19.4	22.0
48-50 mo.	328	16.2	4.80	10.0	13.1	16.1	19.4	22.3
51-53 mo.	379	17.9	4.60	11.6	15.1	17.8	21.4	24.1
54-56 mo.	270	17.7	4.84	11.0	14.3	17.6	21.3	23.8
57-59 mo.	59	18.0	4.55	11.8	14.2	18.4	21.2	23.5
Boys	746	16.7	4.86	10.1	13.6	16.7	20.2	23.0
Girls	665	17.1	4.86	10.5	13.8	17.2	20.9	23.4
Total	1411	16.9	4.86	10.2	13.7	16.9	20.5	23.2

*Range = 0-30.

Table 3

Distributions of Total Score* by Age and Sex, Year 2

Group	N	Mean	SD	Percentiles 10	25	50	75	90
51-53 mo.	81	18.7	4.21	12.8	15.9	18.8	21.2	24.1
54-56 mo.	311	18.7	4.29	13.0	15.6	18.9	21.8	24.5
57-59 mo.	309	19.0	4.54	13.0	15.8	19.6	22.0	24.8
60-62 mo.	352	19.4	4.34	13.6	16.3	19.8	22.6	24.9
63-65 mo.	246	20.4	4.25	15.0	18.0	20.8	23.4	25.4
66-69 mo.	16	20.8	4.02	14.8	19.5	20.5	23.5	26.9
Boys	698	19.1	4.50	12.8	16.0	19.5	22.3	24.9
Girls	617	19.5	4.24	14.0	16.5	19.7	22.4	24.9
Total	1315	19.3	4.38	13.4	16.3	19.6	22.3	24.9

*Range = 0-30.

Analyses of variance were performed on the longitudinal sample; that is, the children tested in both Year 1 and Year 2. In a repeated-measures ANOVA SES* was a significant variable ($F = 63.98$, $df = 2/1045$, $p < .001$) when data were summed across years; high SES children obtained the highest mean score and low SES children, the lowest. No significant sex differences were found. Because children fell in different age categories in the two years, another analysis of variance was performed for each year separately. Using a median age split, there was a significant age difference in both Year 1 ($F = 52.79$, $df = 1/1070$, $p < .001$) and in Year 2 ($F = 17.47$, $df = 1/1129$, $p < .001$) in favor of the older children. The repeated-measures ANOVA also showed a significant increase in scores from Year 1 to Year 2 ($F = 245.97$, $df = 1/1049$, $p < .001$).

Relationship with Other Measures

Factor analyses of the Year 1 and Year 2 child test data did not yield any clusterings of perceptual tasks. Rather, the Johns Hopkins Perceptual Test had its highest loading on the first factor, best defined as general cognitive skills or "g." (See Shipman, 1971, 1972, for a further description of these analyses.)

Correlations of the Johns Hopkins Perceptual Task were moderate to high with measures tapping both general mental ability and perceptual discrimination. In Year 1 (using the total Year 1 sample)** it correlated .39 with the Preschool Inventory, .35 with Peabody B (productive language), and .33 with Peabody A (receptive language), all of which tap general information and verbal skills; and -.52 with errors on the Matching Familiar Figures

---

*Mother's education was the index of SES: below 10th grade; grades 10-12; above 12th grade.

**Correlations for the three-site longitudinal sample were lower by .01 to .10.

Test and -.41 with quickest time to correct solution on the Seguin Form Board Test. Other high correlations were .43 with ETS Story Sequence and .41 with Sigel Grouping Responses, both of which had high loadings on the general ability factor and also contain perceptual elements.

Year 2 correlations (using the three-site longitudinal sample) showed a similar pattern, the higher relationships occurring with general ability and perceptual measures. Correlations were .48 with the Preschool Inventory; -.42 with errors on the Matching Familiar Figures test; .42 with TAMA, a measure of general knowledge; .40 with ETS Matched Pictures, tapping language skills using a picture-choice format; .39 with Form Reproduction; .37 with Peabody A; -.36 with quickest time to correct solution on the Seguin Form Board Test; and .33 with Peabody B.

Summary

The Johns Hopkins Perceptual Test was found to relate to other measures of general ability as its developers intended, and also to tasks with a high perceptual component. SES and age were significant variables in both study years, and there was also a significant increase in performance from Year 1 to Year 2. The SES finding indicates that experiential factors continued to influence results on this measure, which had a minimal verbal element and therefore was hypothesized to be less susceptible to SES.

Future investigation will be directed to performance on this task as it relates to behavior on perceptual and other measures included in subsequent years of the study. Particular attention will be paid to those tasks requiring integration of perceptual processes with implications for successful school performance.

References

Braine, L. G. Age change in mode of perceiving geometric forms. Psychonomic Science, 1965, 2, 155-56.

Brown, D. R. & Goldstein, J. A. Stimulus correlates of the discrimination behavior of children. Psychonomic Science, 1967, 9(4), 177-78.

Cronin, V. Mirror-image reversal discrimination in kindergarten and first-grade children. Journal of Experimental Child Psychology, 1967, 5, 577-85.

Enterline, E. G. Form discrimination: Spatial relationships between a standard and comparison figure. Perceptual and Motor Skills, 1970, 30(3), 959-69.

Gaines, R. The discriminability of form among young children. Journal of Experimental Child Psychology, 1969, 8(3), 418-31.

Gibson, E. J., Gibson, J. J., Pick, A. D. & Osser, H. A. A developmental study of the discrimination of letter-like forms. Journal of Comparative and Physiological Psychology, 1962, 55(6), 897-906.

Gordon, G. Personal correspondence, 1969.

Jeffrey, W. E. Variables in early discrimination learning: Motor responses in the training of a left-right discrimination. Child Development, 1958, 29(2), 269-75.

Ricciuti, H. N. Geometric form and detail as determinants of comparative similarity judgments in young children. In A basic research program on reading, Cornell University, Ithaca, N. Y. 1963.

Robinson, J. S. & Higgins, K. E. The young child's ability to see a difference between mirror-image forms. Perceptual and Motor Skills, 1967, 25, 893-97.

Rosenberg, L. A., Rosenberg, A. M. & Stroud, M. The Johns Hopkins Perceptual Test: The development of a rapid intelligence test for the preschool child. Paper presented at the meetings of the Eastern Psychological Association, New York, April 1966.

Rosenberg, L. A. Personal communication, 1972.

Rudel, R. G. & Teuber, H. L. Discrimination of direction of line in children. Journal of Comparative and Physiological Psychology, 1963, 56, 892-98.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Taylor, J. A. & Wales, R. J. A developmental study of form discrimination in preschool children. Quarterly Journal of Experimental Psychology, 1970, 22(4), 720-34.

Wohlwill, J. F. & Wiener, M. Discrimination of form orientation in young children. Child Development, 1964, 35, 1113-25.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

YOUNG CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

A LONGITUDINAL STUDY

Technical Report 10

Massad Mimicry Test I and II

Virginia C. Shipman
Karla S. Goldman
Susan Simosko

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Massad Mimicry Test I and II

Background

It has been hypothesized that children initially learn language through imitation. Research by Slobin and Welsh (1967), Keeney (1969), and Fraser, Bellugi and Brown (1963) indicates that children's linguistic competence may be assessed through controlled, elicited imitation and that the ability to imitate may be separate from understanding or producing language.

Fraser, Bellugi and Brown (1963), studying imitation and recall in white 3-year-olds, found better repetition of an utterance than identification of the picture belonging to it, and better identification than giving the pictures sentence names. The authors claim that these experimental results indicate that imitation is a perceptual-motor skill that is dissociated from comprehension and that acquisition of structures occurs sequentially from imitation to comprehension and finally to production of these structures. The possibility also exists, however, that repetition is dependent on comprehension rather than the reverse, and that imitation is not merely a perceptual-motor skill.

Analyzing language samples obtained from young children, Ervin-Tripp (1964) found that depending on the situational variables and the structure and length of the utterance, imitation of utterances may be simply a perceptual-motor skill or a process involving comprehension. Menyuk (1971) suggests that the child may merely imitate the phonetic string he hears or may regenerate the sentence by using structures in his own grammatical system. She questions whether it is necessary for the child, in the acquisition of grammatical structures, to go through the process of (uncomprehending) phonetic imitation first. Imitation has been stressed as a necessary prerequisite for comprehension of the meaningfulness of the signal and as an indication that the signal is indeed comprehended. Although Osser et al. (1969) obtained a significant rank order correlation between comprehension errors and critical structure errors, Hall and Turner (1971) did not replicate this finding and claim it merely indicates that poor imitators were also poor comprehenders. The important question, though, is whether a child who is imitating a specific sentence is also attending to the meaning of it or merely making a psychomotor response. It is also entirely possible that the child comprehends the meaningfulness of some utterances before he attempts either to imitate or to produce them. Menyuk (1971) asserts that the facts are as yet unavailable. It is hoped that the present study will be able to look at the rate of development of imitation skills compared to comprehension skills.

An added difficulty in the assessment of these processes is the comparison of groups who use standard English with others who use non-standard English. Baratz (1969) studied imitation skills in urban lower-class black and suburban lower-middle-class white third and fifth graders, asking them to repeat 15 sentences in standard English and 15 sentences in black dialect English. White Ss imitated standard English significantly better than black Ss, and black Ss imitated black dialect significantly better than white Ss. Osser, Wang and Zaid (1969) used two psycholinguistic tasks to assess speech imitation and speech comprehension abilities in 16 lower-class black and 16 middle-class white kindergarten boys. The middle-class Ss had significantly better scores than the lower-class Ss on all indicators for the two tasks. The authors suggest that the task was more difficult for the black lower-class Ss who had to recode the test sentences in the imitation task.

This is in contrast to Fraser, Bellugi and Brown (1963) and Lovell and Dixon (1967) who suggest that in imitating a sentence the child does not process it through his own meaning system and that imitation involves mechanical rather than constructive processes. Osser et al. (1969) also found that for black lower-class <u>Ss</u>, whether comprehension occurs or not, the imitation response will tend to conform to the child's own dialect. Thus, these findings give evidence that sentences are processed through some structural and phonological components of the child's linguistic system, and suggest that this is a function of class rather than race (Hall & Turner, 1971). This interpretation is also supported by Menyuk's study of imitation in nursery school children (1963).

In addition, it is not clear what role memory, auditory and articulatory skills play in this process. Van Riper (1972), Menyuk (1971) and others have considered the lack of adequate listening skills to be a primary factor in misarticulation, which would suggest that auditory and articulatory skills are highly related. While little research exists linking memory with imitation skills, there is evidence (Goetzinger, Dirks & Baer, 1960) that limitations on memory may affect imitation. They may account for those aspects of an utterance to which a child <u>can</u> pay attention and code but not necessarily to those aspects to which he <u>does</u> pay attention, as represented by the utterances he produces and the structure to which he responds. However, what Brière (1967) attributed to memory limitations, Kamil and Rudegeair (1972) posit as merely the result of attentional processes or some simple form of task learning.

Unfortunately, there is a dearth of successful measures in this field. Friedlander (1965) attempted to describe the articulatory and intelligibility status of 150 disadvantaged preschool children. Using the Templin-Darley

Test of Articulation and an experimental vocabulary test, he found no significant differences in intelligibility between black and white subgroups or between black and Spanish subgroups. There were, however, statistically significant differences between Spanish and white Ss. Further, father's occupation and family income were not significantly correlated with children's articulation, intelligibility, or verbal skill. The language behavior of their parents (on the Templin-Darley Test) and not SES was the significant factor in language development of these children.

In an attempt to provide a better measure of imitation and articulation in children, Stern (1967) developed the Echoic Response Inventory. She hoped to determine if disadvantaged four-year-olds would perform better on echoed or imitated items when the items were presented in speech characteristic of the community in which the children were raised. It was found that white children did significantly better on the test when the items were presented in standard English than when presented in dialect. For the black children there was no significant difference, suggesting that they do as well with standard English as they do with dialect at least in this testing context.

Task Description and Administration

Similar to the Echoic Response Inventory, the Massad Mimicry Test is an individually administered task for 3 1/2-4 1/2-year-old children. Part I evaluates the child's ability to reproduce phonemes in thirty nonsense words upon hearing each no more than three times from a tape-recorded model. Similarly, Part II assesses the child's ability to reproduce meaningful words and phonemes as they occur in word phrases and simple sentences. In Year 1, Part II consisted of 13 phrases and 2 simple sentences. In Year 2, Part II consisted of 9 sentences. The test takes approximately 10 minutes to administer. The tester uses two tape recorders, one for playing the model (stimulus) tape, and the other for recording the child during the testing session. Both the model utterances and the child's responses are recorded on the latter tape.

Prior to testing, it is important to establish rapport with the child so that he talks comfortably, and understands that he is making a recording and must not play with the equipment during the testing session. The test is proposed as a game of "Follow the Leader" in which the model utterance is the Leader and is to be followed exactly by the child. Preceding the beginning of each part of the test there is a warm-up of three sample utterances, which are also recorded. Positive reinforcements of the child's responses are given only at four designated times during the actual testing. He is encouraged, however, to speak loudly and directly into the microphone. To avoid the possibility of the tester's remarks being confused with the model tape, only nonverbal communications should be used while it is playing. Each test item may be repeated only twice after the child does not respond.

In order to administer this task, testers must acquire a high level of dexterity in using the two tape recorders. They must be able to use both simultaneously and be able to rerun words on the model tape with facility and accuracy. During the testing session the tester must assure a good recording by using maximally quiet surroundings and by restricting the child's tactile curiosity. Since young children in testing situations tend to speak softly, the volume should be adjusted accordingly.

Scoring

All scoring of the finished tapes was done at the ETS Princeton office by staff who had obtained high agreement with the test developer in determining whether a child correctly matched the sound on the master tape. Part I is composed of three primary scores: Initial sounds (possible score, 30), Medial sounds (possible score, 28), and Final sounds (possible score, 30). Part II provides two scores: Final sounds (possible score: 10 in Year 1, 9 in Year 2) and Model word or some semblance of it (possible score: 35 in Year 1, 23 in Year 2). The three scores for Part I may be totaled for a score on nonsense words; however, each score in Part II is independent and should not be totaled. In addition, the Medial sounds include twelve long vowels and thirteen short vowels which may be studied independently. However, if a score is given for each of these two types of vowels, it must be remembered that the scores are inter-dependent with the score for Medial sounds. Copies of the actual Scoring Guides for each year are included in the Appendix.

Scoring may be difficult if the taping is done under adverse conditions, i.e., interference from extraneous noise, poorly placed microphone, etc.

Interscorer Reliabilities

Interscorer reliability in Year 1 was determined for three scorers (A, B and C). Test tape recordings of 300 children were selected randomly, the proportion from each site reflecting the sample size per site. Each tape was scored twice by independent judges.

Table 1 indicates interscorer reliabilities for Part I and its three subsections and the two subsections of Part II. The reliabilities given in the first column are not adjusted for differences in means between judges, whereas those in the second column are adjusted. The latter refer to interscorer reliabilities adjusted for differences in means that systematically vary among judges and should not be considered part of child-error variance. The analysis indicates that when adjustments were not made for differences in means of judges, only two scorers in Year 1 tended to show

Table 1

Interscorer Reliabilities

Part		Subsection		N	Reliability Unadjusted	Reliability Adjusted
Year 1: Judges A and B						
I.	Nonsense Words			289	.71	.86
		A.	Initial Sounds	289	.81	.82
		B.	Medial Sounds	291	.71	.84
		C.	Final Sounds	289	.45	.77
II.	Meaningful Words in Phrases					
		A.	Final Sounds	273	.66	.67
		B.	Model Word or Some Semblance of It	277	.78	.80
Year 1: Judges A and C						
I.	Nonsense Words			137	.05	.85
		A.	Initial Sounds	137	.60	.85
		B.	Medial Sounds	139	.07	.72
		C.	Final Sounds	137	-.46	.74
II.	Meaningful Words in Phrases					
		A.	Final Sounds	123	.54	.60
		B.	Model Word or Some Semblance of It	101	-.04	.81
Year 2: Judges A and B						
I.	Nonsense Words			494	.75	.87
		A.	Initial Sounds	494	.83	.84
		B.	Medial Sounds	494	.18	.78
		C.	Final Sounds	496	.80	.83
II.	Meaningful Words in Phrases					
		A.	Final Sounds	382	.94	.94
		B.	Model Word or Some Semblance of It	180	.81	.84

agreement at a significant level for Part I and all subsections. However, when adjustments were made for differences in means for judges, agreement was shown to be at a significant level. Seventy-seven percent of all tests in the sample studied were scored by that pair of scorers that tended to show agreement at a significant level even when scorer variation was accounted for. Nevertheless, in analyzing the data appropriate adjustments were made for differences in means for the three judges.

In Year 2, there were two judges and they demonstrated high reliabilities for adjusted scores and also for unadjusted scores except on Part IB. Improved intensive training resulted in higher inter-rater agreement in Year 2.

Score Properties

To identify trends in children's ability to reproduce initial, medial and final phonemes, it was necessary to obtain separate scores for each of these phoneme positions. In addition to looking at specific phoneme production, children's ability to reproduce a meaningful word or some semblance of it was measured since such knowledge contributes to the total picture of language development, particularly in reference to meaningful communication. Thus, the eight scores derived were I. Nonsense Words: total sounds, initial sounds, medial sounds, final sounds, long vowels and short vowels; II. Meaningful Words in Phrases: final sounds, and model word or some semblance of it.

The three experimentally independent scores were Nonsense Words (total sounds) and Meaningful Words in Phrases (final sounds, and model word). In Year 1 Nonsense Words (total sounds) correlated .56 with Meaningful Words in Phrases: final sounds, and .53 with Meaningful Words in Phrases: model word or some semblance of it; Meaningful Words in Phrases: final sounds, and Meaningful Words in Phrases: model word or some semblance of it correlated .47. In Year 2 these scores correlated .38, .62, and .46, respectively.

Table 2 reports estimated score reliabilities in Year 1 and in Year 2. The data indicate that, for an experimental measure, a satisfactory degree of internal consistency exists within Part I and the various subsections. Parts ID and IE, consisting of 12 and 13 items, respectively, each independent of the other but included in Part IB, necessarily reflect lower reliabilities than the longer subsection to which they belong. The low reliability of Part IIA (final sounds) may also be attributable to the fact that it contains only ten items whereas all other subsections, except ID and IE, contain no less than twenty-eight.

Table 2

Estimated Reliabilities for Internal Consistency*

Part			Reliability Year 1	Year 2
I. Nonsense Words			.91	.91
	A.	Initial Sounds	.75	.67
	B.	Medial Sounds	.76	.80
	C.	Final Sounds	.83	.85
	D.	Long Vowels	.59	.59
	E.	Short Vowels	.61	.76
II. Meaningful Words in Phrases				
	A.	Final Sounds	.63	.77
	B.	Model Word or Some Semblance of It	.90	.76

*The Kuder-Richardson (Formula 20) estimate of reliability.

Examining again the three scores of primary importance, their reliabilities in Year 1 and in Year 2 were Nonsense Words (total sounds) .91, .91; Meaningful Words in Phrases: final sounds .64, .77; Meaningful Words in Phrases: model word or some semblance of it .90, .76. The first two scores were used in subsequent structural analyses. The correlation across years for Nonsense Words was .26 and for Meaningful Words in Phrases: final sounds .21.

Sample Performance

Although 80-90% of the sample attempted to respond to each item, there was a tapering off near the end of the task (Part IIB, model word or some semblance of it). Nonetheless this subsection had the highest mean percent passing each item (81%) in Year 1, with 50% of Ss passing it in Year 2. In Year 1 at least, this may be an indication that meaningful words were easier stimuli than nonsense words. Similar findings were reported for the Children's Auditory Discrimination Inventory (CADI) also used in this study. Final sounds appear to be the hardest. The data on mean percent passing items within each subsection are reported in Table 3.

A more detailed item analysis in Year 1 indicated that the most difficult sounds to reproduce were those that may be classified as blends: sc, sn, st, th, dl, lv, ch, rl, ng. For reproduction of both nonsense and meaningful words, these sounds consistently had a smaller proportion passing than any other sounds. Although the intercorrelations among items with similar sound and placement generally were significant at the .01 level, they tended to be low. The data appear to support the generally held notion that a given sound presented in the context of a word, rather than isolated, will be modified in production by adjoining sounds.

Table 3

Percent Passing Items

Part	Subsection	N	Mean	Range
Year 1				
I. Nonsense Words				
	A. Initial sounds	1106-1213	64.2	32-85
	B. Medial sounds	1097-1199	72.8	35-91
	C. Final sounds	1099-1200	42.0	17-74
II. Meaningful Words in Phrases				
	A. Final sounds	1048-1112	36.9	17-52
	B. Model word or semblance of it	889-1114	80.9	61-93
Year 2				
I. Nonsense Words				
	A. Initial sounds	713-753	73.1	22-92
	B. Medial sounds	713-762	73.0	32-96
	C. Final sounds	715-762	49.5	25-77
II. Meaningful Words in Phrases				
	A. Final sounds	519-641	35.6	22-55
	B. Model word or semblance of it	354-585	50.2	21-83

Language development trends are indicated by the data reported in Tables 4 through 12. Examining the total-group ranges in Table 4, it is clear that in Year 1 the range of scores for Nonsense Words (Part I) remains fairly constant for four of the five subsections. The range for final sounds is smaller than for the others. For Meaningful Words (Part II), the two subsections are different from each other, although final sounds in Part II has a similar range to final sounds in Part I. In Year 2, both subsections of final sounds again have the smallest ranges. These data indicate that it is the lower range that contributes most to this aberration, and may be due to these sections having the smallest number of items.

Table 4

Total-Group Ranges

Part		Subsection	No. Items	N	Range (for adjusted scores)
Year 1					
I. Nonsense Words			88	1098	-4.05 to 2.49
	A.	Initial sounds	30	1101	-4.09 to 2.58
	B.	Medial sounds	28	1105	-.465 to 2.04
	C.	Final sounds	30	1100	-2.63 to 2.95
	D.	Long vowels	12	1101	-4.87 to 2.20
	E.	Short vowels	13	1139	-4.45 to 2.05
II. Meaningful Words in Phrases					
	A.	Final sounds	10	1060	-2.00 to 2.97
	B.	Model Word or Some Semblance of it	35	954	-5.85 to 1.74
Year 2					
I. Nonsense Words			88	709	-2.84 to 2.94
	A.	Initial sounds	30	711	-2.96 to 2.34
	B.	Medial sounds	28	710	-3.62 to 3.00
	C.	Final sounds	30	711	-2.38 to 2.80
	D.	Long vowels	12	715	-3.36 to 2.38
	E.	Short vowels	13	733	-4.40 to 2.42
II. Meaningful Words in Phrases					
	A.	Final sounds	9	556	-1.35 to 2.45
	B.	Model Word or Some Semblance of it	23	267	-2.16 to 1.92

Tables 5 and 6 report the means, standard deviations and ranges for Nonsense Words (total sounds) and Meaningful Words in Phrases: final sounds by three-month age intervals and by sex. Tables A through F in the Appendix report these same data for the other subscores. Data from all of these tables indicate that for this sample, children's ability to reproduce phonemes and meaningful words in phrases increased with age in Year 1 but not in Year 2.

A repeated-measures analysis of variance (age x sex x SES) performed on the longitudinal sample (i.e., those Ss who were tested in both Year 1 and Year 2) revealed main effects on Nonsense Words, total sounds when scores were combined across years for sex ($F = 6.07$, $df = 1/419$, $p < .02$) favoring girls; and for SES, using as an index mother's schooling--below 10 years, 10 to 12 years, above 12 years--($F = 11.30$, $df = 2/419$, $p < .001$), favoring the high SES group. Meaningful Words in Phrases could not be used in this analysis since different forms were used in each year. It is discussed below in the ANOVAS performed separately by year.

An analysis of variance performed on the longitudinal sample, separately for each year, showed age (using a median age split) to be significant for both Meaningful Words, final sounds ($F = 7.32$, $df = 1/809$, $p < .01$) and Nonsense Words, total sounds ($F = 18.33$, $df = 1/840$, $p < .001$) in Year 1, with the older Ss obtaining the higher scores. In Year 2, however, a significant age effect was found only for Nonsense Words, total sounds ($F = 6.44$, $df = 1/563$, $p < .02$), favoring the younger group. At the present time these data are not interpretable.

Since Meaningful Words: final sounds could not be used in the repeated-measures ANOVA, it was examined separately by year. Age was significant only in Year 1, as has been discussed above, and sex was significant only in Year 2 (F = 14.68, df = 1/439, $p < .001$), favoring girls. SES was the only variable significant in both years with higher scores obtained by children whose mothers had more schooling (F = 48.33, df = 2/809, $p < .001$; F = 13.12, df = 2/439, $p < .001$).

Table 5

Nonsense Words, Total Sounds: Means, Standard Deviations and Range for Year 1 and Year 2 by Age and Sex

Group	N	Mean	SD	Range
Year 1				
42-44 mo.	62	-0.35	1.05	-4.05 to 1.58
45-47 mo.	211	-0.23	1.09	-3.97 to 2.13
48-50 mo.	265	-0.07	1.00	-3.34 to 2.13
51-53 mo.	292	0.14	0.92	-2.28 to 2.44
54-56 mo.	222	0.14	0.93	-2.84 to 2.49
57-59 mo.	46	0.34	0.94	-1.54 to 2.20
Boys	569	-0.10	1.01	-4.05 to 2.49
Girls	529	0.11	0.98	-3.97 to 2.44
Total	1098	0.00	1.00	-4.05 to 2.49
Year 2				
51-53 mo.	52	0.18	0.83	-1.66 to 1.84
54-56 mo.	167	-0.08	1.06	-2.84 to 2.49
57-59 mo.	146	0.02	0.99	-2.20 to 2.58
60-62 mo.	198	0.07	0.99	-2.81 to 2.94
63-65 mo.	137	-0.10	1.02	-2.29 to 2.20
66-69 mo.	9	0.00	0.82	-1.09 to 1.75
Boys	371	-0.10	1.00	-2.38 to 2.49
Girls	338	0.11	0.99	-2.84 to 2.94
Total	709	0.00	1.00	-2.84 to 2.94

Table 6

Meaningful Words in Phrases, Final Sounds: Means, Standard Deviations and Range for Year 1 and Year 2 by Age and Sex

Group	N	Mean	SD	Range
Year 1				
42-44 mo.	62	-0.13	0.92	-1.49 to 1.74
45-47 mo.	198	-0.06	1.00	-1.95 to 2.39
48-50 mo.	254	-0.04	0.99	-2.00 to 2.83
51-53 mo.	285	0.02	1.01	-1.95 to 2.97
54-56 mo.	215	0.08	1.01	-1.95 to 2.97
57-59 mo.	46	0.12	1.07	-1.95 to 1.98
Boys	545	-0.07	0.98	-1.95 to 2.83
Girls	515	0.07	1.01	-2.00 to 2.97
Total	1060	0.00	1.00	-2.00 to 2.97
Year 2				
51-53 mo.	35	-0.12	0.94	-1.35 to 1.58
54-56 mo.	127	-0.08	1.04	-1.35 to 2.42
57-59 mo.	117	-0.10	0.99	-1.35 to 2.42
60-62 mo.	156	0.10	1.02	-1.35 to 2.04
63-65 mo.	114	0.06	0.95	-1.35 to 2.44
66-69 mo.	7	0.44	1.06	-0.93 to 2.04
Boys	281	-0.16	0.95	-1.35 to 2.44
Girls	275	0.16	1.03	-1.35 to 2.42
Total	556	0.00	1.00	-1.35 to 2.44

Relationships with Other Measures

In Year 1 the Massad Mimicry Test showed little communality with other tasks in the test battery. Using data based on all subjects tested in Year 2, the highest correlations with other tasks in the study ranged from .30 to .38 with similar values for both Nonsense Words and Meaningful Words. These tasks were the Preschool Inventory (.37, .38), Peabody Picture Vocabulary

Test (PPVT), Form A (.33, .33), Peabody Picture Vocabulary Test (PPVT), Form B (.33, .38), and Children's Auditory Discrimination Inventory (CADI), total correct (.31, .30), Matching Familiar Figures, errors (-.30, -.27). These correlations were even lower, particularly fo the Nonsense Words score, when using data based only on Ss tested in both Years 1 and 2. Although these other measures all had high loadings on a first factor which appeared to be tapping general information-processing skills, Mimicry itself had only moderate loadings. These data suggest the association of general ability, verbal comprehension and auditory discrimination with successful performance on this task. Age-specific factors, however, also appear to be operating.

In Year 2, correlations based on data from longitudinal Ss only were lower for Nonsense Words and higher for Meaningful Words. Again the highest correlations were with the Preschool Inventory (.35, .53), PPVT, Form A (.25, .44), TAMA (.24, .45), ETS Matched Pictures (.27, .43), PPVT, Form B (.26, .41), and CADI, nonsense words (.26, .41). Tasks tapping perceptual abilities showed the next highest correlations: Form Reproduction (.26, .39), Sigel Object Categorization, grouping responses (.18, .37), ETS Story Sequence, Part I (.22, .36), and Matching Familiar Figures, errors (-.28, -.34). As in Year 1, these tasks had high loadings on a general ability factor as did Mimicry. (See Shipman, 1971, 1972 for further discussion of these structural analyses.)

Thus it appears that the abilities tapped by this task in both years are related to the development of general competence, particularly receptive and productive language abilities, including auditory discrimination skill.

Summary

In Year 1 the Massad Mimicry Test apparently was tapping highly task-specific skills, defining a factor by itself; in Year 2 it loaded primarily on a general competency factor. This seems to support the hypothesis that in Year 1 behaviors were tapped at the beginning of a period of integration rather than during a period of differentiation, while in Year 2 these behaviors became more integrated with other cognitive-perceptual abilities. The pattern of correlations as well as factor loadings suggest the association of performance with general ability, verbal comprehension and auditory discrimination. However, it is not clear from the present data whether imitation is dependent upon comprehension or vice versa, nor can one delineate the role of attentional processes in the child's performance.

In both years, final sounds, whether in nonsense or meaningful words, appeared to be the most difficult items. Overall nonsense words were more difficult than meaningful words, and usually had lower correlations with other tasks.

Repeated-measures ANOVAS revealed significant effects for sex and SES for Nonsense Words, with girls and children whose mothers had more schooling obtaining higher scores. Separate ANOVAS performed on the Year 1 and Year 2 Meaningful Words data revealed significant effects for SES in both years, but significant sex differences only in Year 2. Further analyses, particularly those delineating home and school environments, should clarify these differences. The fact that girls generally performed better may reflect differences in task orientation and attending as suggested by other data in the study.

There is little doubt that the children's scores may have been affected by the testing situation itself. Problems were created by the demand for manual dexterity of the tester, and the child's adaptation to the strange talking machines. Children may have been too intimidated to perform at their normal level. Although this is often true in any testing situation, this particular situation appears to accentuate it. Many children during this age period required extensive instruction before completing the task. It should also be noted that scoring is a complex and lengthy procedure.

Appendix A

Scoring Guide for Year 1 and Year 2

Scoring Guide
(Year 1)

The total stimulus utterance is indicated at each number; however, the scorer is to listen for whether or not the examinee has correctly reproduced that part of the stimulus utterance that is underlined. Blacken space A if it was correctly reproduced; blacken space B if it was not correctly reproduced. For word reproduction a comprehensible form is acceptable.

Part I (1-88)

1. meb
2. kij
3. feg
4. pul
5. dit
6. bok
7. tan
8. rudl
9. hoot
10. wug
11. sach
12. zef
13. gish
14. kodl
15. nelv
16. yotl
17. jerl
18. scouv
19. sning
20. stogz
21. lal
22. hining
23. gats
24. buthing
25. tafs
26. kelvz
27. thadl
28. terling
29. chitlz
30. thood

*Rewind tape here to beginning of Part I

31. meb
32. kij
33. feg
34. pul
35. dit
36. bok
37. tan
38. rudl
39. hoot
40. wug
41. sach
42. zef
43. gish
44. kodl
45. nelv
46. yotl

jerl (on tape but no score)

47. scouv

sning (on tape but no score)

48. stogz
49. lal
50. hining
51. gats
52. buthing
53. tafs
54. kelvz
55. thadl
56. terling
57. chitlz
58. thood

*Rewind tape here to beginning of Part I

59. meb
60. kij
61. feg
62. pul
63. dit
64. bok
65. tan
66. rudl
67. hoot
68. wug
69. sach

70. zef
71. gish
72. kodl
73. nelv
74. yotl
75. jerl
76. scouv
77. sning
78. stogz
79. al
80. hining
81. gats
82. buthing
83. tafs
84. kelvz
85. thadl
86. terling
87. thitlz
88. thood

Part II (89-133)

89. girl with
90. is going
91. little girl

shelves are (on tape but no score)

92. big bottle
93. booking smells
94. saddle was
95. green elves

books on (on tape but no score)

96. playing games
97. are calling
98. only little

bottle top
Tom was lifting it.
It was lifted by Tom.
(on tape but no score)

---

*Rewind tape here to beginning of Part II. (If the word is repeated or some form of it, give credit. Examples: girl, girls, gels, etc., shelves, shelf, shevs, etc.)

99. girl with
100. is going
101. little girl
102. shelves are
103. big bottle
104. cooking smells
105. saddle was
106. green elves
107. books on
108. playing games
109. are calling
110. only little
111. bottle top

*Rewind tape here to beginning of Part II

112. girl with
113. is going
114. little girl
115. shelves are
116. big bottle
117. cooking smells
118. saddle was
119. green elves
120. books o
121. playing games
122. are calling
123. only little
124. bottle top
125. Tom was lifting it.

126. It was lifted by Tom.

*Rewind tape to:

127. Tom was lifting it.

128. It was lifted by Tom.

*Rewind tape to:

129. Tom was lifting it.

130. It was lifted by Tom.

*Rewind tape to:

131. Tom was lifting it.

132. It was lifted by Tom.

*Rewind tape to:

133. It was lifted by Tom.

Scoring Guide
(Year 2)

The total stimulus utterance is indicated at each number; however, the scorer is to listen for whether or not the examinee has correctly reproduced that part of the stimulus utterance that is underlined. Blacken space A if it was correctly reproduced; blacken space B if it was not correctly reproduced. For sentence reproduction (Part II, 89-97) a comprehensible form of words is acceptable, but the total sentence must be given and the order of words must be the same as that indicated.

Part I (1-88)

1. mēb
2. kĭj
3. fĕg
4. pūl
5. dīt
6. bŏk
7. tān
8. rŭdl
9. hōōt
10. wŭg
11. săch
12. zēf
13. gĭsh
14. kōdl
15. nĕlv
16. yŏtl
17. jĕrl
18. scouv
19. snĭng
20. stŏgz
21. lāl
22. hīnĭng
23. găts
24. būthĭng
25. tāfs
26. kĕlvz
27. thădl
28. tĕrling
29. chĭtlz
30. thōōd

*Rewind tape here to beginning of Part I

31. mēb
32. kĭj
33. fĕg
34. pūl
35. dīt
36. bŏk
37. tān
38. rŭdl
39. hōōt
40. wŭg
41. săch
42. zēf
43. gĭsh
44. kōdl
45. hĕlv
46. yŏtl

jĕrl (on tape but no score)

47. scouv

snĭng (on tape but no score)

48. stŏgz
49. lāl
50. hīnĭng
51. găts
52. būthĭng
53. tāfs
54. kĕlvz
55. thădl
56. tĕrlĭng
57. chĭtlz
58. thōōd

*Rewind tape here to beginning of Part I

59. mēb
60. kĭj
61. fĕg
62. pūl
63. dīt
64. bŏk
65. tān

66.	rŭdl	72.	kōdl	78.	stōgz	84.	kĕlvz
67.	hôot	73.	melv	79.	lăl	85.	thădl
68.	wŭg	74.	yŏtl	80.	hĭnĭng	86.	tĕrlĭng
69.	sāch	75.	jerl	81.	gāts	87.	chitlz
70.	zĕf	76.	scouv	82.	būthĭng	88.	thood
71.	gĭsh	77.	snĭng	83.	tāfs		

Part II (89-120)

89. The girl with brown hair is going.

90. Betty is playing the very pretty songs.

91. Pretty Jean Ann is a little girl.

92. Some shelves are tall, pretty and new.

93. Jim will give Jane the big bottle.

94. Cooking smells make the boys very hungry.

95. The saddle was lifted by the boy.

96. The boy was lifting the big saddle.

97. Are the books on the shelves?

*Rewind tape here to beginning of Part II

98. The girl with brown hair is going. (ĕrl)

99. Betty is playing the very pretty songs. (z)

100. Pretty Jean Ann is a little girl. (z)

101. Some shelves are tall, pretty and new. (s)

102. Jim will give Jane the big bottle. (tl)

103. Cooking smells make the boys very hungry. (ng)

104. The saddle was lifted by the boy (dl)

105. The boy was lifting the big saddle. (z)

106. Are the books on the shelves? (s)

*Rewind tape here to beginning of Part II

107. The girl with brown hair is going. (z)

108. Betty is playing the very pretty songs. (ng)

109. Pretty Jean Ann is a little girl. (tl)

110. Some shelves are tall, pretty and new. (lvz)

Jim will give Jane the big bottle. (on tape but no score)

111. Cooking smells make the boys very hungry. (z)

112. The saddle was lifted by the boy. (z)

113. The boy was lifting the big saddle. (ng.)

114. Are the books on the shelves? (lvz)

*Rewind tape here to beginning of Part II

115. The girl with brown hair is going. (ng)

116. Betty is playing the very pretty songs. (s)

117. Pretty Jean Ann is a little girl. (èrl)

Some shelves are tall, pretty and new. (on tape but no score)

Jim will give Jane the big bottle (on tape but no score)

118. Cooking smells make the boys very hungry. (z)

The saddle was lifted by the boy. (on tape but no score)

119. The boy was lifting the big saddle. (dl)

Are the books on the shelves? (on tape but no score)

*Rewind tape to beginning of Part II

The girl with brown hair is going. (on tape but no score)

120. Betty is playing the very pretty songs. (z)

Pretty Jean Ann is a little girl. (on tape but no score)

Some shelves are tall, pretty and new. (on tape but no score)

Jim will give Jane the big bottle. (on tape but no score)

Cooking smells make the boys very hungry. (on tape but no score)

The saddle was lifted by the boy. (on tape but no score)

The boy was lifting the big saddle. (on tape but no score)

Are the books on the shelves? (on tape but no score)

Appendix B

Supplementary Tables

Table A

Nonsense Words, Initial Sounds: Means, Standard Deviations, and Range for Year 1 and Year 2 by Age and Sex

Group	N	Mean	SD	Range
Year 1				
42-44 mo.	62	-0.39	1.12	-4.09 to 1.76
45-47 mo	211	-0.21	1.06	-4.09 to 1.76
48-50 mo.	265	-0.07	1.01	-3.01 to 2.00
51-53 mo.	294	0.16	0.94	-2.79 to 2.58
54-56 mo.	223	0.13	0.89	-3.02 to 2.20
57-59 mo.	46	0.27	1.01	-2.36 to 2.00
Boys	571	-0.10	1.02	-4.09 to 2.58
Girls	530	0.11	0.96	-4.09 to 2.38
Total	1101	0.00	1.00	-4.09 to 2.58
Year 2				
51-53 mo.	52	0.13	1.00	-2.89 to 1.61
54-56 mo.	168	-0.08	1.30	-2.96 to 1.89
57-59 mo.	146	0.04	0.94	-2.71 to 1.89
60-62 mo.	199	0.03	1.07	-2.96 to 2.34
63-65 mo.	137	-0.03	0.99	-2.96 to 1.89
66-69 mo.	9	-0.01	0.61	-1.19 to 0.83
Boys	371	-0.13	1.02	-2.96 to 2.09
Girls	340	0.15	0.96	-2.96 to 2.34
Total	711	0.00	1.00	-2.96 to 2.34

Table B

Nonsense Words, Medial Sounds: Means, Standard Deviations, and Range for Year 1 and Year 2 by Age and Sex

Group	N	Mean	SD	Range
Year 1				
42-44 mo.	62	-0.37	1.07	-4.40 to 1.54
45-47 mo.	214	-0.24	1.08	-4.65 to 1.78
48-50 mo.	265	-0.06	1.05	-4.15 to 1.89
51-53 mo.	294	0.15	0.92	-2.68 to 1.89
54-56 mo.	224	0.14	0.89	-2.67 to 2.04
57-59 mo.	46	0.33	0.85	-1.63 to 2.04
Boys	572	-0.08	1.02	-4.40 to 2.04
Girls	533	0.08	0.97	-4.65 to 2.04
Total	1105	0.00	1.00	-4.65 to 2.04
Year 2				
51-53 mo.	52	0.11	0.75	-1.50 to 1.64
54-56 mo.	168	-0.10	1.04	-2.73 to 2.73
57-59 mo.	146	0.02	0.98	-2.56 to 3.00
60-62 mo.	198	0.08	0.97	-3.62 to 2.45
63-65 mo.	137	-0.06	1.10	-2.91 to 2.45
66-69 mo.	9	0.06	0.90	-1.09 to 1.91
Boys	371	-0.06	1.00	-2.91 to 2.73
Girls	339	0.06	1.00	-3.62 to 3.00
Total	710	0.00	1.00	-3.62 to 3.00

Table C

Nonsense Words, Final Sounds: Means, Standard Deviations, and Range for Year 1 and Year 2 by Age and Sex

Group	N	Mean	SD	Range
Year 1				
42-44 mo.	62	-0.18	0.96	-2.68 to 1.94
45-47 mo.	211	-0.17	1.09	-2.68 to 2.54
48-50 mo.	265	-0.03	0.97	-2.49 to 2.66
51-53 mo.	293	0.06	0.93	-2.49 to 2.66
54-56 mo.	222	0.10	1.03	-2.09 to 2.95
57-59 mo.	47	0.29	0.94	-1.56 to 2.29
Boys	570	-0.09	1.00	-2.68 to 2.95
Girls	530	0.10	1.00	-2.31 to 2.73
Total	1100	0.00	1.00	-2.68 to 2.95
Year 2				
51-53 mo.	53	0.16	0.95	-2.38 to 2.39
54-56 mo.	167	-0.02	1.07	-2.38 to 2.39
57-59 mo.	146	-0.02	1.00	-2.17 to 2.39
60-62 mo.	199	0.09	0.97	-2.38 to 2.80
63-65 mo.	137	-0.14	0.98	-2.38 to 2.19
66-69 mo.	9	-0.01	0.96	-1.12 to 1.78
Boys	372	-0.07	0.99	-2.38 to 2.39
Girls	339	0.08	1.01	-2.38 to 2.80
Total	711	0.00	1.00	-2.38 to 2.80

Table D

Nonsense Words, Long Vowels: Means, Standard Deviations, and Range for Year 1 and Year 2 by Age and Sex

Group	N	Mean	SD	Range
Year 1				
42-44 mo.	61	-0.40	1.10	-3.30 to 2.06
45-47 mo.	214	-0.16	1.06	-4.35 to 2.20
48-50 mo.	264	-0.07	1.08	-4.87 to 2.20
51-53 mo.	294	0.10	0.94	-2.98 to 2.06
54-56 mo.	223	0.13	0.86	-2.98 to 2.06
57-59 mo.	45	0.38	0.75	-1.19 to 1.55
Boys	568	-0.04	1.02	-4.87 to 2.20
Girls	533	0.04	0.98	-4.35 to 2.20
Total	1101	0.00	1.00	-4.87 to 2.20
Year 2				
51-53 mo.	51	0.00	0.95	-2.66 to 1.81
54-56 mo.	171	-0.10	1.10	-2.75 to 1.81
57-59 mo.	149	-0.03	1.06	-3.36 to 2.38
60-62 mo.	199	0.13	0.82	-1.61 to 2.38
63-65 mo.	136	-0.06	1.06	-3.36 to 2.38
66-69 mo.	9	-0.27	0.85	-1.04 to 1.52
Boys	373	-0.02	0.97	-3.36 to 2.38
Girls	342	0.02	1.03	-2.75 to 2.38
Total	715	0.00	1.00	-3.36 to 2.38

Table E

Nonsense Words, Short Vowels: Means, Standard Deviations, and Range for Year 1 and Year 2 by Age and Sex

Group	N	Mean	SD	Range
Year 1				
42-44 mo.	67	-0.26	1.10	-4.45 to 1.32
45-47 mo.	223	-0.19	1.07	-3.98 to 1.78
48-50 mo.	274	-0.04	1.03	-3.98 to 2.05
51-53 mo.	301	0.13	0.91	-2.30 to 1.79
54-56 mo.	226	0.10	0.95	-3.50 to 1.79
57-59 mo.	48	0.19	0.84	-1.39 to 1.78
Boys	589	-0.07	1.01	-4.45 to 2.05
Girls	550	0.07	0.98	-3.98 to 1.79
Total	1139	0.00	1.00	-4.45 to 2.05
Year 2				
51-53 mo.	54	0.17	0.81	-2.55 to 1.13
54-56 mo.	174	-0.11	1.06	-3.17 to 2.42
57-59 mo.	153	0.05	0.93	-2.55 to 2.42
60-62 mo.	203	0.02	1.00	-4.40 to 2.42
63-65 mo.	140	-0.02	1.07	-2.55 to 1.98
66-69 mo.	9	0.08	0.92	-1.09 to 1.98
Boys	386	-0.06	1.05	-3.28 to 2.42
Girls	347	0.07	0.95	-4.40 to 2.42
Total	733	0.00	1.00	-4.40 to 2.42

Table F

Meaningful Words in Phrases, Model Word or Some Semblance of It:
Means, Standard Deviations, and Range
for Year 1 and Year 2 by Age and Sex

Group	N	Mean	SD	Range
Year 1				
42-44 mo.	51	-0.53	1.20	-4.47 to 1.28
45-47 mo.	173	-0.25	1.11	-5.85 to 1.51
48-50 mo.	232	-0.13	1.01	-5.45 to 1.61
51-53 mo.	255	0.18	0.89	-3.39 to 1.74
54-56 mo.	199	0.18	0.89	-4.24 to 1.47
57-59 mo.	44	0.41	0.73	-1.23 to 1.47
Boys	485	-0.08	1.01	-4.24 to 1.74
Girls	469	0.08	0.98	-5.85 to 1.61
Total	954	0.00	1.00	-5.85 to 1.74
Year 2				
51-53 mo.	18	-0.18	1.11	-1.66 to 1.33
54-56 mo.	56	0.05	1.05	-1.91 to 1.92
57-59 mo.	48	-0.06	0.92	-1.84 to 1.83
60-62 mo.	81	0.02	0.96	-2.16 to 1.92
63-65 mo.	59	-0.08	0.94	-2.16 to 1.68
66-69	5	-0.16	1.30	-2.16 to 1.45
Boys	116	-0.24	0.90	-2.16 to 1.92
Girls	151	0.14	1.01	-2.16 to 1.92
Total	267	-0.02	0.98	-2.16 to 1.92

References

Baratz, J. A., A bi-dialectical task for determining language proficiency in economically disadvantaged Negro children. *Child Development*, 1969, *40*, 889-901.

Briere, E. J. Phonological testing reconsidered. *Language Learning*, 1967, *17*, 163-171.

Ervin-Tripp, S. M. Imitation and structural changes in children's language. In Lenneberg, E. H., (Ed.), *New directions in the study of language*. Cambridge, Mass.: Massachusetts Institute of Technology Press, 1964.

Fraser, C., Bellugi, U., & Brown, R. Control of grammar in imitation, comprehension and production. *Journal of Verbal Learning and Verbal Behavior*, 1963, *2*, 121-135.

Friedlander, G. H. Report on the articulatory and intelligibility status of socially disadvantaged preschool children, Head Start Research and Evaluation Office, 1965. (ED 014 321)

Goetzinger, C. P., Dirks, D. D. & Baer, C. J. Auditory discrimination and visual perception in good and poor readers. *Annals of Otology, Rhinology, and Laryngology*, 1960, *69*, 121-136.

Hall, V. C. & Turner, R. R. Comparison of imitation and comprehension scores between two lower-class groups and the effects of two warm-up conditions on imitation of the same groups. *Child Development*, 1971, *42*, 1735-1750.

Kamil, M. L. & Rudegeair, R. E. Methodological improvements in the assessment of phonological discrimination in children. *Child Development*, 1972, *43*, 1087-1091.

Keeney, T. Young children's imitation of nonsense syllable strings, word strings, and grammatical sentences. Princeton, N. J.: Educational Testing Service, [illegible], RB-69-70.

Lovell, K. & Dixon, E. The growth of the control of grammar in imitation, comprehension, and production. *Journal of Child Psychology and Psychiatry*, 1967, *8*, 31-40.

Menyuk, P. A preliminary evaluation of grammatical capacity in children. *Journal of Verbal Learning and Verbal Behavior*, 1963, *2*, 429-439.

Menyuk, P. *The acquisition and development of language*, Englewood, N. J.: Prentice-Hall, 1971.

Osser, H., Wang, N. D., & Zaid, F. The young child's ability to imitate and comprehend speech: A comparison of two subcultural groups. *Child Development*, 1969, *40*, 1063-1075.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Slobin, D. I. & Welsh, C. A. Elicited imitation as a research tool in developmental psycholinguistics. An unpublished paper. University of California at Berkeley, 1967.

Stern, C. Comparative effectiveness of echoic and modeling procedures in language instruction with culturally disadvantaged children. Head Start Research and Evaluation Office, 1967. (ED 025 314)

Van Riper, C. Speech correction. Principles and methods. 5th Edition. Englewood Cliffs, N. J.: Prentice-Hall, 1972.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

YOUNG CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

A LONGITUDINAL STUDY

Technical Report 11

Matching Familiar Figures Test

William C. Ward

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Matching Familiar Figures Test

Background

The Matching Familiar Figures Test is a measure of the response style "reflection-impulsivity." On tasks where there are several response alternatives and some uncertainty as to which is correct, some individuals--reflectives--typically take time to consider their possible responses, and therefore have a relatively low error rate; others--impulsives--respond quickly and with a higher proportion of errors (Kagan, Rosman, Day, Albert & Phillips, 1964).

Several classes of explanations have been proposed for these differences. The child's concern over the quality of his performance may be crucial for reflective performance; this possibility has been expressed both in positive terms, as degree of ego-involvement (Hess, Shipman, Brophy & Bear, 1969), and negatively, as anxiety over the possibility of failure (Kagan, 1965a). It has also been suggested that impulsive performance is a cognitive manifestation of an early, perhaps genetic, dimension of response tempo, seen in the first years of life in such variables as rapid habituation to a repeated stimulus and in frequent changes of activity in a free-play situation (Kagan & Kogan, 1970). The former type of explanation has received some support from studies showing that failure experiences lead to relatively more reflective behavior in both reflective and impulsive children (Messer, 1970; Reali & Hall, 1970), and by a demonstration of more "nonmeaningful" responding by impulsives (Hess et al., 1969). With respect to the genetic explanation, only speculative interpretations (Kagan & Kogan, 1970) of limited data are available. Moreover, the lack of relation of impulsivity scores to response latency in tasks not involving simultaneously available response alternatives (Kagan et al., 1964), and the lack of strong relations to several less cognitively based indices of impulse control (Hess et al., 1969; Ward, 1968a), raise some doubts concerning this explanation.

Response latency on tests of reflection-impulsivity has been found to be nearly independent of IQ, although errors are a function both of the stylistic variable and of ability. There appears to be a positive association between reflectiveness and social class (Eska & Black, 1971; Hess et al., 1969). It has been suggested that the dimension is of greater relevance to cognitive performance for males than for females (Lewis, Rausch, Goldberg & Dodd, 1968), but the evidence is weak (Eska & Black, 1971).

Reflectiveness is related to performance on tests of reasoning (Kagan, Pearson & Welch, 1966) and of word reading (Kagan, 1965b) in early elementary children. Its implications for performance in children below school age remain largely unexplored, but the dimension has been found to be present in kindergarten children of diverse SES (Ward, 1968b), and in middle-class nursery school children (Katz, 1971; Lewis et al., 1968). A measure of the dimension was included in the present battery, along with several other measures of impulse expression and control, to enable assessment of the generality and dimensionality of impulsivity in young disadvantaged children, and of its implications for cognitive performance at this age.

Task Description

The version of the test used in the present battery was developed by Lewis et al. (1968), and used by them with middle-class three-year-olds. The test consists of two practice and eighteen test items. On each item the child is shown one standard and four comparison figures. Figures are simple line drawings done in black on a white background; five items show animals, five show humans, seven use common objects, and three use geometric designs. In each case one of the comparison figures is identical to the standard, while each of the remaining figures differs from the standard in some detail.

As the test has been given in the Longitudinal Study, on each item the child is first shown the set of comparison figures, and is asked to look in turn at each figure. He is then given the standard and must point to the one figure among the four which is identical to it. Latency to first choice and number of errors (to a maximum of two per item) are recorded. Feedback as to errors is given after the first error for an item, but the child's second choice is accepted regardless of accuracy.

Testing time is about ten minutes. The test is not difficult to administer, but requires accurate use of a stopwatch, careful avoidance of giving the child cues as to which of the alternatives is correct, and establishment of a rhythm in item presentation which assures that the instruction is completed, the standard presented, and the stopwatch started, all simultaneously.

Scoring

Two scores are obtained: mean response time and mean number of errors. In the present data, the latencies were transformed by log (X + 1) before averaging, since their distributions were positively skewed, and it appeared desirable to decrease the effect of occasional very long latencies on the score. (Year 1 latencies were Windsorized, large values reduced to an arbitrary value, 20 seconds, before the log transformation was applied. Since response times longer than 20 seconds were very rare, about 0.1% of all responses, this precaution was unnecessary, and it was not repeated with Year 2 data.) In both years mean errors were expressed on a per-item basis, so that spoiled items could be eliminated from the average for a subject without affecting his possible error score.

Score Characteristics

For the various items, in Year 1 data the number of subjects whose first response was correct ranged from 40 to 84 percent, with a median of 49 percent. In Year 2, the range was from 39 to 90 percent, with a median of 59 percent. The correct alternative was the modal first response for all eighteen test items in Year 1, and for seventeen of them in Year 2. The test, therefore, appeared to possess an appropriate difficulty level for the present sample, though several of the items had very low error rates in the Year 2 data.

The mean latency and error scores possessed satisfactory internal consistency; the former was the more reliable score. For the sample as a whole, coefficient alpha for latencies was .90 in Year 1 and .91 in Year 2. For errors, the alphas were .70 and .71. When the sample was broken down by testing site, (each year), or by sex, race, age, or SES level (Year 1 only), subgroup reliabilities remained high, ranging from .87 to .92 for latencies and .63 to .76 for errors.

Sample Performance

Means and percentiles for the total group and for age and sex subgroups are presented in Tables 1-4. Tables 1 and 2 contain latency data for Years 1 and 2, respectively. The overall mean in each year, 0.60 in log transformed form, is equivalent to a mean of 2.93 seconds, indicating that these children spent little time in "reflecting" over alternative solutions. The percentile data in Tables 1 and 2 show that the logarithmic transformation did succeed in eliminating the skewness in the latency distribution.

Table 1

Mean Response Time by Sex and Three-Month Age Intervals (Transformed by T = Log[T + 1]) in Year 1

Group	N	Mean	SD	10	25	Percentiles 50	75	90
42-44 mo.	83	0.61	0.11	0.45	0.54	0.62	0.69	0.74
45-47 mo.	293	0.60	0.12	0.46	0.52	0.59	0.68	0.75
48-50 mo.	332	0.60	0.12	0.45	0.51	0.59	0.67	0.75
51-53 mo.	369	0.59	0.12	0.44	0.51	0.59	0.67	0.75
54-56 mo.	261	0.58	0.11	0.45	0.51	0.59	0.66	0.72
57-59 mo.	61	0.56	0.10	0.43	0.48	0.60	0.64	0.69
Boys	732	0.60	0.12	0.46	0.52	0.60	0.68	0.75
Girls	667	0.59	0.11	0.44	0.50	0.58	0.66	0.73
Total	1399	0.60	0.12	0.45	0.51	0.59	0.67	0.74

Table 2

Mean Response Time by Sex and Three-Month Age Intervals (Transformed by T = Log[T + 1]) in Year 2

Group	N	Mean	SD	10	25	Percentiles 50	75	90
51-53 mo.	81	0.58	0.10	0.45	0.50	0.58	0.65	0.72
54-56 mo.	313	0.60	0.11	0.47	0.53	0.60	0.68	0.74
57-59 mo.	303	0.61	0.12	0.45	0.53	0.60	0.68	0.76
60-62 mo.	347	0.60	0.11	0.46	0.53	0.60	0.68	0.75
63-65 mo.	244	0.61	0.12	0.46	0.54	0.61	0.68	0.76
66-69 mo.	16	0.57	0.11	0.43	0.47	0.57	0.66	0.74
Boys	692	0.60	0.12	0.45	0.52	0.59	0.67	0.75
Girls	612	0.61	0.11	0.47	0.54	0.61	0.68	0.75
Total	1304	0.60	0.11	0.46	0.53	0.60	0.68	0.75

Tables 3 and 4 contain error data for Years 1 and 2, respectively. The mean error scores were .61 per item in Year 1, and .46 in Year 2.

Table 3

Mean Number of Errors per Item* by Sex and Three-Month Age Intervals in Year 1

				Percentiles				
Group	N	Mean	SD	10	25	50	75	90
42-44 mo.	83	0.72	0.35	0.26	0.42	0.70	1.00	1.18
45-47 mo.	293	0.66	0.29	0.32	0.42	0.63	0.86	1.11
48-50 mo.	332	0.66	0.31	0.28	0.40	0.62	0.84	1.12
51-53 mo.	369	0.56	0.29	0.23	0.35	0.54	0.75	0.93
54-56 mo.	261	0.55	0.30	0.19	0.31	0.49	0.78	0.93
57-59 mo.	61	0.52	0.26	0.19	0.32	0.51	0.70	0.83
Boys	732	0.62	0.30	0.25	0.38	0.61	0.82	1.07
Girls	667	0.60	0.30	0.25	0.37	0.57	0.79	1.05
Total	1399	0.61	0.30	0.25	0.37	0.59	0.81	1.07

*Range = 0-2.

Table 4

Mean Number of Errors per Item* by Sex and Three-Month Age Intervals in Year 2

				Percentiles				
Group	N	Mean	SD	10	25	50	75	90
51-53 mo.	81	0.52	0.25	0.24	0.32	0.47	0.64	0.85
54-56 mo.	313	0.52	0.28	0.20	0.30	0.46	0.71	0.91
57-59 mo.	303	0.46	0.25	0.17	0.26	0.41	0.62	0.82
60-62 mo.	347	0.44	0.26	0.16	0.24	0.38	0.60	0.81
63-65 mo.	244	0.40	0.26	0.13	0.22	0.34	0.55	0.78
66-69 mo.	16	0.40	0.24	0.16	0.21	0.32	0.55	0.72
Boys	692	0.49	0.28	0.17	0.27	0.42	0.67	0.87
Girls	612	0.43	0.24	0.16	0.24	0.39	0.58	0.77
Total	1304	0.46	0.27	0.17	0.26	0.40	0.62	0.83

*Range = 0-2.

Latencies and errors were each examined by analysis of variance to determine whether there were differences associated with age (median split within each year's data), sex, or SES (mother's education level--less than 10, 10 to 12, or more than 12 years of schooling) in the data. For latencies, repeated-measures analysis of variance, including all children tested in both years on this measure, showed significant effects of sex (F = 11/62, df = 1/1042, $p < .001$) and of SES (F = 5.70, df = 2/1042, $p < .005$); males and children of higher SES had longer latencies. In each case, however, the differences converted to seconds are of trivial magnitude, less than 0.1 seconds. There was no significant change in latency from the first to the second year of testing. Age by sex by SES analyses for each year's data separately showed marginal age (F = 5.40, df = 1/1074, $p < .025$) and SES (F = 3.64, df = 2/1074, $p < .05$) effects, but a highly significant sex difference (F = 10.95, df = 1/1074, $p < .001$) in Year 1, with younger children, boys and higher SES children obtaining the longer latencies. Again these differences were of trivial magnitude. No significant differences in latency were found in Year 2.

Repeated-measures analysis of the error data showed significant effects attributable to sex (F = 9.98, df = 1/1042, $p < .005$), SES (F = 70.44, df = 2/1042, $p < .001$), and year (F = 348.05, df 1/1046, $p < .001$); more errors were found in males, low SES children, and data obtained in Year 1. There was, in addition, a year by SES interaction (F = 6.17, df = 2/1042, $p < .005$); while the ordering of means for the three SES groups did not change from one year to the next, the lowest SES group showed the least change in error scores, and the middle group the most. Age by sex by SES analyses on each year's data showed significant effects for all three variables in both years, with older children (F = 45.21, df = 1/1074,

$p < .001$; $F = 30.33$, $df = 1/1123$, $p < .001$, in Years 1 and 2, respectively), girls ($F = 5.62$, $df = 1/1074$, $p < .025$ in Year 1 and $F = 41.07$, $df = 1/1123$, $p < .001$ in Year 2) and higher SES children ($F = 41.07$, $df = 2/1074$, $p < .001$; $F = 72.79$, $df = 2/1123$, $p < .001$, in Years 1 and 2, respectively) making fewer errors.

Score intercorrelations based on the three-site longitudinal sample are presented in Table 5. Within years, latencies and errors had little relation in Year 1 ($r = -.07$) and a low but significant negative correlation in Year 2 ($r = -.27$). Between-years consistency was low for latencies ($r = .22$) and somewhat higher for errors ($r = .43$).

Table 5

Correlations Among Latency and Error Scores

	Year 1 Errors	Year 2 Latency	Year 2 Errors
Year 1 Latency	-.07	.22***[a]	-.00
Year 1 Errors		-.13***	.43***
Year 2 Latency			-.27***

Note.--N's range from 727 to 790.

[a]*** $p < .001$

Construct Validity

Year 1. Despite the internal consistency of the latency and error measures, there was no evidence for the existence of the reflection-impulsivity dimension in the present data. Latencies and errors had a negligible correlation. Thus, those children whose response times were longer were not using the additional time for more thorough processing of the information available. This result could be explained simply if the task had been too easy for these children; however, the mean error score (.61 errors per item) corresponded very closely to the error rate obtained by Lewis et al. (1968) in a sample of bright middle-class three-year-olds--12.9 errors over 20 items, or .65 errors per item.

As in previous work, MFF errors showed significant relations to measures of intellectual competence and achievement; e.g., correlating -.42 with Preschool Inventory scores. They were associated with both verbal and nonverbal indices, correlating for example -.43 with both the Peabody Picture Vocabulary Test, Form A, and the Johns Hopkins Perceptual Test (whose format is virtually identical to that of MFF). These correlations fall within the range previously reported for error-IQ association in other studies (Eska & Black, 1971). They provide further evidence that the task was of an appropriate difficulty level for the children in the present sample.

There was no evidence for clustering of the three instruments--MFF, Motor Inhibition Test, and Mischel Technique (Delay of Gratification)--included in the test battery as possible indices of a broader impulse-control dimension. MFF errors correlated significantly ($r = -.22$) with the measure of motor control obtained from the Motor Inhibition Test, but this relation apparently derived from the correlation of each of these measures with the larger general ability factor. Partialling out the Preschool Inventory score, for example, reduced their correlation to -.08. Otherwise, Motor Inhibition, Delay of Gratification, and MFF latencies and errors all had essentially zero intercorrelations.

There was, however, evidence for consistency in response tempo which was not associated with competence of performance. MFF latencies correlated .47 with latency to choice of first object in the Sigel Object Categorization Task, and .19 with latency of first response in the Preschool Embedded Figures Test, while these measures correlated .22 with one another. In the factor analyses for the total sample, and for the analyses of sub-samples, these three measures defined a second factor which was independent of the first, competence-of-performance factor. Thus, somewhat consistent with Kagan's speculations as to the existence of an early-appearing tempo dimension, it appears that a cognitive tempo dimension is to be found in young disadvantaged children prior to the point in development at which reflection-impulsivity has implications for quality of performance. The meaning of this finding, and its integration with the different results obtained in the several studies of reflection-impulsivity in middle-class preschool children, remains to be clarified.

<u>Year 2</u>. As in Year 1 data, MFF errors correlated significantly with other measures of intellectual competence and achievement. This score had its highest correlations with the Preschool Inventory ($\underline{r}$ = -.54), with the Peabody Picture Vocabulary Test, Forms A and B (-.50), and with fastest time on the Seguin Form Board (.49). The same variables, including both verbal and perceptual measures, tended to show significant

correlations with errors in the two years of testing, though the Year 2 correlations were generally higher. For example, only four scores correlated as high as .4 with errors in Year 1 data, while 10 (including these four) correlated .4 or better in Year 2.

Year 2 results showed a significant negative relation between latencies and errors, indicating that longer response times were associated with more careful decisions as to which alternative is correct. Unlike the Year 1 data, therefore, these children are showing individual differences along the reflectiveness-impulsiveness dimension. It should be noted, however, that the error correlations with ability measures, along with the relatively low level of intercorrelation of latencies with errors, suggest that stylistic factors are only one, and perhaps not the major, determinant of performance on the test at this age. Moreover, while individuals can be classified as relatively reflective or impulsive, the majority of individuals were still responding very quickly in absolute terms; for example, 95% of the sample had mean latencies of less than 5.4 seconds.

The search for a general impulse-regulation dimension in the Year 2 data produced results parallel to those from Year 1--Motor Inhibition performance showed a significant correlation with MFF errors (r = -.25), which was reduced to negligible magnitude (-.07) when the Preschool Inventory score, serving as an index of general cognitive functioning, was partialled out. Otherwise, near-zero correlations were found among Motor Inhibition Test, Mischel Delay of Gratification and MFF scores.

There was some replication of the finding from Year 1 of a response tempo dimension independent of competence: latency to first choice on the Sigel Object Categorization Task and latency to first response on the Preschool Embedded Figures Test correlated .20 with one another. These scores correlated with MFF latencies, however, only .09 and .07, respectively. It may be that the lack of latency relations between MFF and the other tests is another index that in Year 2, in contrast to the earlier results, MFF latencies are beginning to serve as an index of cognitive style. In any case, the low interrelations among latency measures in Year 2, combined with low stability coefficients for all the latency measures over the two years (r's across years of .22, .06, and .17, respectively, for MFF, Sigel, and PEFT latencies), suggests that the tempo dimension is a transitory and particular, rather than a stable and broad, aspect of the performance of these children.

Summary

The form of the Matching Familiar Figures Test employed here is of appropriate difficulty level for three- and four-year-old children, and provides both latency and error scores of high reliability. At least some of the items in this version, however, would be too easy for children older than those in the present sample.

In the first year's data from this study, the MFF did not show evidence of measuring the response style reflection-impulsivity; rather, latencies were associated with latencies on several other tasks, suggesting a [illegible]ral response tempo dimension unrelated to competence of performance, while errors were strongly related to other measures of competence and achievement. In the second year of the study, errors remained an index of general level of functioning, even increasing their correlations with other competence measures; but there was also some evidence for the emergence of the response style the test was designed to measure. Data from future years of the study will be examined with the expectation that even stronger evidence of stylistic variance will be found, suggesting that the cognitive style is developing at a slower rate in these children than in others that have been reported (generally with bright, more advantaged samples) using preschool children.

No evidence was found for a general dimension of impulse regulation consistent across diverse measures. Other investigators have reported significant correlations, for example, between MFF errors and Motor Inhibition Test performance (e.g., Ward, 1968a); but the correlations have never been of impressive magnitude. In this respect the present data are at least not strongly discrepant with others' results.

Further analyses will be directed at discovering whether reflectiveness has implications for cognitive performance which do not appear in the single-variable correlations examined thus far (e.g., in studying children classified as "slow," "fast," "reflective" and "impulsive"), and to the study of personal-social, family background, and instructional correlates of a reflective or impulsive cognitive style.

References

Eska, B. & Black, K. N. Conceptual tempo in young grade-school children. Child Development, 1971, 42, 505-516.

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, 1969.

Kagan, J. Impulsive and reflective children: The significance of conceptual tempo. In J. D. Krumboltz (Ed.), Learning and the educational process. Chicago: Rand-McNally, 1965. (a)

Kagan, J. Reflection-impulsivity and reading ability in primary grade children. Child Development, 1965, 36, 609-628. (b)

Kagan, J. & Kogan, N. Individuality and cognitive performance. In Paul H. Mussen (Ed.), Carmichael's manual of child psychology. 3rd Ed. (Vol. 1). New York: Wiley, 1970.

Kagan, J., Pearson, L. & Welsh, L. Conceptual impulsivity and inductive reasoning. Child Development, 1966, 37, 583-594.

Kagan, J., Rosman, B. L., Day, D., Albert, J. & Phillips, W. Information processing in the child: Significance of analytic and reflective attitudes. Psychological Monographs, 1964, 78, (1, Whole No. 578).

Katz, J. M. Reflection-impulsivity and color-form sorting. Child Development, 1971, 42, 745-754.

Lewis, M., Rausch, M., Goldberg, S. & Dodd, C. Error, response time and IQ: Sex differences in the cognitive style of preschool children. Perceptual and Motor Skills, 1968, 26, 563-568.

Messer, S. The effect of anxiety over intellectual performance on reflection-impulsivity in children. Child Development, 1970, 41, 723-735.

Reali, N. & Hall, V. Effect of success and failure on the reflective and impulsive child. Developmental Psychology, 1970, 3, 392-402.

Ward, W. C. Creativity in young children. Child Development, 1968, 39, 737-754. (a)

Ward, W. C. Reflection-impulsivity in kindergarten children. Child Development, 1968, 39, 867-874. (b)

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

YOUNG CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

A LONGITUDINAL STUDY

Technical Report 12

Mischel Technique

David R. Lindstrom
Virginia C. Shipman

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Mischel Technique

Early theorizing concerning the ability to delay gratification grew out of psychoanalytic thought and clinical judgments regarding personality development. Specifically, the ability to delay gratification was theorized to be related to the transition from "primary process thinking" to governance of the "reality principle" and development of ego boundaries. Psychoanalytic theory suggests that inhibition or delay of expression of basic impulses is an important factor in early adjustment. If such theorizing is correct, the ability to delay gratification should be related to socializing influences and other indexes of impulsivity.

Typically, experimental investigation of the ability to delay gratification has involved presenting Ss with a choice situation: S can choose to have a small reward immediately or a larger reward at a later promised time. In general, the studies reported below followed this procedure. Variations include length of time intervals, group administrations and the value of reward.

As expected, older Ss have been found to choose the delayed reward more often than younger Ss (Bialer, 1961, ages 7-14; Hess, Shipman, Brophy & Bear, 1969, ages 7, 8; Klein, 1967, ages 8, 12 and adult; Mischel, 1958, ages 7-9; Mischel & Metzner, 1962, ages 5-12; Shipe & Lazare, 1969, ages 4 to 6). IQ and mental age have been shown to relate positively with choice of the delayed reward (Bialer, 1961; Hess et al., 1969; Mischel & Metzner, 1962), although no relationship with IQ was found in one study (Klein, 1967). This latter finding may have been due to the wide age range of the sample (second grade to adult) which also included institutionalized schizophrenic Ss. Social class differences have been reported by Hess et al., 1969;

MacKenzie, 1970; Walls and Smith, 1970; and Weintraub, 1969, with lower-class Ss tending to make more immediate choices than middle-class Ss. However, using a very crude index of status (tin vs. tiled roof in Trinidad), Mischel (1958) and Kunz and Moyer (1969) report no social class differences in choice behavior. Sex differences appear to be negligible (Mischel, 1961; Mischel & Metzner, 1962; Strickland, 1972), with the exception of one study in which girls chose the delayed reward more than boys (Wohlford, Clark & Trunfio, 1972). Mischel (1958) postulated that delay of gratification may be related to the absence of a father in the home and found that Ss from intact homes made more delay choices than Ss from fatherless homes. Similar results are reported by Hess et al. (1969) with SES controlled. Trunfio and Wohlford (1972) reported father-absent boys were less able to delay than father-present boys, but no differences were reported for girls.

If the ability to delay gratification is viewed as an index of a unitary trait of impulsivity, it should relate to other indices of impulse control. Hess et al. (1969) and Kahana and Kahana (1970) report low but positive correlations between preference for delayed reward and measures of impulse control (e.g., Draw-A-Circle Slowly); however, MacKenzie (1970) reports no relationship between choice and ability to inhibit motor activity. Choice of the delayed reward also has been shown to be related to internal locus of control in three studies (Bialer, 1961; Phypers, 1970; Walls & Smith, 1970), but has shown no relationship in another (Zytkoskee, Strickland & Watson, 1971). Strickland (1972), however, found this relationship to hold for white Ss only. Mothers' internal locus of control also has been related to the child's preference for the delayed reward (Hess et al., 1969).

Academic variables studied in relation to delay of gratification include achievement scores (Hess et al., 1969; Mendell, 1967) and ratings of school adjustment (Phypers, 1970). These studies indicated that preference for delayed reward is associated with higher grades and better school adjustment. Social variables reflecting social responsibility (scores on the Harris "Social Attitudes Scale") have also been shown to relate to delay of gratification (Mischel, 1961).

Individual consistency of preference is reported in very few studies. One study by Mischel (1961) does indicate consistency of choice for an adolescent Trinidadian sample across three choice tasks (two written questions expressing willingness to delay and one actual choice situation). Shipe and Lazare (1969) report a correlation of .20 between two actual choice situations for a group of four- to six-year-old children. An attempt to change Ss' choices from immediate to delayed preferences is reported by Staub (1972). Exposure to a model who displayed delay behavior was effective in changing preferences in males only. Clearly, more research is needed to examine individual consistency in response and determinants of change over time.

Thus, some studies have reported positive relationships between the ability to delay gratification and age, IQ, social class, absence of father from the home, internal locus of control, academic variables and other measures of impulsivity. The measure of delay of gratification was included in the Longitudinal Study to permit further exploration of developmental change from immediate to delayed reward, processes concomitant with choice behaviors, and its relationship with other purported measures of impulse control.

Task Description

The procedure used in Years 1 and 2 of the present study was an adaptation (Hess et al. 1969) for young children of the technique devised by Mischel (1958). Adaptations involved (1) asking the child to identify the larger of two pieces of candy to facilitate comprehension of the rewards; (2) specifying a standard time limit for receipt of the delayed reward which would be comprehensible to a young child; and (3) asking for a rationale for the choice in order to further understand the dynamics of preferences.

E presented the child with a large and a small piece of Saran-wrapped candy (four and two sections of Tootsie Roll). S was asked to identify the one with "more to eat." E then told S that he could "have this little one right now," or wait and receive the large piece when it was time for him to go home. Following his choice, S was asked the reason for his choice. If the child chose both, or wanted the big one immediately, further explanation was given and a second trial presented.

To standardize the length of delay interval and to facilitate the child's understanding of it, choices were presented early in the test battery and E stressed the immediate ("right now") and delayed ("time to go home") time points. Ss who chose the delayed reward were asked to recall E's instructions when the reward was presented.

Tester Training and Administration

Tester training on the Mischel Technique is relatively easy. Recording of identification and choice is a simple matter; rationales, however, must be recorded verbatim. The candy should always be kept wrapped for hygenic reasons, but cellophane should be used to enable S to see the reward. Testers must remember to give the child the reward when the child makes a delayed choice. It also is essential that testers check first with the mother that the child is allowed candy and that he does not have allergic reactions to it. Testers reported very few problems in administering the test.

Scoring

Scores were obtained for correctness of the child's identification of the large piece, for his choice, the reason for this choice, and for memory of the instructions.

Sample Performance

Inspection of recognition responses indicated that 95% of the Year 1 sample and 98% of the Year 2 sample correctly identified the larger piece. There was no relationship with age or sex.

Tables 1 and 2 present distribution data by age and sex for percent of children making an immediate or delayed choice in Year 1 and Year 2, respectively. There was no relationship with age as defined by three-month intervals in either year.

Table 1

Percentage of Children Choosing Immediate or Delayed Reward by Age and Sex for Year 1

Group	N	Immediate Reward	Delayed Reward	Other*
42-44 mo.	91	37.4	60.4	2.2
45-47 mo.	323	37.5	59.8	2.7
48-50 mo.	340	33.2	64.4	2.4
51-53 mo.	383	36.3	59.8	3.4
54-56 mo.	271	31.7	67.2	1.1
57-59 mo.	63	33.3	65.1	1.6
Boys	785	33.8	63.2	3.0
Girls	686	36.6	61.7	1.7
Total	1471	35.1	62.5	2.5

*Includes wanted both or large one now with no final choice, refusals and indeterminate responses.

Table 2

Percentage of Children Choosing Immediate or Delayed Reward by Age and Sex for Year 2

Group	N	Immediate	Delayed	Other*
51-53 mo.	76	27.6	72.4	0.0
54-56 mo.	211	34.6	64.0	1.4
57-59 mo.	213	42.3	57.3	0.5
60-62 mo.	249	41.8	57.4	0.8
63-65 mo.	146	46.6	52.1	1.4
66-69 mo.	9	33.3	55.6	11.1
Boys	474	39.9	59.3	0.8
Girls	430	39.5	59.3	1.2
Total	904	39.7	59.3	0.9

*Includes wanted both or large one now with no final choice, refusals and indeterminate responses.

Similarly, for longitudinal Ss (those tested in both years), separate ANOVAS for Year 1 and Year 2 data showed no significant differences in choice behavior for children above and below the median age. Repeated-measures analysis of variance (sex x age x SES) for the longitudinal sample indicated no significant differences in choice behavior for sex or Ss classified according to mother's education (more than 12 years of schooling, between 10-12 years, less than 10 years). The correlation between Year 1 and Year 2 choice score was -.02.

Tables 3 and 4 present percentages of children using different choice rationales in Years 1 and 2, respectively. In Year 1 egocentric choice rationales were most commonly given, as might be expected. There was a tendency for test-defined responses to be given more by the older age group

(51-59 mos), but this appears to be the only age trend apparent in the data. Sex differences were negligible with the exception that boys gave more "seeming irrelevance" responses and girls made more references to home in their rationales. Approximately 72% of the sample did produce a relevant rationale which suggests task demands were readily understood.

Table 3

Percentage of Children Using Different Choice Rationales by Age and Sex for Year 1

Group	N	Rationale Categories* 2	3	4	5	6	7	8	9
42-44 mo.	89	32.6	5.6	4.5	0.0	21.3	5.6	2.2	28.1
45-47 mo.	320	35.3	1.6	3.1	1.6	16.6	8.1	5.0	28.8
48-50 mo.	338	39.3	0.9	3.3	0.3	20.4	6.8	4.1	24.9
51-53 mo.	378	39.9	0.8	5.6	0.3	24.1	7.9	2.1	19.3
54-56 mo.	270	29.6	1.9	3.7	0.0	29.6	5.6	5.9	23.7
57-59 mo.	63	33.3	1.6	1.6	0.0	27.0	6.3	1.6	28.6
Boys	777	35.8	1.4	3.6	0.4	22.4	7.2	4.5	24.7
Girls	681	36.6	1.6	4.3	0.6	22.8	6.9	3.2	24.1
Total	1458	36.1	1.5	3.9	0.5	22.6	7.1	3.9	24.4

*Rationale categories

2 = Egocentric ("I like it." "I want to.")

3 = Family member/tester used as determinant ("my mother/tester told me to")

4 = Home (to share with or show to others)

5 = Hunger reference

6 = Test-defined response ("it's bigger," "to eat now/later")

7 = Nonexclusive response ("it tastes good" "it's brown")

8 = Seeming irrelevance

9 = "don't know;" "because;" no answer

Table 4

Percentage of Children Using Different Choice Rationales by Age and Sex for Year 2

		Rationale Categories*							
Group	N	2	3	4	5	6	7	8	9
51-53 mo.	74	33.8	4.1	5.4	0.0	35.1	5.4	0.0	16.2
54-56 mo.	209	38.8	1.4	6.7	0.5	31.1	3.8	4.3	13.4
57-59 mo.	206	38.6	0.0	9.2	0.5	34.0	6.3	8.3	13.1
60-62 mo.	245	35.9	0.4	4.1	0.4	40.0	4.9	4.5	9.8
63-65 mo.	143	28.0	0.7	8.4	0.0	41.3	6.3	5.6	9.8
66-69 mo.	9	22.2	0.0	0.0	0.0	33.3	0.0	11.1	33.3
Boys	466	3[illegible] 0	0.9	6.4	0.6	38.2	4.7	4.9	11.2
Girls	420	33.6	1.0	6.9	0.0	34.0	5.7	5.5	13.3
Total	886	33.3	0.9	6.7	0.3	36.2	5.2	5.2	12.2

*See Table 3 for rationale definitions.

In Year 2, test-defined responses ("it's bigger," 'to eat now/later") was the most frequently used category. Egocentric responses (category 2) decreased slightly from Year 1. Approximately 82% of the total sample did give relevant reasons for their choices, a 10% increase over the Year 1 total sample.

The frequency of different rationales did, however, vary with the nature of the child's choice. These data are presented in Table 5 for Year 1 and in Table 6 for Year 2. In Year 1, egocentric and test-defined responses accounted for the majority of rationales for both choices. For immediate choice in Year 2, there was an increase over Year 1 in test-defined responses and a decrease in egocentric responses; for delayed choice, test-defined and egocentric responses increased as did responses in which the child made a reference to home (e.g., "to share with others"). Thus, in both years the size of the reward (larger) and its accessibility (to eat now/later) appeared to influence the child's choice; however, the differential effect of each of these on choice cannot be determined in the context of the present scoring system. The meaning of egocentric responses also is vague, given the scoring categories. The decreases in category 9 ("don't know," "because," no answer) suggest an increase in the child's comprehension of the task and his ability to account for choice during this age period.

Table 5

Percent Use of Different Rationales Classified by Choice for Year 1

Choice	N	Rationale Categories 2	3	4	5	6	7	8	9
Immediate	529	36.1	1.7	1.5	0.8	20.6	8.5	4.3	26.5
Delayed	919	35.7	1.3	5.3	0.3	23.7	7.6	3.5	22.5

Table 6

Percent Use of Different Rationales Classified by Choice for Year 2

Choice	N	Rationale Categories 2	3	4	5	6	7	8	9
Immediate	357	27.7	0.6	1.7	0.6	42.6	6.4	7.0	13.4
Delayed	522	37.0	1.0	10.2	0.2	32.2	4.4	3.8	11.3

Ss who made a delayed choice were asked to recall E's instructions when the reward was presented. Data for memory of instructions are presented in Tables 7 and 8. Less than one-half of the total Year 1 sample was able to recall E's instructions, while slightly more than one-half of the Year 2 sample correctly recalled the instructions. Differences according to sex, and mother's education were not significant in repeated-measures ANOVAS, and age differences were significant only in Year 1 (F = 12.26, df = 1/580, p < .001), favoring older children.

Table 7

Percentage of Correct Recall Responses by Age and Sex for Year 1 (delayed choice only)

Age	N	Correct	Incorrect	Other*
42-44 mo.	53	49.1%	18.9%	32.0%
45-47 mo.	169	35.5	27.8	37.7
48-50 mo.	196	39.8	30.6	39.5
51-53 mo.	218	55.0	20.2	24.9
54-56 mo.	167	48.5	21.0	30.6
57-59 mo.	35	51.4	11.4	37.1
Boys	462	42.4	28.1	29.4
Girls	376	49.7	18.6	31.7
Total	838	45.7	23.9	30.4

*Includes no answer, "don't know," and indeterminate responses.

Table 8

Percentage of Correct Recall Responses by Age and Sex for Year 2 (delayed choice only)

Age	N	Correct	Incorrect	Other*
51-53 mo.	50	62.0%	20.0%	18.0%
54-56 mo.	132	54.5	16.7	28.8
57-59 mo.	116	52.6	16.4	31.0
60-62 mo.	132	52.3	31.8	15.9
63-65 mo.	73	47.9	28.8	23.3
66-69 mo.	5	40.0	40.0	20.0
Boys	269	51.7	26.0	22.3
Girls	239	54.8	19.2	26.0
Total	508	53.1	22.8	24.1

*Includes no answer, "don't know," and indeterminate responses.

Relationship with Other Measures

In the Longitudinal Study, other purported measures of impulsivity included average time on slow-instructed trials on the Motor Inhibition Test and average time to first response on the Sigel Object Categorization Test, Matching Familiar Figures Test (MFF), and the Preschool Embedded Figures Test (PEFT). Intercorrelations among these variables are presented in Table 9*. These data indicate no statistically significant relationships between choice and other purported measures of impulsivity. Similar results were reported by MacKenzie (1970) For this sample, at this age, quickness of response, ability to control response and delay of gratification do not have the same relationships as found in previous studies using older Ss which have reported low, but significant, relationships between these variables (Hess et al., 1969; Kahana & Kahana, 1970).

Table 9

Relationships Among Impulsivity Measures for Years 1 and 2

	Mischel	MFF	Sigel	PEFT	Motor Inhibition
Mischel - (Delay choice)	--	.01	-.02	.00	.07
MFF Latency (mean Log [x+1])	.00	--	.08	.07	.15
Sigel Latency (Log 10)	-.03	.46	---	.20	.07
PEFT Latency (Log 10)	.06	.23	.23	--	.02
Motor Inhibition: Average Time, Trial 2, for the Walking and Drawing Subtests	-.09	.10	.09	.06	--

Note.--Values to the right of the diagonal represent Year 2 data (N = 709-783). Values to the left of the diagonal represent Year 1 data (N = 505-738). r .001 = .145 for N = 500.

*Correlations are based on data for the longitudinal sample.

The structural analyses of Year 1 and 2 child test data (Shipman, 1971, 1972) indicated a "response tempo" factor was formed by the latency scores described above, but immediate-delay choice did not load onto this factor. Choice and memory also did not load onto the "information-processing skills" factor. Instead, analysis of the longitudinal sample data indicated that Mischel choice defined a separate factor, while the memory score had low correlations with several factors. These findings "could be interpreted as reflecting special abilities limited to one task and/or incomplete sampling of the processes reflected by tasks [Shipman, 1972, p. 56]." The failure of Mischel choice to load onto a factor with other tests, to correlate with other tests of impulsivity, or to show any stability over this age period, suggest that in young children impulsivity is a multi-dimensional trait which may have various components that are as yet not understood. For the young child, it appears that the relationship between preference for immediate or delayed reward and impulse control has not yet stabilized (cf. Shipe & Lazare, 1969).

Summary

Data from the present study indicate that use of the Mischel Technique with preschool-aged children is a feasible method for establishing baseline information for studying possible later changes in delay of gratification. Developmental changes in reward preference studied over a period of time might reveal integration of reward choice into more general impulse control; however, at the young age levels studied here the test seems to be measuring something other than impulse control. Young children appear to comprehend the task readily as evidenced by high percentages of the samples making a choice, with more than two-thirds of them able to give meaningful rationales for their choices. The task is simple to administer, and testers' comments indicated the children enjoyed it.

Although the Mischel scores showed no relationship with other purported measures of impulsivity, it must be remembered that the behaviors required for performance on these measures are quite different. The research evidence presented earlier, indicating that delay of gratification is related to other measures of impulsivity, was conducted with Ss older than those in this study. At the early age levels with which this study dealt, the data indicate that delay of gratification does not have the same implication that it has for older Ss. The data do indicate, however, that young children can delay impulse to receive a larger reward at a later time point when the delay interval is relatively short and clearly specified. Stability of choice across years was low, however, as were correlations with other measures; the fact that the test involves only one item may account for this. Group comparisons of data rather than individual differences would seem more appropriate for understanding this test.

In using the Mischel Technique with young children several aspects of the procedure need to be considered. One point which needs to be stressed is the child's trust in the tester and his environment (cf. Mischel, 1958; Strickland, 1972). If the child does not trust adults, especially strangers, or "promises," he may make an immediate choice which should be interpreted in light of these considerations and impulse control. It was noted during the testing session that many children who chose the immediate reward saved the candy to take home later. This would suggest caution in interpreting an immediate choice as inability to delay oral gratification. Consideration needs to be given to relationships between choice and characteristics of the delay agent, including the possiblity of differential race effects between examiner and child (cf. Strickland, 1972). Inclusion of an initial size identification question and a concrete specification to the child of the length of the delay interval seem to be important procedural modifications for use with young children. Also to be considered is the differential appeal of the reward to the child; candy may not have been the best choice, as hunger could have affected responses.

References

Bialer, I. Conceptualization of success and failure in mentally retarded and normal children. Journal of Personality, 1961, 29, 303-320.

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.

Kahana, B. & Kahana, E. Roles of delay of gratification and motor control in the attainment of conceptual thought. Proceedings of the 78th Annual Convention of the American Psychological Association, 1970, 5, 287-288.

Klein, R. H. The effects of length of delay interval and expectancy control on preference for delayed rewards in children and normal and schizophrenic adults. (Doctoral dissertation, State University of New York at Buffalo) Ann Arbor, Michigan: University Microfilms, 1967, No. 66-7974. (Abstract)

Kunz, J. & Moyer, J. E. A comparison of economically disadvantaged and economically advantaged kindergarten children. Journal of Educational Research, 1969, 62, 392-395.

MacKenzie, C. S. Motor inhibition and capacity for delay of gratification in two classes of Negro boys. (Doctoral dissertation, University of Maryland) Ann Arbor, Michigan: University Microfilms, 1970, No. 70-13, 721. (Abstract)

Mendell, D. Relating delay of gratification to achievement in young children. (Doctoral dissertation, Yeshiva University) Ann Arbor, Michigan: University Microfilms, 1967, No. 67-9668. (Abstract)

Mischel, W. Preference for delayed reinforcement: An experimental study of a cultural observation. Journal of Abnormal and Social Psychology, 1958, 66, 57-61.

Mischel, W. Preference for delayed reinforcement and social responsibility. Journal of Abnormal and Social Psychology, 1961, 62, 1-7.

Mischel, W. & Metzner, R. Preference for delayed reward as a function of age, intelligence, and length of delay interval. Journal of Abnormal and Social Psychology, 1962, 64, 425-431.

Phypers, J. M. Delay of gratification and adjustment in elementary school. (Doctoral dissertation, University of Colorado) Ann Arbor, Michigan: University Microfilms, 1970, No. 70-5883. (Abstract)

Shipe, D. & Lazare, S. Two measures of delay of gratification: Age and socioeconomic status in young children. Paper presented at the meeting of the American Psychological Association, Washington, D. C., September 1969.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Staub, E. Effects of persuasion and modeling on delay of gratification. Developmental Psychology, 1972, 6, 166-177.

Strickland, B. R. Delay of gratification as a function of race of the experimenter. Journal of Personality and Social Psychology, 1972, 22, 108-112.

Trunfio, J. A., Jr. & Wohlford, P. Mother's personal time and their children's ability to delay gratification. In Head Start parents in participant groups: The parent project. Final report submitted to the Office of Economic Opportunity, 1972.

Walls, R. T. & Smith, T. S. Development of preference for delayed reinforcement in disadvantaged children. Journal of Educational Psychology, 1970, 61, 118-123.

Weintraub, S. A. Cognitive and behavioral impulsivity in internalizing, externalizing, and normal children. (Doctoral dissertation, University of Minnesota) Ann Arbor, Michigan: University Microfilms, 1969, No. 69-11, 471. (Abstract)

Wohlford, P., Clark, L. C. & Trunfio, J. A., Jr. Effects of interviewers' race and sex on black mothers' child - rearing attitudes and their Head Start children's delay of gratification preferences. In Head Start parents in participant groups: The parent project. Final report submitted to the Office of Economic Opportunity, 1972.

Zytkoskee, A., Strickland, B. R. & Watson, J. Delay of gratification and internal versus external control among adolescents of low socioeconomic status. Developmental Psychology, 1971, 4, 93-98.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

YOUNG CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

A LONGITUDINAL STUDY

Technical Report 13

Motor Inhibition Test

William C. Ward

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Motor Inhibition Test

Background

The Motor Inhibition Test was one of several measures of "impulse control" administered in this study. As a group, these measures permit investigation of the dimensionality of self-regulatory behavior and of its implications for intellectual performance in young economically disadvantaged children. The child's score on this task is the time taken to perform several motor acts, under instructions to proceed as slowly as he can. Thus, in contrast to measures like the Matching Familiar Figures Test, this task investigates the child's ability rather than his stylistic preference in self-regulation.

Maccoby, Dowley, Hagen, and Degerman (1965), who introduced the task, found that Motor Inhibition scores in middle-class nursery school children were highly consistent across acts, and that they were positively related (r = .44) to IQ. Support for the generality of these findings has been found with lower-class preschool children (Massari, Hayweiser, & Meyer, 1969), in another sample of middle-class nursery school children (Loo & Wenar, 1971), and in eight-year-old middle-class boys (Ward, 1968a). Massari et al. (1969) also provided evidence that the association with IQ was not based on intelligence differences in the child's understanding of the instruction to perform the act slowly; for example, they showed that time scores when the child was instructed to act fast were not associated with IQ. Mumbauer and Miller (1970) in their study of advantaged and disadvantaged preschoolers found no significant correlation between Motor Inhibition and IQ. For advantaged children, a trend in the direction reported by Maccoby et al. (1965) was found, but the correlation was not statistically significant ($p > .05$). Similarly, Huron Institute (Walker, Bane & Bryk, 1973) reported low correlations with Binet IQ for the Fall 1970 Head Start Planned Variation (HSPV) sample.

Several investigators have reported significant positive correlations between Motor Inhibition scores and response times on tests of reflection-impulsivity (Banta, 1970; Hess, Shipman, Brophy, & Bear, 1969; Kagan, Rosman, Day, Albert, & Phillips, 1964; Ward, 1968b). The relationship is weak, but provides some evidence for the existence of generality in the impulse control or self-regulation domain.

Maccoby et al. (1965) reported the short-term retest reliabilities of times on the several motor acts to be high, ranging from .77 to .89. (However, the retest interval was not reported, and only ten subjects were employed.) Hess et al. (1969) found the score from one act (draw-a-circle slowly) to be correlated .28 (N ca. 153; $p < .01$) over a one year period in children first tested at first grade, while a fall-to-spring study of Head Start children found a correlation of only .06 (N = 1457; $p < .05$) for the total score (Stanford Research Institute, 1971). Walker et al. (1973) reported test-retest reliability coefficients of .30 to .71 for the "slow" times (log transformed) after two weeks for approximately 20 Head Start children tested by testers with differential testing experience. Bissell (1972) reported significant increases (fall to spring testing) in motor inhibition scores for Ss in Head Start Planned Variation programs. The largest gains were made by Ss in programs low in structure (e.g., programs emphasizing exploration and self-expression).

Task Description and Administration

The test requires the child to perform three motor acts. The first involves walking a distance of six feet on a five-inch wide runway. In the second, the child must draw a line, using ruler and pencil, between two points 8 1/2 inches apart. The line completes a missing segment of wire between two telephone poles shown on the paper. In the third task, the child is given a toy tow truck and a jeep which is "broken;" he is to wind the jeep up to the rear of the tow truck, a distance of 30 inches, so that it can be taken to the repair shop.

Each act is introduced with a practice trial, on which no reference to rate of performance is made; immediately afterwards, the trial is given on which the child is instructed to perform the act as slowly as possible.

The "truck" subtest did not prove satisfactory in Year 1 and therefore was eliminated from subsequent years of the study. Its scores were not used in analyses of the interrelations of measures in Year 1.

A second change from Year 1 to Year 2 administration concerned the walkway used for the walking subtest. In Year 1, fiberboard walkways, hinged at the middle for easier handling, were used; these proved unstable, leading to occasional balancing problems for children. In the second year, therefore, the runway was laid out on the floor in masking tape, as had been done in most previous studies with this task.

While the tasks the child is asked to perform are simple, the tester's job is made somewhat difficult by the need simultaneously to time the child's behavior and to monitor, and assist with, his performance. On the walking subtest, the chief problem is to assure that the child starts at the beginning of the walkway and does not fall off. On the drawing subtest, the tester must give the child practice in drawing, and during both practice and test trials, must, if necessary, assist by holding the ruler steady. On the truck subtest, finally, the tester must hold the truck steady and be prepared to redirect the jeep if it begins to roll to the side rather than directly towards the tow truck.

Scoring

The data consist of six scores--for each of the three subtests, the time taken on the practice trial and on the "slow" instruction trial. Because of their skewed distribution, these scores were transformed by log (x + 1) for all analyses.

Score Properties

Means and percentiles for the subtests administered in Years 1 and 2 are presented in Table 1.

Table 1

Means, Standard Deviations, and Percentile Distributions*
for Trials 1 and 2 in Years 1 and 2

Score	N	Mean	SD	Percentiles 10	25	50	75	90
Year 1								
Walking - Practice	1467	0.76	0.16	0.57	0.65	0.74	0.84	0.98
Slow	1471	0.87	0.21	0.61	0.72	0.85	1.01	1.15
Drawing - Practice	1392	0.68	0.22	0.40	0.53	0.66	0.81	0.96
Slow	1406	0.84	0.29	0.48	0.62	0.81	1.01	1.23
Tow Truck - Practice	1448	1.62	0.17	1.42	1.52	1.62	1.72	1.83
Slow	1448	1.71	0.18	1.49	1.59	1.70	1.82	1.94
Year 2								
Walking - Practice	1117	0.75	0.15	0.57	0.64	0.74	0.84	0.96
Slow	1114	0.94	0.22	0.66	0.78	0.93	1.08	1.24
Drawing - Practice	1289	0.65	0.20	0.40	0.51	0.63	0.76	0.90
Slow	1290	0.95	0.30	0.57	0.74	0.93	1.14	1.35

*Data are times in seconds transformed by log (x + 1).

Correlations among the slow administration time scores were examined to determine whether all three subtests did in fact contribute to a single dimension of ability to inhibit response. For two of the subtests--walk-a-line and draw-a-line--the intercorrelation was .50 (Year 1) and .53 (Year 2), indicating the presence of such a dimension. This is almost identical to that reported by Walker et al. (1973) for the Fall 1970 HSPV Sample (r = .46, N = 1043). The third subtest, in which the child had to wind up a toy jeep to the back of a tow truck, showed lesser correlations with the first two, approximately .25 (Year 1 only). (Walker et al. [1973] also report correlations in the 20's.) The lower relation may have been due to a combination of the greater demands this subtest made on the child's coordination--the winch of the tow truck was difficult to manipulate smoothly--and on the tester's skill--the truck had to be held steady, and children had to be kept from reversing the direction in which they were winding, at the same time the tester was attempting to time the task.

Practice and slow times from each subtest were related around .50 in Year 1, and .40 in Year 2, reflecting shared method variance; but there was little consistency among practice times: those from the walking and drawing subtests correlated .17 in each year's data, while time from the truck subtest had near zero correlations with time from each of the others.

The stability of performance over two administrations was examined for a single score, the average of standardized "slow" administration time scores for the walking and drawing subtests. This score showed a correlation of .39 over years.

Sample Performance

Descriptive data for age and sex results are presented in Tables 2 and 3, for average transformed "slow" times over the walking and drawing subtests. For purposes of analysis, these two subtests were combined into an average time score for each subject, and these data were then examined by analysis of variance (age x sex x SES), using only those subjects tested in both Years 1 and 2 of the study.

Table 2

Means, Standard Deviations, and Percentile Distributions
by Age and Sex in Year 1

Walking Subtest - Slow Administration

				Percentiles				
Group	N	Mean	SD	10	25	50	75	90
42-44 mo.	96	0.84	0.18	0.61	0.70	0.82	0.97	1.10
45-47 mo.	312	0.87	0.21	0.61	0.72	0.84	1.00	1.13
48-50 mo.	345	0.85	0.21	0.60	0.70	0.83	0.98	1.12
51-53 mo.	385	0.88	0.22	0.60	0.72	0.87	1.01	1.15
54-56 mo.	274	0.92	1.22	1.65	1.76	1.88	1.04	1.21
57-59 mo.	59	0.87	0.23	0.56	0.76	0.84	1.00	1.13
Boys	773	0.86	0.21	0.61	0.72	0.84	0.99	1.14
Girls	698	0.88	0.21	0.61	0.72	0.86	1.02	1.17
Total	1471	0.87	0.21	0.61	0.72	0.85	1.01	1.15
		Drawing Subtest - Slow Administration						
42-44 mo.	84	0.76	0.25	0.47	0.58	0.73	0.94	1.04
45-47 mo.	294	0.81	0.29	0.47	0.59	0.77	0.99	1.18
48-50 mo.	328	0.80	0.27	0.47	0.61	0.77	0.97	1.16
51-53 mo.	373	0.85	0.28	0.50	0.64	0.83	1.04	1.25
54-56 mo.	270	0.89	0.31	0.51	0.67	0.88	1.09	1.32
57-59 mo.	57	0.90	0.28	0.56	0.69	0.84	1.11	1.24
Boys	742	0.83	0.29	0.48	0.62	0.80	1.00	1.21
Girls	664	0.84	0.29	0.49	0.63	0.82	1.02	1.25
Total	1406	0.84	0.29	0.48	0.62	0.81	1.01	1.23

Table 3

Means, Standard Deviations, and Percentile Distributions by Age and Sex in Year 2

Walking Subtest - Slow Administration

Group	N	Mean	SD	Percentiles 10	25	50	75	90
51-53 mo.	71	0.91	0.21	0.67	0.75	0.87	1.05	1.23
54-56 mo.	270	0.90	0.21	0.64	0.76	0.88	1.02	1.19
57-59 mo.	249	0.95	0.22	0.66	0.78	0.93	1.10	1.25
60-62 mo.	306	0.95	0.22	0.67	0.79	0.96	1.09	1.23
63-65 mo.	206	0.97	0.22	0.69	0.81	0.96	1.13	1.29
66-69 mo.	12	0.97	0.32	0.71	0.80	0.90	1.05	1.33
Boys	587	0.93	0.22	0.66	0.78	0.92	1.07	1.24
Girls	527	0.95	0.22	0.67	0.78	0.93	1.09	1.25
Total	1114	0.94	0.22	0.66	0.78	0.93	1.08	1.24
Drawing Subtest - Slow Administration								
51-53 mo.	73	0.95	0.27	0.64	0.74	0.92	1.16	1.32
54-56 mo.	315	0.89	0.31	0.51	0.66	0.88	1.11	1.31
57-59 mo.	290	0.94	0.30	0.58	0.72	0.91	1.13	1.36
60-62 mo.	344	0.97	0.30	0.60	0.77	0.95	1.16	1.37
63-65 mo.	252	0.99	0.30	0.62	0.79	0.98	1.16	1.38
66-69 mo.	16	0.89	0.33	0.56	0.62	0.87	1.10	1.29
Boys	683	0.93	0.29	0.56	0.72	0.92	1.11	1.31
Girls	607	0.97	0.31	0.59	0.75	0.94	1.19	1.40
Total	1290	0.95	0.30	0.57	0.74	0.93	1.14	1.35

Age analyses, performed separately for both years on the data split at the median into an "older" and a "younger" group, were highly significant in both Year 1 ($\underline{F}$ = 27.14, $\underline{df}$ = 1/1118, $\underline{p} < .001$) and Year 2 ($\underline{F}$ = 15.47, $\underline{df}$ = 1/1119, $\underline{p} < .001$), favoring the older subjects, although the differences were very small in absolute magnitude in both years. Over all trials in both years, age correlated with log time from -.08 to .14. Results of a repeated-measures analysis of variance (age x sex x SES) on data from children tested in both years showed SES, as indexed by mother's education (below 10 years of schooling, 10-12 years, above 12 years), to have a highly significant effect across years (F = 52.74, df = 2/1089; $p < .001$), with scores increasing as mother's education increased.

Children in the present sample performed the motor acts relatively quickly. The mean number of seconds to complete the walking and drawing subtests under slow instructions was 6.4 and 5.9, respectively, in Year 1, and 7.7 and 7.8, in Year 2. It is clear, therefore, that there is ample opportunity for further development in these children of the ability to slow down a motor response.

Finally, instructions to perform the act slowly did lead children to perform more slowly on the second trial than the practice trial for each task. Mean time scores under slow instructions represented an increase over practice times of 36% for the walking subtest and 54% for the drawing task in Year 1, and of 65% and [illegible]%, respectively, in Year 2 data. Moreover, when the sample was divided into six three-month age groups, an increase in mean times from first to second trial was found on each subtest for every age level in both years. Thus, although the change in performance under the slow instruction was not large in absolute terms, it was highly consistent, and even the youngest children were able to conform to the task demand.

Construct Validity

Results with the present sample were, in general, consistent with those that have been found by other investigators using this task. There was evidence for consistency in "slow" administration times across the tasks though, as discussed above, and presumably because of the poor design of the subtest, scores from the tow truck subtest were only moderately correlated to those from the other two subtests. Motor inhibition ability (average standardized time, after log transformation, on the slow administration of the walking and drawing subtests) correlated* positively with various measures of competence of performance, for example, .36 and .37 with the Preschool Inventory in Years 1 and 2, respectively; and .34 and .36 with the Peabody Picture Vocabulary Test, Form A. In the various breakdowns of the data it loaded consistently on the first, competence-related factor (Shipman, 1971, 1972). Practice times, on the other hand, showed little internal consistency, little relation to slow times, and no relation to ability measures, supporting the interpretation of the task as measuring the child's self-regulation ability rather than a preferred response style.

There was, however, no relation to other measures in the "impulsivity" domain. Of the three latency measures which might, at least in Year 1, contain impulsivity variance (those on the MFF, Sigel, and PEFT), none correlated higher than .15 with MIT performance in either year. Likewise, ability to delay gratification (Mischel choice) had correlations of .01 in Year 1, and .07 in Year 2, with MIT performance. Motor Inhibition ability did correlate -.22 with Matching Familiar Figures error scores in Year 1, and -.25 with errors in Year 2. Presumably, however, this correlation was mediated by an ability rather than a stylistic component of MFF performance; partialling out Preschool Inventory scores, for example, reduced these correlations to -.08 and -.07, respectively.

Summary

These results suggest that the most appropriate motor inhibition score

---

*All correlations are based on data from the longitudinal subjects only.

from this test is the average of standardized (and log transformed) slow times from the walking and drawing subtests. The truck subtest results will be omitted in further analyses, and the test as given in future years of this study will not include this subtest. (Huron Institute [Walker et al., 1973], from its reanalysis of 1970 HSPV data, comes to a similar recommendation.) It remains to be seen whether the truck subtest is simply unreliable because of equipment and administration problems, or whether additional processes are brought into play during the substantially longer time period required for response on this subtest than on the other two employed.

The lack of intertask consistency in practice time indicates that there is no need to "correct" the motor inhibition score for practice time. Moreover, the absence of a factor common to practice and slow administration times suggests that the test provides a measure of the ability to regulate one's own behavior, rather than a stylistic preference for a fast or slow tempo of response, or an index of motor coordination.

Motor Inhibition ability correlated significantly with other measures of competence and achievement in the battery and, like these, was positively associated with age and SES The results of Massari et al. (1969) suggest that the relationship is not simply due to a failure of less able or less advantaged children to understand the instruction to perform slowly; and in the present data the finding that even the youngest age groups showed some tendency to slow response on instruction is compatible with the same conclusion. Perhaps inhibition of response improves performance on ability tests, or perhaps more able children are more advanced in a developmental sequence, one of whose attributes is increasing self-regulation ability; the present data do not provide the information needed to explain what processes underlie the relation.

A general "impulsivity" dimension did not emerge in the first two years' data of this study. Results from future years will be examined to discover whether such a dimension develops at a later age in these children and the extent to which it is associated with different teaching behaviors; for example, the extent to which the child is encouraged to reflect and consider alternative responses.

References

Banta, T. J. Tests for the evaluation of early childhood education: The Cincinnati Autonomy Test Battery (CATB). In J. Hellmuth (Ed.), Cognitive studies. Vol. 1. New York: Brunner/Mazel, 1970.

Bissell, J. S. Planned variation in Head Start and Follow Through. Paper presented at the Second Annual Blumberg Symposium, Baltimore, Maryland, 1972.

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, 1969.

Kagan, J., Rosman, B. L., Day, D., Albert, J. & Phillips, W. Information processing in the child: Significance of analytic and reflective attitudes. Psychological Monographs, 1964, 78 (1, Whole No. 578).

Loo, C. & Wenar, C. Activity level and motor inhibition: Their relation to intelligence-test performance in normal children. Child Development, 1971, 42, 967-971.

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 intellectual functioning in deprived preschool children. Developmental Psychology, 1969, 1, 286-290.

Mumbauer, C. C. & Miller, J. O. Socioeconomic background and cognitive functioning in preschool children. Child Development, 1970, 41, 471-480.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

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). Interim report: first year of evaluation. Menlo Park, California, May 1971.

Walker, D. R., Bane, M. J., & Bryk, T. The quality of the Head Start Planned Variation data. Cambridge, Mass.: Huron Institute, 1973, in press.

Ward, W. C. Creativity in young children. Child Development, 1968, 39, 737-754. (a)

Ward, W. C. Reflection-impulsivity in kindergarten children. Child Development, 1968, 39, 867-874. (b)

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

A LONGITUDINAL STUDY

Technical Report 14
Open Field Test

William C. Ward

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Open Field Test

Background

Most tests require the child to perform a narrowly defined task, and provide for step-by-step control over his activity by the tester. It is likely that there are important dimensions of behavior which might be assessed by observing the child in a less structured, more "open" play setting. Such dimensions would include both cognitive variables (e.g., complexity and duration of play activities) and personal-social ones (e.g., style in coping with an unfamiliar situation). The Open Field Test provided such a setting.

Virtually nothing is known concerning the importance of situational parameters for the manifestation of individual differences in play. Therefore, it seemed advisable to provide as "neutral" a setting as possible. The child was placed in a room containing a number of toys, varying in familiarity and complexity; he was given minimal instructions; and minimal social stimulation was provided--the tester was present in the room, but unobtrusive and minimally responsive. Several investigators have employed similar settings, Goldberg and Lewis (1969), Messer and Lewis (1970), and Kagan (1971) with infants of both sexes and of both middle and lower class; and Hess, Shipman, Brophy and Bear (1969), with first grade children covering a broad SES range. In these studies, in contrast to the present one, the child's mother was present in the room, and recording was done by observers watching through one-way mirrors. Interestingly, both sets of studies found SES differences to be largely nonexistent in such a setting, though there were sex differences in both social and nonsocial behaviors. With young children there is also some evidence that the meaning of one frequent measure, the number of changes in play objects in a fixed time, differs across sex, relating positively to rate of habituation to visual stimuli in boys, and negatively in girls (Kagan, 1969; Messer, Kagan & McCall, 1970). With first graders, finally, there is evidence for relations between measures of level and length of sustained play activity, on the one hand, and of maternal control and teaching strategies and of child competence and cognitive style, on the other (Hess et al., 1969).

Task Description and Administration

After a child was halfway through one conventional test battery, he was brought into a new testing room. He was shown ten standard play objects arranged around the room including two dolls (one dark-skinned, one light), a truck, alphabet blocks, Rising Towers (complex plastic building blocks), clay, crayons, felt-tipped markers, plain paper, and a four-page coloring book. He was told that he could do anything he wanted with the toys. The tester seated herself in one corner of the room and remained there for ten minutes, initiating no interaction with the child and responding minimally to any overtures he made.

During each thirty-second period of the test, the tester recorded and described every play activity involving each object, along with a variety of nonplay activities--child verbalizations, divided into those directed toward the tester and those apparently made for the child's own benefit; approaches to the tester; active attempts to leave the testing situation; and inactivity. Each behavior was tallied in a matrix in which columns indicated the 20 time periods into which the task was divided, and in which each row indicated a particular object or nonplay activity. (See answer sheet attached.) Entries in the matrix were also numbered so as to be coordinated with descriptions the activity entered on lines below the matrix. Every play activity occurring during an interval was entered, no matter how brief. The tester also indicated, by drawing a circle to include the two or more tallies, whenever the child combined the use of several objects in his play.

Scoring

The initial step in data analysis was to recode the data from the answer sheet onto a 15 x 20 matrix. As before, columns indicated time periods and rows indicated specific objects or nonplay activities. Here, however, for the 10 objects the number entered in each cell of the matrix referred to a given level of play activity. Nonplay activities were coded for presence-absence in each cell, except for the two verbal activities. "Talks to tester" was coded into attempts to draw the tester's attention to the task, seeking help or direction, attempts to withdraw from the play situation, and other or ambiguous verbalizations. "Talks to self" was categorized as related to the play activity, unrelated to the play activity, and ambiguous. Details are given in the coding manual attached.

Coders were trained through cell-by-cell discussion of their practice coding, and reached a level of agreement of around 90% with a "standard" coding of twenty protocols, chosen to be of greater than average difficulty, by the end of the training period. (An entry here was counted as an agreement only if the object, the time period, and the level or subcategory entered were all exactly correct; thus, many "disagreements," such as those credited to an incorrect time period, would generally have no effect on total scores.)

Three aspects of the child's play activities were scored. The first of these was the number of half-minute periods out of the twenty during which he engaged in any play activities. The second was mean complexity of play. All activities with the objects were coded into one of four "levels."

Level 1 play involved only attending to a play oject; level 2, holding or manipulating it; level 3, playing with one object alone; and level 4, using two or more objects in an integrative activity. The complexity score is the mean level taken over all play activities recorded. This procedure made possible an objective and relatively straightforward approach to complexity of play, yielding scores whose ranking of subjects closely agreed with intuitive judgments of complexity. Third, the duration of sequences of activity engaged in by the child was measured. A "simple" sequence was defined as a series of half-minute periods during which the child continued without interruption to play with the same object.

Number of simple sequences, mean length, and length of longest and of shortest simple sequence were all obtained. Comparable scores were obtained for "complex" sequences, those in which at least part of the time was spent in play involving an integration of two or more of the objects. The three types of scores indicated above all represent summaries of the child's play activity, without respect to the particular play objects employed. In addition, scores were obtained for the number of periods during which the child employed each object in play, and for the mean level of complexity of his play with each object used.

The remaining scores obtained concerned the child's nonplay activities. Verbalizations directed toward the tester and verbalizations intended for the child himself were scored (number of periods in which each of these occurred), along with counts of the number of periods in which each subcategory of verbalization of those listed above was found. Finally, the number of time periods in which the child approached (or remained with) the tester, made active attempts to leave the testing situation, and engaged in no overt activity, were each [illegible]red.

Score Properties

Examination of the data showed that a number of the scores listed above represented rare events. While some of these might be of interest in future, specialized analyses, they did not possess suitable distributions for inclusion in the factor analyses and analyses of variance conducted thus far. Table 1 contains the means and medians for these scores, expressed in terms of the

Table 1

Scores Possessing Insufficient Variance to Support Further Analysis

	Year 1 (N = 1471)				Year 2 (N = 893)			
Score	Mean	SD	Median	Percent of Subjects	Mean	SD	Median	Percent of Subjects
No. of Play Periods	18.61	3.90	19.83	75[a]	19.44	2.08	19.89	82[a]
Verbalizations to Tester:								
-Seek attention	0.64	2.00	0.12	19	0.64	1.76	0.16	24
-Seek help	0.32	0.93	0.10	17	0.29	0.78	0.10	17
-Stop activity	0.12	0.63	0.03	6	0.09	0.58	0.03	5
Verbalizations to Self:								
- Related to Play	0.71	2.32	0.09	16	1.05	2.57	0.18	27
- Unrelated to Play	0.24	1.35	0.03	6	0.33	1.36	0.06	11
Approach Tester	0.14	0.71	0.04	7	0.18	0.69	0.06	10
Attempt to Leave	0.17	0.99	0.03	6	0.10	0.59	0.03	5
Do Nothing	0.80	3.20	0.06	11	0.37	1.36	0.05	9

Note.--Means and medians are for number of periods in which the subject engaged in the activity in question. Percent of subjects is percent who engaged in the activity at all.

[a]Percent of subjects playing with objects in 20 of 20 play periods.

number of time periods out of 20 in which the child engaged in the activity, and the per cent of children who engaged in the activity at any time during the 10-minute observation. Number of periods of play is included in this table since there was virtually no failure of the child to engage in play activities for all, or nearly all, of the time spent in the situation. Totals for verbalization to tester and to self possessed an adequate range for use in analysis, but the subcategories into which each of these was divided did not; the median occurrence of each such category was less than .2 time periods. Approaching the tester, attempting to leave the situation, and engaging in no activity, finally, all had median frequencies of less than .1 time periods in both years of testing.

Examination of scores for sequences of activities showed that these were highly interrelated. Number of simple sequences, longest simple sequence, and shortest simple sequence, for example, had intercorrelations on the order of .70 (ignoring sign; number of sequences was negatively correlated with each of the other two measures) in each year of testing. One score, number of simple sequences, was chosen to represent simple sequences in further analyses. Complex sequences form a subset of simple ones with only about one-third of the sample showing any such sequences in either year. No score for complex sequences was included in the analyses.

Four scores appropriate for use in structural analyses remained after this process of score reduction; these were mean complexity of play, number of periods child talked to tester, number of periods child talked to self, and number of simple sequences. An estimate of the reliabilities of these scores was obtained by scoring the first versus the second half of the test for a randomly drawn sample of 100 cases in each year. Split-half reliabilities for mean complexity of play were .61 in Year 1 and .52 in Year 2; they were .81 and

.72, respectively, for verbalization directed at the tester, and .73 and .75 for self-directed verbalization. This procedure was not appropriate for assessing the reliability of the number of sequences score. These are conservative estimates of the reliability of the child's behavior, since any systematic changes in behavior over time within the session might result in lowered coefficients.

In addition, scores for play with each object were retained for examination. Finally, scores for "masculine" and "feminine" play were derived empirically, based on play with those objects on which clear sex differences in amount of play were obtained. Given the lack of any very satisfactory procedure for obtaining reliability estimates on these time-based scores, especially for events which were relatively low in rate of occurrence for the typical subject, no reliability assessments were made for these scores.

Sample Performance

Attention in the analysis was directed toward the four "major" scores listed above. Means for each of these scores are presented in Tables 2 and 3 for Years 1 and 2, respectively, by three-month age intervals, sex, and total sample. For each of these scores, age (median split on age at time of Year 1 testing) x sex x SES (three levels of mother's education) repeated-measures analysis of variance was performed. For mean complexity of play a significant effect of sex was present; males tended to engage in more complex play ($\underline{F}$ = 46.87, $\underline{df}$ = 1/678, $\underline{p} < .001$). The sex effect was also significant at the .001 level in age x sex x SES analyses conducted separately for the longitudinal sample each year. In Year 1, males had a mean complexity score of 3.19 and females, 3.05; in Year 2, the respective means were 3.12 and 3.01. These analyses showed no significant changes over years in mean complexity, [illegible]d no effects attributable to age or SES.

Table 2

Means and Standard Deviations for Selected Scores by Age and Sex for Year 1

Score		42-44	45-47	48-50	51-53	54-56	57-59	Boys	Girls	Total
		Age in Months								
Mean	N	94	307	331	383	265	59	768	671	1439
Complexity	M	3.08	3.10	3.10	3.10	3.13	3.05	3.16	3.03	3.10
of Play[a]	SD	0.41	0.31	0.35	0.29	0.28	0.40	0.35	0.27	0.32
Number of	N	95	308	332	378	266	59	769	669	1438
Simple	M	2.97	3.00	2.86	2.90	2.87	2.54	2.77	3.05	2.90
Sequences[a]	SD	2.19	2.23	2.15	2.22	2.08	1.75	2.06	2.26	2.16
Talk to	N	96	318	337	388	271	61	784	687	1471
Self (Number	M	0.98	1.92	1.94	1.89	1.68	1.54	1.96	1.60	1.79
of Periods)	SD	2.43	3.87	3.93	3.80	3.54	3.36	3.94	3.43	3.71
Talk to	N	96	318	337	388	271	61	784	687	1471
Tester (Number of	M	2.27	2.10	1.98	2.13	2.34	2.28	2.04	2.26	2.14
Periods)	SD	3.99	3.47	3.56	3.86	4.00	3.69	3.53	3.95	3.74

[a]Subjects who showed no play activities eliminated from sample for this score.

In the analyses for the number of periods in which the child talked to the tester, the repeated-measures analysis showed a marginal effect of SES ($F = 4.32$, $df = 1/699$, $p < .02$) and a marginal year by SES interaction ($F = 3.04$, $df = 2/699$, $p < .05$). The analyses for individual years showed a significant SES effect in Year 1 ($F = 9.99$, $df = 2/1121$, $p < .001$), with means of 1.5, 2.3, and 3.2 periods, respectively, for levels 1, 2, and 3 of SES, while there was no effect in Year 2 (means of 1.1, 1.4, and 1.3 periods). Age and sex had no significant effects on talking to the tester.

Number of periods child talked to himself showed no significant effects in the repeated-measures analysis or in the analysis of Year 2 data alone. The Year 1 analysis showed a significant effect of sex, with males engaging in more such talk ($\underline{F}$ = 9.60, $\underline{df}$ = 1/1121, $\underline{p} < .005$). Means were 2.0 periods for males and 1.4 for females.

Number of simple sequences, finally, showed only an increase in mean score over years in the repeated-measures analysis ($\underline{F}$ = 15.00, $\underline{df}$ = 1/682, $\underline{p} < .001$). In the analysis of Year 1 data there was a marginally significant

Table 3

Means and Standard Deviations for Selected Scores by Age and Sex for Year 2

Score		Age in Months						Boys	Girls	Total
		51-53	54-56	57-59	60-62	63-65	66-69			
Mean	N	75	207	209	244	146	9	464	426	890
Complexity	M	3.10	3.07	3.12	3.07	3.09	3.01	3.15	3.02	3.09
of Play[a]	SD	0.28	0.21	0.29	0.25	0.24	0.37	0.28	0.21	0.26
Number of	N	75	207	209	244	146	9	464	426	890
Simple	M	3.00	3.44	3.16	3.59	3.37	3.33	3.49	3.24	3.37
Sequences[a]	SD	2.15	2.34	2.03	2.38	1.88	2.12	2.18	2.21	2.20
Talk to	N	75	209	209	245	146	9	465	428	893
Self (Number	M	1.64	1.70	2.03	1.76	1.90	1.00	1.80	1.83	1.82
of Periods)	SD	3.71	3.53	3.81	3.24	3.58	2.29	3.41	3.66	3.53
Talk to	N	75	209	209	245	146	9	465	428	893
Tester (Number of	N	1.52	1.44	1.43	1.32	1.25	0.11	1.19	1.55	1.36
Periods)	SD	3.40	3.02	2.53	2.34	2.33	0.33	2.35	2.92	2.64

[a] Subjects who showed no play activities eliminated from sample for this score.

SES effect ($F = 3.46$, $df = 2/1036$, $p < .05$), with means of 3.0, 3.0 and 2.5 obtained by the lower, middle and higher SES children, respectively. There were no significant effects in the analysis of Year 2 data alone.

For these four scores, low but significant correlations across years were obtained. Mean complexity of play correlated .19 from Year 1 to Year 2; number of periods talking to the tester, .17; number of periods talking to self, .15; and number of simple sequences, .21. With N's of 744 or larger, each of these coefficients is significant at .01.

Despite a high degree of consistency in relative frequency of play with the objects (see next section), males and females showed several significant differences in play choices. In both years, males spent more time with the truck and blocks, while females spent more time with the dolls, crayons, and coloring book ($p < .01$ for each comparison). An index of "masculine" play was derived by summing the number of periods of play with the truck and blocks, while an index of "feminine" play was obtained by summing play activities with the dolls, crayons, and coloring book. The means of these scores in each year are presented in Table 4. In age x SES

Table 4

Sex Typing in Mean Number of Periods of Play with Objects Preferred by Each Sex

	Year 1		Year 2	
	Masculine Play	Feminine Play	Masculine Play	Feminine Play
Boys	9.46	3.82	9.22	4.54
Girls	3.78	8.13	3.35	8.46

analysis of variance, feminine play by males showed an effect for SES (more feminine play associated with lower SES, $p < .025$). No other significant age or SES effects were found.

Play with Individual Objects

Results for the child's play with individual objects are presented in Tables 5 through 8. These data are presented as possibly useful descriptive information; they have not been examined in detail or used in structural analysis. Year 1 data are presented in Table 5 for the mean number of periods of play with each object, for each age group, boys, girls, and total sample; and in Table 6, by the same breakdowns, for the number of children playing with each object and the mean complexity of play taken over those subjects who played with the object in question. Year 2 data are presented analogously in Tables 7 and 8.

These data were examined to determine whether boys and girls showed consistency in their play with the 10 objects. There was agreement across sex in the relative number of periods in which each object was used (the rank order correlation, rho, was .66 for Year 1 data and .60 for Year 2), in the number of children playing with each object (rho = .70 and .76, respectively), and in the mean complexity of play with each object (rho = 1.00 and .75, respectively). For five of these coefficients, $p < .05$. Blocks, truck, and Rising Towers elicited the most complex play from children of each sex. These, along with the magic markers, coloring book, and (in Year 2) the clay, were most often played with. The dolls were less popular with children of both sexes and tended to elicit the least complex play activity.

A similar analysis was used to examine the consistency of play behaviors across age groups within each year. The index here was the coefficient of

Table 5

Mean Number of Periods of Play with Each Object*
(Year 1)

	Age in Months																	
	42-44		45-47		48-50		51-53		54-56		57-59		Boys		Girls		Total	
Object	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Black Doll	0.94	1.89	0.50	1.38	0.74	2.11	0.67	2.14	0.72	2.36	0.82	2.68	0.28	0.95	1.14	2.74	0.68	2.05
White Doll	0.80	1.57	0.48	1 31	0.70	2.15	0.69	2.30	0.51	1.97	0.48	1.49	0.27	0.92	1.04	2.61	0.63	1.95
Truck	4.33	6.03	4.74	6.40	3.92	5.94	4.41	6.22	3.84	5.90	3.92	6.40	6.62	6.98	1.55	3.38	4.24	6.14
Blocks	2.12	3.93	2.76	4.72	2.84	4.75	2.91	5.12	2.90	4.62	3.08	5.70	3.03	4.89	2.58	4.72	2.82	4.81
Rising Towers	6.40	6.52	6.49	6.59	6.85	6.86	6.75	6.85	6.28	6.65	5.74	6.45	6.60	6.66	6.53	6.79	6.56	6.72
Clay	1.25	2.30	1.46	3.57	1.55	3.87	1.28	3.35	1.47	3.45	1.66	4.33	1.36	3.49	1.52	3.65	1.43	3.56
Crayons	1.93	4.68	1.13	3.18	1.37	3.73	1.33	3.78	1.33	3.87	1.73	4.17	0.87	3.04	1.90	5.35	1.35	3.75
Magic Markers	4.29	6.62	4.25	6.28	4.59	6.78	4.63	6.68	5.05	6.94	4.32	6.27	4.02	6.28	5.23	6.98	4.58	6.64
Paper	2.25	4.78	1.21	3.52	1.49	4.02	1.64	4.31	1.62	4.27	1.00	3.48	1.32	3.86	1.74	4.31	1.52	4.08
Coloring Book	3.45	6.50	2.93	5.71	3.75	6.46	3.48	6.32	4.26	6.93	4.11	6.78	2.73	5.75	4.57	6.91	3.59	6.38

*N's: 96, 320, 338, 391, 276, 62, 787, 696, and 1483, respectively.

Table 6

Mean Complexity of Play with Each Object
(Year 1)

	Age in Months																	
	42-44		45-47		48-50		51-53		54-56		57-59		Boys		Girls		Total	
Object	N	M	N	M	N	M	N	M	N	M	N	M	N	M	N	M	N	M
Black Doll	28	2.57	60	2.75	74	2.60	73	2.71	46	2.68	12	2.52	96	2.762	197	2.617	293	2.66
White Doll	28	2.59	57	2.80	70	2.63	69	2.75	41	2.72	10	2.69	89	2.771	186	2.673	275	2.70
Truck	52	3.19	170	3.19	166	3.28	208	3.26	135	3.29	26	3.36	532	3.295	225	3.159	757	3.25
Blocks	34	3.36	120	3.30	144	3.25	153	3.31	115	3.37	20	3.39	335	3.402	251	3.18[illegible]	586	3.31
Rising Towers	62	3.19	229	3.15	233	3.16	273	3.13	183	3.12	39	3.25	546	3.225	473	3.062	1019	3.15
Clay	26	2.99	91	2.84	91	2.79	97	2.77	76	2.84	14	2.52	205	2.818	190	2.801	395	2.81
Crayons	29	2.93	64	2.92	71	2.93	73	2.90	56	3.02	13	2.92	121	2.915	185	2.951	306	2.94
Magic Markers	35	2.98	146	2.88	145	2.90	176	2.89	129	2.92	31	2.77	319	2.896	343	2.892	662	2.89
Paper	27	2.95	51	2.94	67	3.02	74	3.02	50	3.04	8	3.00	131	2.983	146	3.020	277	3.00
Coloring Book	28	3.06	104	2.92	114	3.00	129	2.96	99	3.01	22	2.99	209	2.964	287	2.986	496	2.98

Table 7

Mean Number of Periods of Play with Each Object*
(Year 2)

	Age in Months																	
	51-53		54-56		57-59		60-62		63-65		66-69		Boys		Girls		Total	
Object	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Black Doll	0.76	1. 9	0.56	1.49	0.95	2.36	0.86	1.94	0.84	1.85	3.44	5.62	0.47	1.45	1.21	2.45	0.82	2.02
White Doll	0.59	1.59	0.48	1.23	0.86	2.17	0.69	1.68	0.80	2.13	1.44	1.94	0.39	1.18	1.04	2.24	0.70	1.80
Truck	4.04	5.81	3.45	4.98	4.67	5.91	3.97	5.60	4.16	5.48	2.89	2.21	6.23	6.12	1.64	3.41	4.04	5.51
Blocks	2.76	4.73	3.02	4.79	3.65	5.25	3.27	4.98	3.33	4.71	6.11	6.29	4.15	5.35	2.37	4.30	3.30	4.95
Rising Towers	4.95	6.18	5.41	6.38	5.45	6.18	5.75	6.16	6.83	6.43	8.89	8.52	5.92	5.92	5.55	6.70	5.74	6.30
Clay	3.84	5.69	3.42	4.98	2.93	5.14	2.99	4.61	2.60	4.47	0.11	0.33	2.70	4.38	3.45	5.36	3.06	4.89
Crayons	0.65	2.30	0.96	3.13	0.80	2.60	0.73	2.83	0.98	3.07	1.67	5.00	0.38	1.85	1.35	3.62	0.84	2.87
Magic Markers	5.57	7.22	4.68	6.80	4.44	6.85	4.59	6.85	3.65	6.35	0.00	0.00	3.45	5.93	5.57	7.43	4.46	6.77
Paper	1.53	4.32	1.34	3.79	1.83	4.41	1.85	4.34	1.38	3.73	0.00	0.00	1.22	3.50	2.02	4.66	1.60	4.11
Coloring Book	3.67	6.42	4.21	6.88	2.85	5.85	3.22	6.20	3.16	5.95	1.78	4.97	2.37	5.22	4.48	7.07	3.38	6.26

*N's: 75, 210, 210, 248, 147, 9, 470, 429, and 899, respectively.

Table 8

Mean Complexity of Play with Each Object
(Year 2)

	Age in Months																	
	51-53		54-56		57-59		60-62		63-65		66-69		Boys		Girls		Total	
Object	N	M	N	M	N	M	N	M	N	M	N	M	N	M	N	M	N	M
Black Doll	18	3.05	46	2.69	55	2.88	65	2.64	39	2.75	5	2.57	83	2.87	145	2.69	228	2.76
White Doll	16	3.20	43	2.71	51	2.98	56	2.71	38	2.71	4	2.25	78	2.98	130	2.70	208	2.81
Truck	37	3.26	110	3.18	132	3.22	137	3.18	88	3.19	7	3.39	357	3.24	154	3.10	511	3.20
Blocks	30	3.19	102	3.10	103	3.23	122	3.12	71	3.18	7	3.23	268	3.21	167	3.08	435	3.16
Rising Towers	46	3.20	142	3.05	142	3.11	170	3.08	115	3.13	6	2.94	353	3.14	268	3.05	621	3.10
Clay	39	2.83	105	2.91	85	2.87	117	2.76	66	2.74	1	2.00	210	2.83	203	2.82	413	2.82
Crayons	8	2.71	31	2.96	32	2.96	31	2.83	26	2.85	1	3.00	43	2.86	86	2.91	129	2.89
Magic Markers	36	2.99	89	2.94	82	2.91	102	2.96	47	2.92	0	0.00	162	2.92	194	2.96	356	2.94
Paper	12	3.01	35	2.88	42	3.06	61	2.96	28	2.96	0	0.00	83	2.95	95	3.00	178	2.97
Coloring Book	23	3.00	75	2.94	56	2.99	77	2.94	45	2.98	2	2.50	116	2.95	162	2.96	278	2.96

concordance, W, which is a generalization of rho. For number of periods of play with each object, W was .96 in Year 1, .94 in Year 2; for number of children playing with each object, it was .90 and .98, respectively; and for mean complexity of play, it was .93 and .73, respectively. All of these coefficients are significant at the .01 level. (For Year 2, the analysis used only five of the age periods; the number of children falling into the 66-69 month age interval was too small to provide reliable numbers.)

Finally, comparisons were made across years of testing. For the total sample tested in the two years, all three measures showed high consistency: rho was .83 for number of periods of play with each object, .85 for number of subjects playing with each object, and .99 for mean complexity of play with each object. Each of these coefficients is significant at the .01 level.

In summary, there was a moderate degree of consistency in play preferences across sex, and a high degree of agreement across age within year and across years of testing.

Relationship with Other Measures

In the factor analyses of both Year 1 and Year 2 data (13-factor solution, Varimax rotation; see Shipman, 1972) mean complexity of play and mean number of simple sequences defined a test-specific factor on which no other variable had a substantial loading. These two scores correlated with one another -.19 in each year's data for the longitudinal sample; they did not correlate as high as .20 with any other measure in the battery.

In the analysis of Year 1 data, a large general ability factor emerged. In the Varimax rotation, this factor split to some extent into perceptual and verbal abilities. Talks-to-tester loaded substantially on factor 4, the verbal component of ability, which contained moderately high loadings as well for the Sigel Object Categorizing Test (SOCT) grouping responses, and for the Talks-to-self measure. Its only correlation in the battery as high as .2, however, was with the PPVT, Form A (r = .23). Talks-to-self did not correlate this high with any other measure.

In the Year 2 data, in which there was no separation of general ability into verbal and perceptual components, there was no relation between Open Field verbalization measures and any ability measure. Factor 2 in the Varimax analysis was defined by the Talks-to-tester and Talks-to-self measures, along with the verbal elaboration score on the Preschool Inventory. In extension analyses, elaborations on the Social Schemata Test also correlated with this factor. The correlations between the two Open Field verbalization measures and the two verbal elaboration measures ranged from .25 to .31. Thus, a tendency toward spontaneous verbalization in a test situation, unrelated to measures of competence, appears to underlie the Open Field verbalization measures in Year 2.

Measures of sex typing in play showed little consistent relation to other measures in the battery. In Year 1, masculine play was associated with longer latencies on the SOCT, and feminine play, with shorter latencies, for both sexes; Matching Familiar Figures reaction time was negatively related to feminine play for males. In Year 2, masculine play for males was negatively associated with Massad Mimicry scores, while masculine play for females related negatively to the Mischel memory score. These correlations were all modest in magnitude and, given their apparent lack of systematic meaning, do not merit further interpretation at this time.

Comments

The Open Field Test was included in the Longitudinal Study test battery in the hope of assessing a number of variables which might not be manifested in a more standard testing situation. There is negative evidence--a general lack of correlation with the test measures, along with low communalities in the factor matrices--suggesting that what was measured was indeed different. In part, the separation of Open Field from other testing measures may be attributable to the difference in format between a "convergent" testing situation, with limited response alternatives and overall direction of activity by the tester, and the "divergent" play situation, with much broader freedom of activity for the child. In part, however, it may also be attributable to the choice of measures to be employed in the test batteries. To a large extent, only general information-processing skills were successfully measured by the battery; attempts to sample stylistic variables and personal-social characteristics were less successful. Several measures which, from earlier work, might have related to Open Field measures were included in the test battery (e.g., measures of habituation and of reflection-impulsivity) but the construct validity of these was in doubt in the Year 1 and 2 data. Thus, it is not clear how much of the separation is attributable to the situation and how much to the content of the test battery. Given this, and the generally low stability of Open Field measures from one year to the next, it is not now possible to determine how useful a source of correlates to other sources of data, more stylistic and personality in orientation, will be found in Open Field measures.

References

Goldberg, S. & Lewis, M. Play behavior in the year-old infant: Early sex differences. Child Development, 1969, 40, 21-31.

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, 1969.

Kagan, J. The three faces of continuity in human development. In D. Goslin (Ed.), Handbook of socialization theory and research. New York: Rand McNally, 1969.

Kagan, J. Change and continuity in infancy. New York: John Wiley & Sons, 1971.

Messer, S. B., Kagan, J. & McCall, R. B. Fixation time and tempo of play in infants. Developmental Psychology, 1970, 3, 406.

Messer, S. B. & Lewis, M. Social class and sex differences in the attachment and play behavior of the year-old infant. Paper presented at the annual meeting of the Eastern Psychological Association, Atlantic City, April 1970.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

YEARS 1 and 2

OPEN FIELD TEST

CODING MANUAL

In this test, the child is given 10 minutes of free play with a standard set of toys. The tester sits in the room, recording what the child does during each of the 20 1/2-minute periods while the child is playing. Each column on the answer sheet is used to record what happens in a specific time period; each row is used to record a particular object played with or a specific nonplay activity. The tester was instructed to put a check in each cell, as appropriate, and also to provide numbers for the cells. The numbers refer to descriptions, at the bottom of the page, of just what activity is going on.

Where two or more objects were played with in a single time period, the tester should have drawn a circle around the checks for those two objects if they were used together, but no circle if the child played with first one and then the other. The description at the bottom of the page should also indicate whether they were used together or separately. However, testers tended to be inconsistent or idiosyncratic in how they numbered and described activities; it is necessary to read the record as a whole before trying to interpret what went on.

You will find cases where the descriptions indicate that not all the appropriate cells were checked; for example, the tester may check "talks to E" and comment at the bottom that the child said he wanted to leave the room. Here both "talks to E" and "wants out" should have been checked. Another common error is to record "Other," when the comments indicate that the child was doing "Nothing." (Where there are discrepancies between the cells checked and the descriptions, you will code according to the descriptions, if they give a clear indication of what went on.)

In coding, you will be doing three jobs at once.

1. You will be eliminating the differences between testers in the ways in which behavior was recorded. This is mainly a case of reading the record carefully to make sure the various cells are checked appropriately.

2. You will be reducing the variety of behaviors observed to a relatively small number of categories, ones which will let us look at the complexity and the sequencing of the child's activity without being concerned with every detail of what he is doing.

3. You will be recording this cleaned-up and simplified set of observations in machine-readable form, so that the computer can be used to derive the various summary scores we want from the data.

The categories to be used are described below:

Playing with the Toys

Here, for each of the 10 play objects, you will be concerned with the level of complexity of the child's activity. When one or more of these objects are used in a time period, this will be coded as follows:

Code 1

If the child just looks at the object without handling or playing with it.

Code 2

If the child touches or handles the object without using it in any play activity. This often happens when the child seems to be examining a toy before deciding whether or not to play with it.

Examples:
a. Holds blocks or rising towers.
b. Holds a doll without doing anything more active (like hugging or rocking).
c. Examines crayons or magic markers but is not using paper or coloring book yet. (Child may use markers as a toy to stack or roll--this activity is coded 3.)
d. Child merely sets an object on the floor without beginning a play activity.
e. When a child examines a truck to see how it works or examines a doll's clothes or hair, this is coded 3 since this would be a normal part of play for these toys.

Code 3

If the child uses the object in any play activity, so long as another object is not used at the same time.

a. The 3 category is also used when the description makes it unclear what the level of activity is, or when there is no description of the activity.

b. If the activity is coloring, whether with crayons or magic markers, and whether on paper or coloring book, the objects are coded 3 (rather than 4) since this is the expected activity with crayons and coloring book.

Code 4

If the child uses the object in a play activity that also involves some other object or objects.

Examples:
a. Loads blocks onto truck (code 4 for blocks <u>and</u> truck).
b. Colors rising towers with crayons (since this is a more complex activity than coloring on paper).
c. Interaction with both dolls--in this case it must be clear that the dolls are interacting (ex. making them kiss, dance, etc.). Merely holding or rocking both dolls is <u>not</u> a 4 level activity.
d. Stacking magic markers, rising towers, and blocks together.
e. If the tester notes that the child pushed the truck <u>to</u> the blocks, only code 3 since the blocks are not actually involved in an activity.

General Comments on Play Activity

a. Only one code number may be used for any object in a single time period. If there is more than one level of play in a single time period, use the highest number that applies.

b. Any number of objects may be scored in a time period; all that is necessary is that each be used, however briefly, in that period.

c. The test is concerned only with the toys (and their containers) as listed. If the child plays with a chair or something else in the room there is no code recorded for that time period; it is simply left blank. imilarly, if the child stacks the rising towers on a chair, he gets a 3 for rising towers luring that time period.

Talks to E

Code 1

For verbalizations that try to draw the tester's attention to the activity.

Examples: a. "Look at this picture."
b. "I want to build a high tower."
c. "Blocks are fun to play with."

Code 2

If the child asks for help or direction.

Examples: a. "Can I play with this?"
b. "What are these?"
c. "How does this truck work?"
d. "Do you have any more red blocks?"

Code 3

If the child wants to leave or to stop the play activity.

Examples: a. "I want to eat my candy now."
b. "I have to go to the bathroom."
c. "I'm tired of playing."
d. "Where is my mother?"

If the tester marks "wants out" and writes that the child wants to go home, it is assumed that the child talked to the tester and this should be marked as 3 under "talks to E" in addition to being marked under "wants out."

Code 4

For any other verbalization. This will be primarily talk which is conversational in nature, without involving the child's play activity or a request for help.

Examples: a. "My house is full of toys."
b. "I will go to the circus today."

Also code 4 when the tester does not describe the verbalization.

General Comments on "Talks to E"

a. If the child talks about one of the play objects but is not marked as playing with it during the period in question, credit him with looking at the object during this period.

b. However, if the comment is "Looked at all the toys before starting to play," do not record all the toys as 1--just leave the time blank until the child begins another level of play.

Talks to Self

Code 1

If the talk is unrelated to play activity. This level includes singing or humming unless this is connected to the play as when singing to the dolls.

Code 2

If the talk is related to the play activity.

Examples: a. Making truck or motor noises while playing with the truck.
b. Talking to the dolls.
c. Planning the play activity aloud.
d. Counting the blocks as they are stacked.

Code 3

If there is no description of the child's verbalization.

Approaches E

Code 1

If the child moves toward or stays with the tester, giving her his attention.

Wants Out

Code 1

If the child indicates that he wants to leave the play situation; indication may be by saying he wants to leave, by trying to get out, etc. (A verbalization means that both "talks to E" and "wants out" should be recorded.)

Nothing

Code 1

If the child does nothing at all, for example, just sits and stares at the wall or out the window. If the child is doing anything--e.g., talking to E, walking around the room, trying to get out, etc., do not record as "nothing."

Other

Not being used for coding; if some behavior is described under "Other" check that it doesn't belong in one of the other categories; if it does not, do not record the activity.

Example: If the child does handstands or skips around the room, there is no activity recorded on the score sheet.

Miscellaneous

Code 1

If the child breaks the play activity by leaving the room (for example, to go to the bathroom).

Code 2

If the tester stops the play activity.

Code 3

If there is someone else in the room during the play activity.

Code 4

If the child refuses to do the test, either all or part of it.

REVISED AS OF MAY 1, 1964

FOR KEYPUNCH USE ONLY

EDUCATIONAL TESTING SERVICE

DATA PROCESSING DIVISION

PUNCHED CARD LAYOUT

Program/Project Longitudinal Study

P/J ______ Study Code ______ JR No. ______

UNITS ______ TOTAL ______ Card Description Open Field Test Card No. 01

______ Alpha (______ Columns) Issue Date ______ No. of Cases ______

______ Numeric (______ Columns) Job Initiator ______ Section Head ______

1	CHILD ID#		1	41	WHITE DOLL 10	41
2			2	42	1/2	42
3			3	43		43
4			4	44		44
5			5	45		45
6			6	46		46
7			7	47		47
8	Inst #	Administration (1)	8	48		48
9		Task (20)	9	49		49
10			10	50		50
11	Card # (01)		11	51		51
12			12	52		52
13	Examiner ID		13	53		53
14			14	54		54
15			15	55		55
16	Day of Testing	Month	16	56		56
17			17	57		57
18		Day	18	58		58
19			19	59		59
20		Year	20	60		60
21			21	61	10	61
22	BLACK DOLL	1/2	22	62	(blank)	62
23		1	23	63		63
24		1 1/2	24	64		64
25		2	25	65		65
26		2 1/2	26	66		66
27		3	27	67		67
28		3 1/2	28	68		68
29		4	29	69		69
30		4 1/2	30	70		70
31		5	31	71		71
32		5 1/2	32	72		72
33		6	33	73		73
34		6 1/2	34	74		74
35		7	35	75		75
36		7 1/2	36	76		76
37		8	37	77		77
38		8 1/2	38	78		78
39		9	39	79		79
40		9 1/2	40	80		80

EXPLANATION

REVISED AS OF MAY 1, 1964

EDUCATIONAL TESTING SERVICE

DATA PROCESSING DIVISION

PUNCHED CARD LAYOUT

FOR KEYPUNCH USE ONLY

Program/Project Longitudinal Study

P/J ______ Study Code ______ JR No. ______

UNITS ______ TOTAL ______ Card Description Open Field Test Card No. 02

______ Alpha (______ Columns) Issue Date ______ No. of Cases ______

______ Numeric (______ Columns) Job Initiator ______ Section Head ______

1–10	(Same as Card 01)	
11–12	Card # (02)	
13–32	TRUCKS	1/2 ↓ 10
33–40	BLOCKS	1/2 ↓
41–52	RISING TOWERS	↓ 10
53–72	RISING TOWERS	1/2 ↓ 10
73–80	(blank)	

EXPLANATION

REVISED AS OF MAY 1, 1964

FOR KEYPUNCH USE ONLY

EDUCATIONAL TESTING SERVICE

DATA PROCESSING DIVISION

PUNCHED CARD LAYOUT

Program/Project Longitudinal Study

P/J ______ Study Code ______ JR No. ______

UNITS ______ TOTAL

Card Description Open Field Test Card No. 03

______ Alpha (____ Columns) Issue Date ______ No. of Cases ______

______ Numeric (____ Columns) Job Initiator ______ Section Head ______

1–10	(Same as Card 01)
11–12	Card # (03)
13–32	CLAY 1/2 → 10
33–40	CRAYONS 1/2 → 10
41–52	10
53–72	MAGIC MARKER 1/2 → 10
73–80	(blank)

EXPLANATION

REVISED AS OF MAY 1, 1964

EDUCATIONAL TESTING SERVICE

DATA PROCESSING DIVISION

PUNCHED CARD LAYOUT

FOR KEYPUNCH USE ONLY

Program/Project: Longitudinal Study

P/J ____ Study Code ____ JR No. ____

UNITS ____ TOTAL ____ Card Description: Open Field Test Card No. 04

____ Alpha (____ Columns) Issue Date ____ No. of Cases ____

____ Numeric (____ Columns) Job Initiator ____ Section Head ____

1			1	41			41
2			2	42			42
3			3	43			43
4	(Same as Card 01)		4	44			44
5			5	45	B		45
6			6	46	O		46
7			7	47	O		47
8			8	48	K		48
9			9	49			49
10			10	50			50
11	Card # (04)		11	51			51
12			12	52			52
13	Time of Day		13	53		10	53
14	Sequence		14	54			54
15		1/2	15	55			55
16			16	56			56
17			17	57			57
18			18	58			58
19			19	59			59
20			20	60			60
21	P		21	61	(blank)		61
22	A		22	62			62
23	P		23	63			63
24	E		24	64			64
25	R		25	65			65
26			26	66			66
27			27	67			67
28			28	68			68
29			29	69			69
30			30	70			70
31			31	71			71
32			32	72			72
33			33	73			73
34		10	34	74			74
35		1/2	35	75			75
36	C		36	76			76
37	O		37	77			77
38	L		38	78			78
39	O		39	79			79
40	R		40	80			80

EXPLANATION

REVISED AS OF MAY 1, 1964

EDUCATIONAL TESTING SERVICE

DATA PROCESSING DIVISION

PUNCHED CARD LAYOUT

FOR KEYPUNCH USE ONLY

Program/Project Longitudinal Study

P/J ______ Study Code ______ JR No. ______

UNITS ______ TOTAL

Card Description Open Field Test Card No. 05

______ Alpha (______ Columns) Issue Date ______ No. of Cases ______

______ Numeric (______ Columns) Job Initiator ______ Section Head ______

1–10	(Same as Card 01)	
11–12	Card # (05)	
13–32	TALK TO E	1/2 … 10
33–52	TALK TO SELF	1/2 … 10
53–72	TOWARD E	1/2 … 10
73–80	(blank)	

EXPLANATION

REVISED AS OF MAY 1, 1964

FOR KEYPUNCH USE ONLY

EDUCATIONAL TESTING SERVICE

DATA PROCESSING DIVISION

PUNCHED CARD LAYOUT

Program/Project Longitudinal Study

P/J ______ Study Code ______ JR No. ______

UNITS ______ TOTAL

Card Description Open Field Test Card No. 06

______ Alpha (______ Columns) Issue Date ______ No. of Cases ______

______ Numeric (______ Columns) Job Initiator ______ Section Head ______

Col.	Field		Col.
1			1
2			2
3			3
4			4
5			5
6	(Same as Card 01)		6
7			7
8			8
9			9
10			10
11			11
12	Card # (06)		12
13		1/2	13
14			14
15			15
16			16
17	W		17
18	A		18
19	N		19
20	T		20
21	S		21
22			22
23	O		23
24	U		24
25	T		25
26			26
27			27
28			28
29			29
30			30
31			31
32		10	32
33		1/2	33
34	N		34
35	O		35
36	T		36
37	H		37
38	I		38
39	N		39
40	G		40

Col.	Field		Col.
41			41
42			42
43			43
44			44
45			45
46			46
47			47
48			48
49			49
50			50
51			51
52		10	52
53		1/2	53
54			54
55			55
56	M		56
57	I		57
58	S		58
59	C		59
60	E		60
61	L		61
62	L		62
63	A		63
64	N		64
65	E		65
66	O		66
67	U		67
68	S		68
69			69
70			70
71			71
72		10	72
73			73
74			74
75			75
76	(blank)		76
77			77
78			78
79			79
80			80

EXPLANATION

OPEN FIELD TEST

Answer Sheet

Name________________________________

I.D. Number________________ Examiner I.D.__________________ Date __/__/__

NOW WE HAVE SOME NICE THINGS FOR YOU TO PLAY WITH. I HAVE SOME THINGS TO DO SO I'LL SIT OVER HERE WHILE YOU PLAY. YOU CAN PLAY WITH ANYTHING YOU WANT TO.

Time (Minutes)

	0-½	½-1	1-1½	1½-2	2-2½	2½-3	3-3½	3½-4	4-4½	4½-5
Doll (black)										
Doll (white)										
Truck										
Blocks										
Rising Towers										
Clay										
Crayons										
Magic Markers										
Paper										
Coloring Book										
Talks to E										
Talks to Self										
Approaches E										
Wants Out										
Nothing										
Other (specify)										

Description of Activity

1 ________________________________

2 ________________________________

3 ________________________________

4 ________________________________

5 ________________________________

6 ________________________________

7 ________________________________

8 ________________________________

9 ________________________________

10 ________________________________

11 ________________________________

12 ________________________________

OPEN FIELD TEST

Time (Minutes)

	5–5½	5½–6	6–6½	6½–7	7–7½	7½–8	8–8½	8½–9	9–9½	9½–10	
Doll (black)											
Doll (white)											
Truck											
Blocks											
Rising Towers											
Clay											
Crayons											
Magic Markers											
Paper											
Coloring Book											
Talks to E											
Talks to Self											
Approaches E											
Wants Out											
Nothing											
Other (specify)											

Description of Activity

13 ______
14 ______
15 ______
16 ______
17 ______
18 ______
19 ______
20 ______
21 ______
22 ______
23 ______
24 ______
25 ______
26 ______
27 ______

OPEN FIELD TEST

Description of Product

Blocks

Rising Towers

Clay

Drawing

Other and Comments

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

Technical Report 15

Peabody Picture Vocabulary Test

Judith A. Meissner
Virginia C. Shipman
Lynn E. Gilbert

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health, Education, and Welfare

December 1972

Peabody Picture Vocabulary Test

Background

The "Peabody" is one of the best known and most widely used tests of verbal ability for young children. A measure of receptive language functioning, it is often used to determine a child's readiness for inclusion in a regular classroom setting. There are two alternate forms, A & B, with reported correlations ranging between .61 and .97. The correlation of the Peabody was .67 with the Verbal Scale of the WISC and .83 with the Stanford Binet with normal subjects (Dunn, 1965).

Normative data contained in the manual (Dunn, 1965) were based on a standardization sample (N = 4,012) consisting entirely of white children and youth in and around Nashville, Tennessee, many of whom were mentally, emotionally, or physically handicapped. Conversion tables are provided to transform the raw scores into standard scores or IQs. For young children, however, there are serious problems of interpretation in that children with the same raw score, but very small differences in age, receive widely discrepant IQ scores.

Several investigators have administered the Peabody to disadvantaged preschoolers. With a predominantly disadvantaged sample of 563 urban children, DiLorenzo and Brady (1968) obtained a correlation of .79 between the Peabody and the Stanford-Binet. Milgram and Ozer (1967) investigated the stability of Peabody scores of black urban preschoolers and compared them to scores on the Binet. The correlation between the Binet and Peabody IQ was .65 for 51 5-year-olds. The test-retest reliability of mental age scores based on two administrations of the Peabody was .69 for 5-year-olds tested four weeks apart, and .80 for 3-year-olds tested ten months apart.

Costello and Ali (1971) found the Peabody to correlate .28 with Stanford-Binet IQ in a sample of 60 urban disadvantaged children. Each of these three studies found lower scores on the Peabody than on the Binet. Rieber and Womack (1968) found differences in Peabody scores of low-income preschool children as a function of parents' educational level, family size and maternal employment. Hodges, McCandless and Spiker (1967) studied children from homes characterized by severe psycho-social deprivation and whose initial Stanford-Binet intelligence scores ranged between 50 and 85. Children who participated in an experimental preschool program made significantly higher post-test scores on the Peabody than did children in a control pre-school or children at home.

Working with 4-year-old disadvantaged black children in New York City, John and Goldstein (1964) found qualitative differences in verbal behavior. Analysis of the first 35 items revealed that the children were unable to identify action words, words related to rural living, and words whose referents were rare in low-income families. Jeruchimowicz, Costello and Bagur (1971) administered the Peabody to black preschoolers and found that children of low SES made a significantly higher proportion of errors on verbs than on nouns, whereas the middle-class group showed no difference.

The arrangement of items in ascending order of difficulty might contribute to the child's discouragement and the examiner's low expectancy (Zigler & Butterfield, 1968). A Modified Peabody Picture Vocabulary Test, consisting of the first 70 items randomized for difficulty level, was administered to a group of disadvantaged preschool children by Ali and Costello (1971). Test responses were positively reinforced according to a fixed schedule. Although the differences in scores were not significant, children scored higher on the modified version, and the test-retest reliability was .86 as compared to .77 for the standard instrument. The Huron Institute (Walker, Bane & Bryk, 1973) recommends further use and development of this d version.

Task Description and Administration

The Peabody test booklet consists of a set of bound cards, each containing four black and white drawings. There are three practice and 150 test plates, arranged in an ascending order of difficulty. The same booklet is used for both Form A and Form B, i.e., for a given item one of the four pictures presented is the "right" answer to the word presented on Form A and another picture in the same set is the "right" answer to the word stimulus on Form B. The test is untimed, and is usually administered in 15 minutes or less.

The standard Peabody test booklet (Dunn, 1965) was used in Year 1. In order to provide a more appropriate and less culturally-biased task, a special ETS adaptation of the first 60 items was developed which contained redrawings of a number of the human pictures to present black children and adults in a variety of roles. Race depicted and item choice were counterbalanced, as was portrayal in positive and negative roles. This ETS version was used in Year 2 of the study for both Forms A and B of the test.*

The Peabody was presented in two ways to the Longitudinal Study sample:

1) Form A was administered in the standard way in both Years 1 and 2 to obtain an estimate of receptive vocabulary. The tester presented the stimulus word orally and the child was required to point to one of the four pictured choices. Testing was terminated in Year 1 after S made six errors on eight consecutive items. Since the ETS-adapted test booklet was used in

---

*Shipman and Tanaka (1971) later extended this adaptation to all 150 plates for use in 1971-72 Head Start Planned Variation and Follow Through evaluation research.

Year 2, only the first 60 items were administered to all subjects.

2) Form B was administered in a modified form in both years in order to obtain an estimate of productive vocabulary. On this form, the tester pointed to the stimulus picture on the page and asked the child to tell what it was or, in the case of verbs, to say what the person in the picture was doing. In Year 1, testing continued until the criterion of six errors on eight consecutive items was reached. In Year 2, using the modified booklet, the first 60 items of the test were administered to all subjects.

Training was relatively simple, with cautions concerning inadvertent pointing or otherwise indicating the correct response. Special probes were given for handling multiple answers, elaboration of the various pictures on a page and other non-task defined responses. Considerable practice was needed, however, to train testers to use a criterion cutoff, as is the standard procedure.

Scoring

A total score consisting of one point for each correct item was calculated for each subject for Form A in both years and for Form B in Year 1. In Year 2, a modified scoring system was devised for Form B, in which a correct label to a stimulus picture received 3 points credit, a correct description received 2 points and a correct but vague association 1 point. By using this system, children received some credit for vague or partially correct responses, which had not been the case in Year 1.

In addition to the total score, subscores were obtained based on the percentage of verbs and nouns correctly identified out of those attempted. Because the mean difficulty level of nouns and verbs was not equivalent in the present Longitudinal Study data, and because verbs posed greater scoring difficulty than did nouns on Form B, noun-verb differences were not analyzed, and for the present analyses, only total scores were used.

The Peabody manual (Dunn, 1965) provided the key for correct answers for Form A. Since more than one word could serve as a correct response to an item in the modified version of Form B, lists of acceptable synonyms were generated for each item in Year 1 and for each score category in Year 2, and the child was given credit if he produced any one of the acceptable responses for an item. All scoring of the Peabody was done at the ETS Princeton office. Answer sheets were independently coded by two research assistants, with differences resolved by a senior research staff member.

Score Properties

The coefficient alpha reliability for Form A was .96 in Year 1 and .91 in Year 2. For Form B, the coefficients were .93 and .84, respectively. These coefficients are considerably higher than are the alternate form reliabilities given by Dunn (1965) for his standardization subjects of this same age (.77 for children in the age range 4-0 to 4-5 and .72 for the age range 4-6 to 5-0). Since the present study used a cutoff criterion in Year 1 and only 60 items in Year 2, these reliability estimates are probably inflated. Ali and Costello (1971) also obtained very high reliability for Peabody scores of disadvantaged preschoolers (r = .86), but since they only used the first 70 items of the test, this coefficient may be inflated also.

The coefficient of stability of .69 between Form A administered Year 1 and Year 2 was one of the highest stability coefficients found in the test battery and suggests that even though the test was slightly different in the two years, the same behaviors were being tapped. The Form B coefficient of stability was .58, which, although somewhat lower than that of Form A, reflects considerable consistency in performance over the two years. Thus, in accord with previous research, the test appears to have quite good psychometric properties for this age sample.

Sample Characteristics

A. Form A (Receptive Vocabulary)

Tables 1 and 2 present means, standard deviations and percentile distributions of total score on Form A by age and sex for all subjects for Years 1 and 2, respectively. Mean total correct score for the Year 1 administration of Form A was 26.3 (N = 1198; SD = 12.85); in Year 2, mean number of correct items obtained by 1309 subjects was 41.6, with a standard deviation of 9.75. Caution must be used in comparing these results from Years 1 and 2, due to the somewhat different composition and method of administration of the test in the two years. Within each of the two years, however, a developmental age trend is definitely apparent, the mean scores in Tables 1 and 2 gradually increasing with each older age group.

Table 1

Means, Standard Deviations and Percentile Distributions of Total Score on Form A by Age and Sex in Year 1

				Percentiles				
Group	N	Mean	SD	10	25	50	75	90
42-44 mo.	68	22.8	10.45	10.58	15.30	20.85	30.10	36.39
45-47 mo.	273	23.5	11.95	9.45	14.33	21.83	31.56	40.46
48-50 mo.	276	25.2	12.66	9.93	14.64	23.81	34.17	44.19
51-53 mo.	313	28.3	13.32	11.18	17.86	26.76	38.85	47.01
54-56 mo.	217	28.6	13.43	11.73	17.48	26.50	39.53	48.34
57-59 mo.	51	30.7	11.37	13.63	22.88	31.16	40.92	44.46
Boys	630	25.9	13.10	10.03	15.36	24.10	35.45	45.56
Girls	568	26.8	12.58	11.20	16.47	24.91	36.55	44.77
Total	1198	26.3	12.85	10.55	15.91	24.43	36.05	45.21

Table 2

Means, Standard Deviations and Percentile Distributions of Total Score on Form A by Age and Sex in Year 2

				Percentiles				
Group	N	Mean	SD	10	25	50	75	90
51-53 mo.	79	39.2	8.70	27.4	32.1	40.6	46.6	49.6
54-56 mo.	309	39.5	10.11	26.2	31.8	39.7	47.5	53.3
57-59 mo.	305	40.8	9.67	27.1	33.6	41.8	48.3	53.0
60-62 mo.	350	42.8	9.40	30.3	37.0	43.9	50.3	54.0
63-65 mo.	248	43.8	9.59	30.0	36.7	45.7	51.3	55.1
66-69 mo.	18	45.1	7.61	35.3	41.2	45.3	50.0	53.1
Boys	696	41.3	10.06	26.9	33.8	42.5	49.2	53.7
Girls	613	41.9	9.38	29.0	35.1	42.7	49.5	53.7
Total	1309	41.6	9.75	28.0	34.5	42.6	49.4	53.7

Although the descriptive data in Tables 1 and 2 were obtained by using all subjects in order to gather as much normative data as possible, subsequent analyses utilized only the longitudinal subjects of the study (i.e., the subjects for whom data were available in both Years 1 and 2). Because of the previously-discussed new test version and slight methodological differences in task administration in Years 1 and 2, it was deemed more appropriate to perform separate ANOVAS on the Year 1 and Year 2 data rather than combining them in a repeated-measures analysis.

Age data were divided at the median in order to obtain an "older" and a "younger" age group. The difference between total correct score means for these two groups was highly significant in both Year 1 (32.43 for the older group and 28.10 for the younger; $F = 43.97$, $df = 1/1099$, $p < .001$) and in Year 2 ($M = 44.39$ and 41.57, respectively; $F = 30.62$, $df = 1/1122$, $p < .01$). The normative data for the Peabody (Dunn, 1965), which grouped subjects at age intervals of from six months to one year, also found considerable mean age differences in total score performance.

The Year 1 analysis of variance found a significant difference between the means for boys (29.5) and girls (31.0) on total items correct (F = 5.34, df = 1/1099, $p < .025$), but in Year 2, sex differences were not significant (boys' mean = 42.7; girls' mean = 43.2, F = .35). Apparently, the question of whether there are stable sex differences in receptive vocabulary knowledge (as measured by the Peabody) for this sample will have to await analysis of further longitudinal data, although the above results do suggest that sex differences may be age-specific.

In order to obtain a rough measure of socioeconomic status, the longitudinal subjects were divided into three groups on the basis of their mothers' educational level. High SES subjects were those children whose mothers had more than twelve years of schooling, middle SES subjects had mothers with more than ten but less then twelve years of schooling, and low SES subjects' mothers had fewer than ten years of schooling. SES differences in mean number of items correct on Form A were highly significant in both Years 1 and 2 of the study. In Year 1, the F-value was 92.01 (df = 2/1099, $p < .001$), with means of 23.79 for the low SES subjects, 27.13 for middle SES subjects and 39.87 for the high SES group. In Year 2, the means for these three groups were 36.17, 41.71 and 51.05, respectively, yielding an F of 138.63 (df = 2/1122, $p < .001$).

B. Form B (Productive Vocabulary)

In Tables 3 and 4, means, standard deviations and percentile distributions of total score on the productive vocabulary task, Form B of the Peabody, are presented by age and sex for Years 1 and 2 for all subjects. In Year 1, overall mean number of correctly-described task items for the

991 subjects adminstered the task was 19.1, with a standard deviation of 8.91. In Year 2, with a slightly different method of task administration, mean number of items correct was 31.4, with a standard deviation of 6.33. With the exception of the oldest age group of subjects (the group 57-59 months old in Year 1 and 66-69 months old in Year 2), there was a developmental trend with age, each successive age group obtaining a slightly higher mean number of correct item descriptions.

Table 3

Means, Standard Deviations and Percentile Distributions of Total Score on Form B by Age and Sex in Year 1

				Percentiles				
Group	N	Mean	SD	10	25	50	75	90
42-44 mo.	51	15.5	7.15	7.09	9.67	14.07	20.98	25.38
45-47 mo.	221	17.2	7.84	6.57	11.95	16.52	22.24	26.37
48-50 mo.	223	18.0	8.18	7.32	11.46	18.05	23.50	28.03
51-53 mo.	249	20.1	9.64	8.85	14.12	18.87	24.53	30.95
54-56 mo.	195	21.8	9.35	10.37	15.66	21.91	26.43	35.34
57-59 mo.	52	20.7	8.97	8.27	14.51	21.54	24.55	33.43
Boys	524	19.5	9.54	7.48	12.95	18.71	24.43	32.30
Girls	467	18.7	8.12	8.39	13.02	18.62	23.58	27.50
Total	991	19.1	8.91	7.90	12.99	18.67	23.97	29.36

Table 4

Means, Standard Deviations and Percentile Distributions
of Total Score on Form B by Age and Sex in Year 2

Group	N	Mean	SD	10	25	Percentiles 50	75	90
51-53 mo.	75	29.1	5.91	21.4	24.8	29.3	33.1	35.7
54-56 mo.	205	30.4	6.57	21.9	25.8	30.6	34.9	38.9
57-59 mo.	203	31.0	6.06	23.8	26.9	31.4	35.1	38.4
60-62 mo.	246	32.6	6.03	24.6	29.1	32.9	36.5	40.4
63-65 mo.	150	32.8	6.39	24.2	29.4	32.9	37.8	40.8
66-69 mo.	14	30.9	7.11	20.5	26.9	31.2	35.6	40.2
Boys	470	32.4	6.31	24.2	29.0	32.8	36.7	40.5
Girls	423	30.3	6.17	22.4	25.8	30.4	34.6	38.6
Total	893	31.4	6.33	23.1	27.3	31.6	35.7	39.8

The Year 1 analysis of variance using only the longitudinal subjects of the study (those subjects who were tested in both Years 1 and 2) found no significant sex differences, but mean differences between the older (M = 21.5) and the younger (M = 17.5) age group were highly significant (F = 42.94, df = 1/759, $p < .001$), as were SES differences, defined by the mother's level of schooling (F = 64.73, df = 2/759, $p < .001$). Children of mothers with less than 10 years of schooling obtained a mean score of 14.8 items described correctly, those whose mothers had between 10 and 12 years of schooling had a mean of 18.3 correct items, and children of mothers with more than 12 years of schooling had a mean of 25.3 items correct.

In Year 2, all three variables of age, sex and SES showed differences significant beyond the .001 level. Older children obtained higher mean scores than did younger ones (33.7 and 31.0, respectively; F = 24.86, df = 1/709), boys performed better than girls (obtaining means of 33.3 and 31.4 respectively; = 22.58), and children with more highly-educated mothers scored higher as a group than did children whose mothers had less schooling (means of 36.2, 31.4 and 29.4, resulting in an F of 34.01, df = 2/709).

C. Comparison of Receptive and Productive Scores

As would be expected from previous research (Fraser, Bellugi & Brown, 1963; Meissner, Fish & MacGinitie, 1972; Stern & Bryson, 1970), children performed better when only receptive language skills were required. In Year 1, overall mean score was more than 7 points higher on Form A (the receptive task) than on the modified Form B (productive task); total means were 26.3 and 19.1, respectively. By Year 2, the Form A-Form B discrepancy in mean items correct was larger than 10 points (41.6 and 31.4 points, respectively). T tests comparing these Form A-Form B mean differences were significant beyond the .001 level for both Years 1 and 2 (t = 15.3 for Year 1 and 30.9 for Year 2). Much of this discrepancy in difficulty is probably a result of differences in task requirements between receptive and productive language tasks, the receptive task simply calling for the child to recognize and point to the appropriate picture, but the productive task requiring the child not only to recognize but also to produce a correct label or description of the stimulus object.

Relationship with Other Measures

Evidence for the validity of the two Peabody tasks as measures of receptive and expressive vocabulary knowledge comes from the intercorrelations with other measures used in the Longitudinal Study test battery and with the factor loadings of these two tasks.

Tables 5 and 6 show all correlations above .30 of Peabody Form A and modified Form B with all other tests in the battery, for Years 1 and 2, respectively, for all subjects who were tested both years. It can be seen from these two tables that the correlations of the two forms with other cognitive measures are quite similar, although the correlations with Form A are of slightly higher magnitude than Form B in every case except that of the ETS Story Sequence Task, and the Year 2 correlations are, in most cases, slightly higher than the Year 1 correlations with the same task.

Table 5

Correlations of All Tests Loading Above .30
with Peabody Form A and Form B Total Scores in Year 1

Test and Score	Correlation with Form A	Correlation with Form B
Preschool Inventory (Caldwell): Adjusted total score	.58	.50
TAMA General Knowledge Test: Total score	.52	.42
Children's Auditory Discrimination Inventory: Nonsense words score	.52	.41
Matching Familiar Figures: Mean errors per valid item	-.43	-.32
Seguin Form Board: Log fastest time for correct placement	-.40	-.38
Form Reproduction: Total score	.40	.33
Sigel Object Categorization: Total grouping responses	.40	.34
Hess-Shipman Eight-Block Sorting Task: Total score	.39	.39
Motor Inhibition Test: Average "slow" time	.36	.32
ETS Matched Pictures: Total score	.36	.26
ETS Story Sequence Task: Part I (Receptive Language) score	.35	.36
Johns Hopkins Perceptual Test: Total score	.32	.25
Preschool Embedded Figures Test: Total score	.31	.23
Massad Mimicry: Meaningful word phrases - Final Sounds	.28	.32
Peabody Picture Vocabulary Test, Form A	--	.69

Note.--Sample sizes for the above tasks ranged from 519 to 733.
$r_{.001} = .148$ for $N = 500$.

Table 6

Correlations of All Tests Loading Above .30
with Peabody Form A and Form B Total Scores in Year 2

Test and Score	Correlation with Form A	Correlation with Form B
Preschool Inventory (Caldwell): Adjusted total score	.66	.63
TAMA General Knowledge Test: Total score	.63	.60
Hess-Shipman Eight-Block Sorting Task: Total score	.53	.46
Matching Familiar Figures: Mean errors per valid item	-.50	-.50
Children's Auditory Discrimination Inventory: Nonsense words score	.47	.41
Seguin Form Board: Log fastest time for correct placement	-.46	-.45
Form Reproduction: Total score	.46	.41
ETS Matched Pictures II: Total score	.44	.44
Massad Mimicry Test: Meaningful word phrases - final sounds	.44	.41
Sigel Object Categorization: Total grouping responses	.44	.43
ETS Enumeration II: Counting score (Items 1-4)	.44	.37
Johns Hopkins Perceptual Test: Total score	.37	.33
Motor Inhibition: Average "slow" time	.34	.28
ETS Story Sequence Task: Part I (Receptive Language) score	.33	.35
Spontaneous Numerical Correspondence: Total score	.32	.28
Peabody Picture Vocabulary Test, Form A	--	.73

Note.--Sample sizes ranged from about 800 to more than 1200.
$r_{.001} = .148$ for $N = 500$.

The highest correlations of both forms in both years are with the Preschool Inventory and the TAMA General Knowledge Test, and range from .42 to .66. The correlation obtained with the Preschool Inventory is almost identical to that reported by Datta (1967) for four- and five-year-old Head Start children (r = .69 and .62, respectively). Forms A and B both had moderate correlations (ranging from .26 to .44) with measures of more general language skills such as comprehension of syntax (Matched Pictures) and comprehension of sequence (ETS Story Sequence). Of all the measures used in the test battery, the Peabody, along with the Preschool Inventory, had the highest correlations with other cognitive tasks. The relatively high communality of the Peabody, Form A (.64 in Year 1 and .59 in Year 2) indicated that it had considerable variance in common with the other variables in the structural analyses. The fact that both Form A and Form B were tapping similar verbal abilities is reflected in their intercorrelation of .69 in Year 1 and .73 in Year 2.

Factor analyses of both the Year 1 and Year 2 data revealed the highest loadings of both forms of the Peabody to be on the factor best defined as "g," or general information-processing ability (Shipman, 1971, 1972).

Summary

For this sample during this age period, both Form A and modified Form B of the Peabody were found to have high internal consistency and stability across years, were relatively easy to administer and were enjoyed by most subjects. The modified Form B version used in this test shows that reliable and stable data on productive vocabulary can also be obtained from the test.

The test is not recommended in its original format, however. Many of its stimulus pictures are dated and are particularly inappropriate for minority children. Even more important, it is not known to what extent the depiction of black subjects in only two roles in 150 plates, that of railroad porter and native spear carrier, may be teaching a black child negative feelings about himself as he is being tested.

Both forms of the Peabody were highly correlated with other cognitive-perceptual tests in the Longitudinal Study battery. Like other vocabulary measures, the Peabody reflects one's ability to process general information from the environment. The highly significant SES differences found on both forms in both years suggests, however, that the Peabody taps "ability" only within a particular cultural context. Perhaps the high "g" factor loadings only indicate the importance of verbal comprehension skills as a common dimension in performance on these measures at this age; thus, the test is best viewed as a measure of receptive vocabulary and not general cognitive functioning. The present lack of stable sex differences may change in later study years as a result of differential reinforcement in school and other environmental influences.

References

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

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

Datta, L. Draw-A-Person as a measure of intelligence in preschool children from very low income families. Journal of Consulting Psychology, 1967, 31, 626-630.

DiLorenzo, L. T. & Brady, J. J. Use of the Peabody Picture Vocabulary Test with Preschool Children. Psychological Reports, 1968, 22, 247-251.

Dunn, L. M. Expanded Manual, Peabody Picture Vocabulary Test. Circle Pines, Minn.: American Guidance Service, Inc., 1965.

Fraser, C., Bellugi, U. & Brown, R. Control of grammar in imitation, comprehension and production. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 121-135.

Hodges, W. L., McCandless, B. R. & Spiker, H. H. The development and evaluation of a diagnostically based curriculum for preschool psycho-socially deprived children. U. S. Department of Health, Education, and Welfare, Project No. 5-0350, 1967.

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

John, V. P & Goldstein, L. S. The social context of language acquisition. Merrill-Palmer Quarterly, 1964, 100, 265-275.

Meissner, J., Fish, M. & MacGinitie, W. Concept communication in kindergarten children. Paper presented at the meeting of the Eastern Psychological Association, Boston, Ar 1 1972.

Milgram, N. A. & Ozer, M. N. Peabody Picture Vocabulary Test scores of preschool children. Psychological Reports, 1967, 20, 779-784.

Rieber, M. & Womack, M. The intelligence of preschool children as related to ethnic and demographic variables Exceptional Children, 1968, 34, 609-616.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. & Tanaka, M. Revised Adaptation of the Peabody Picture Vocabulary Test. Princeton, N. J. Educational Testing Service, 1971.

Stern, C. & Bryson, J. Competence vs. performance in the young child's use of adjectival comparatives. Child Development, 1970, 41, 1197-1201.

Walker, D. R., Bane, M. J. & Bryk, T. The quality of the Head Start Planned Variation data. Cambridge, Mass.: Huron Institute, 1973, in press.

Zigler, E. & Butterfield, E. Motivational aspects of changes in IQ test performance of culturally deprived nursery school children. Child Development, 1968, 39, 1-14.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

A LONGITUDINAL STUDY

Technical Report 16

Picture Completion Test
Wechsler Preschool and Primary Scale of Intelligence

Diran Dermen

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Picture Completion Test (WPPSI)

Background

The Picture Completion subtest of the Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 1967) was included in the present study as a measure of analytic functioning, an important aspect of psychological differentiation. Research over a period of more than twenty years has drawn a detailed picture of the nature of psychological differentiation (Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962; Witkin, Oltman, Raskin, & Karp, 1971). Witkin's early work on perceptual field dependence-independence provided the impetus for the original research and, as investigation continued, it became clear that individual consistencies or styles were not limited to the perceptual area, but were related to a complex of psychological, cultural, and even physiological consistencies. A brief presentation of such findings with an explication of the origins and nature of the work of Witkin and his associates is to be found in A Manual for the Embedded Figures Test (Witkin et al. 1971).

The specific choice of the Picture Completion subtest as a measure of analytic functioning was based on factor analytic studies of the WISC by Cohen (1959) and by Goodenough and Karp (1961). In these studies separate factors were found identifiable as "verbal comprehension" (Vocabulary, Comprehension, Information, and Similarities), "attention-concentration" or "memory" (Digit Span, Arithmetic, Coding), and "analytical ability" (Block Design, Object Assembly, Picture Completion). In the Goodenough and Karp study, the "analytical ability" factor was also marked by perceptual field independence measures and by measures of Guilford's adaptive flexibility factor. Unfortunately, at the time the present study was begun there were no factor analytic data available on the WPPSI subtests.

Recent studies (Witkin, personal communication, 1971; Witkin, Faterson, Goodenough, and Birnbaum, 1965) have found among subjects classified as mildly retarded a substantial number of individuals whose "verbal comprehension" factor scores are quite low, but who have near normal prorated IQs on the "analytical" factor. In mentally retarded boys in special public school classes the mean prorated IQ difference between the "verbal comprehension" and "analytical" scores was 13 points; in institutionalized mentally retarded boys the corresponding IQ disparity was 20 points. Because of the often poor performance of culturally disadvantaged children in school, and in view of the apparent importance of verbal as opposed to analytical functioning in determining the application of the "retarded" label with all its implications, it was felt that measures of verbal comprehension and analytic functioning should be obtained separately. The WPPSI Picture Completion subtest and the Preschool Embedded Figures Test were chosen to measure analytic functioning.

Task Description and Administration

The Picture Completion subtest of the Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 1967) is a downward extension of the identically named subtest of the Wechsler Intelligence Scale for Children. Twelve of the 23 items were taken from the WISC and 11 were new. The child is shown a series of 23 pictures, each of which has some important part missing, and is asked to indicate the missing part by either naming it or pointing to where it should be. If the child fails to give the correct answer on either of the first two pictures, he is shown and told the correct answer. No help is given after item two. The test is continued through all 23 pictures or until 5 consecutive failures beginning with card 3. Testing time is approximately five minutes.

The test was given only in Year 1 of the study. Its questionable validity as a measure of analytic functioning (which will be discussed below) led to its exclusion from the Year 2 test battery.

Scoring Procedures

Summed scores were obtained for correct verbal responses (i.e., the child's naming or describing the missing part of the object), correct non-verbal responses (the child's pointing to the missing part) and total correct (including some items scored as correct but where insufficient information had been recorded to determine whether the response had been verbal or non-verbal). Other scores obtained but not analyzed due to their infrequency of occurrence or lack of range were number of items repeated, number of items with elaborations, total task time and number of indeterminate (non-scorable) responses. Incompleteness in the recording of many children's responses and insufficient probing by testers made the coding task quite difficult. Simple labelling responses, for example, were often not questioned and thus had to be scored as indeterminate. In addition, some of the younger children apparently did not understand the word "missing" in the standard test question, "What's missing in this picture?," and Psychological Corporation would not grant permission to modify this formal instruction to the child. The difficulty in scoring suggests a need for verbatim recording of a child's response (which is not required in the WPPSI manual) and a modification of the word "missing" for use with very young subjects.

Score Properties

Score means, standard deviations, percentile distributions and coefficient alpha reliabilities for the total sample are presented in Table 1.

It was originally felt that the division of the total score into verbal correct and nonverbal correct subscores might reflect different degrees of competency and thus be useful for later analyses. Because of difficulties in determining categories for mixed answers (i.e., cases where the child both pointed and verbalized) and of many incompletely-recorded answer sheets, however, these two subscores were dropped from all further analyses.

The coefficient alpha reliability of the total score (.89) compares favorably with the corrected odd-even reliability coefficients reported in the WPPSI manual. These ranged from .81 to .86 for ages 4 through 6 1/2. In the present study, alphas were also obtained separately for those children 42-50 months of age and for those 51-59 months of age. The respective coefficients were .88 and .89. It should be noted, however, that given a performance-dependent cutoff, these alphas are inflated estimates of internal consistency.

Table 1

Means, Percentile Distributions and Reliabilities for Nonverbal, Verbal and Total Correct Scores* (N = 1405)

Score	Mean	SD	Percentiles 10	25	50	75	90	Reliability**
Nonverbal Correct	1.34	1.92	0.0	0.01	0.55	2.10	4.05	.72
Verbal Correct	3.11	3.47	0.0	0.24	1.98	5.09	8.36	.85
Total Correct	5.45	4.71	0.13	1.59	4.34	8.47	12.49	.89

*Range = 0-23.
**Cronbach's coefficient alpha.

Sample Performance

This test was more difficult for the present subjects as a group than it was for Wechsler's standardization sample. The WPPSI manual (Wechsler, 1967) lists a median raw score equivalent of 6 for his four-year-old group, whereas the present subjects (with a mean age of 4 years, 2 months) obtained a median total score of 4.3.

Distributions of the Total Score data by sex and three-month age groups are presented in Table 2. With the exception of the oldest subjects (the group 57-59 months old, which had a relatively small N and included a disproportionate number of low SES urban children), a pronounced age trend is apparent. Confirmation of this age trend is provided by an analysis of variance which showed significant age differences ($F = 72.22$, $df = 1/1057$, $p < .001$) for those children above and below the median age for the total correct score.

Table 2

Means, Standard Deviations and Percentile Distributions for Total Correct* by Age and Sex

				Percentiles					
Group	N	Mean	SD	10	25	50	WPPSI Median Raw Score** by Age	75	90
42-44 mo.	85	2.54	3.14	0.0	0.42	1.63	--	3.58	6.25
45-47 mo.	301	4.30	4.08	0.0	0.89	3.11	--	6.87	10.34
48-50 mo.	332	4.85	4.42	0.08	1.34	3.59	6	7.69	11.39
51-53 mo.	360	6.33	4.74	0.40	2.24	5.72	7	10.00	13.08
54-56 mo.	268	7.02	5.25	0.39	2.39	6.36	8-9	11.25	14.37
57-59 mo.	57	6.37	4.57	0.45	3.06	5.67	9-10	9.85	12.15
Boys	747	5.40	4.73	0.12	1.61	4.31		8.32	12.51
Girls	656	5.50	4.69	0.15	1.57	4.39		8.66	12.48
Total	1403	5.45	4.71	0.13	1.59	4.34		8.47	12.49

*Range = 0-23.

**Taken from WPPSI tables of scaled score equivalents of raw scores. These are approximate figures based on smoothed distributions. Ages for the WPPSI standardization groups are each one month younger than the figures to the left of the above table. That is, the median value listed for 48-50 months is from the conversion table for 47-49 months. N = 200 for each group.

As can be seen by comparing the median scores for the study sample with the median scores in the WPPSI norm-conversion tables, the medians are consistently lower for the present sample. Not shown in Table 2, but apparent from comparison of score distributions for the two subject pools, is the fact that there are considerably higher proportions of very low scores in the present study samples as compared with the WPPSI norm groups but, interestingly, not an appreciably smaller proportion of high scores.

In an Age x Sex x SES ANOVA a statistically significant sex effect was obtained ($F = 7.48$, $df = 1/1057$, $p < .01$), favoring boys. In absolute terms, however, the difference in means was negligible. Using the educational level of the mother as an index of socioeconomic status, the sample was divided into high (more than twelve years of schooling), middle (ten to twelve years of schooling) and low (less than ten years of schooling) SES groups. The F-value obtained from the analysis of variance was 75.08, $df = 2/1057$, $p < .001$, with those children whose mothers had received more schooling obtaining significantly higher scores.

Relationship with Other Measures

The major correlates of the total correct score on the Picture Completion test are given in Table 3. The three highest correlation coefficients (.56 to .59) are with the Preschool Inventory, the standard version of the Peabody Picture Vocabulary Test (A), and a modified Peabody (B) in which the child is asked to label the picture to which the tester points. All three of these tasks are predominantly verbal measures. The next group of correlates (correlations ranging from .40 to .45) involve perceptual or analytic material: Form Reproduction, understanding of negation on Matched Pictures, fast performance on the Seguin Form Board, few errors in Matching Familiar Figures, greater number of appropriate classifications on the Sigel Object Categorization Test, ability to discriminate among commonly confused speech sounds on the Children's Auditory Discrimination Inventory. The remainder the correlates listed fit no particular pattern but appear compatible

Table 3

Correlations of all Test Scores Loading Above .30 with Picture Completion Total Score*

Test and Score	Correlation
Preschool Inventory (Caldwell): Adjusted Total Score (minus Form Reproduction items 52-55)	.59
Peabody Picture Vocabulary Test, Form A: Total Correct	.56
Peabody Picture Vocabulary Test, Form B: Total Correct	.56
Form Reproduction: Total Score	.45
ETS Matched Pictures: Total Score - Negation	.43
Seguin Form Board: Fastest Time for Correct Placement (Log 10)	-.42
Children's Auditory Discrimination Inventory: Total Correct - Nonsense Words	.40
Sigel Object Categorization: Total Grouping Responses	.40
Matching Familiar Figures: Mean Errors per Valid Test Item	-.40
Hess and Shipman Eight-Block Sorting Task: Total Score	.38
Hess and Shipman Toy Sorting Task: Total Score	.36
Massad Mimicry: Word or some semblance of word given	.34
Motor Inhibition Test: Average Time, Trial 2, for the Walking and Drawing Subtests	.34
ETS Story Sequence Task: Test Items 1 and 2 (Receptive Language) Total Score	.33
ETS Matched Pictures: Total Score	.31
Johns Hopkins Perceptual Test: Total Correct	.30
Enumeration Task I: Total Correct (Items 1-12)	.30

*Sample sizes range from about 800 to more than 1200. All correlations are significant beyond the .001 level of confidence ($\underline{r}_{.001} = .148$ for $\underline{N} = 500$).

with either a general functioning or analytical functioning interpretation. Not listed in Table 3 because of its low magnitude is the correlation of Picture Completion with the other measure chosen to represent analytic functioning, the Preschool Embedded Figures Test. The correlation between total correct scores on the two measures was .28, statistically a highly significant correlation ($p < .001$), but one much too low for two reasonably reliable tests supposed to measure the same construct.

The manual for the WPPSI (Wechsler, 1967) gives the correlations among the scale subtests. At the 4- and 4 1/2-year levels Picture Completion correlated as highly or higher with verbal subtests (in particular, Information and Vocabulary) as with performance subtests. In fact, the contamination-corrected correlation of the Picture Completion subtest with the WPPSI performance score is not very different from that with the verbal score (r = .56 and .58, respectively, for age 4, and .50 and .59, respectively, for age 4 1/2).

Factor analysis of the present data showed the Picture Completion subtest to load .67 on Factor 1 (a general competency dimension), but to have loadings of .11 or less on the other factors obtained (Shipman, 1971). Coates and Bromberg (1972) factor analyzed the subtest correlation matrices from the WPPSI at each of the half-year age levels from 4 through 6 1/2, and found the Picture Completion subtest to load onto four differently-interpreted factors at the various age levels. At age 4 the test loaded on a verbal comprehension factor and a perceptual organization factor. At age 4 1/2, it loaded only on an uninterpreted fourth factor along with the Information, Vocabulary and Block Design subtests. Further evidence of the factorial complexity of this subtest and its change with age is shown in a recent factor analytic study of the WPPSI by Hollenbeck & Kaufman (1972), who used Wechsler's (1967) standardization sample of 1200 children, 400 each at ages 4, 5, and 6. At age 4, Picture Completion had high loadings on both the verbal and the performance factor, but at ages 5 and 6 there was a clear shift to the performance factor only. Thus, there was a verbal component at age 4 that diminished with the older subjects.

It seems clear from this pattern of correlations and of factor loadings that the WPPSI Picture Completion subtest at this age in this sample is measuring general competency, a culturally loaded general dimension. Evidence for its success as a distinct measure of analytic functioning at this age is much less clear, however. Its relatively low correlation with the Preschool Embedded Figures Test would seem to indicate that if Picture Completion is measuring analytical functioning, it is not doing so differentially in this sample at this age.

<u>Summary</u>

The WPPSI Picture Completion subtest was included in the test battery of the present study to measure analytic functioning. While it is reasonably easy to administer, great care must be taken by testers to question simple labelling responses and to record answers completely. Some children did not understand the meaning of the term "missing" in the instructions. The task was found to be sensitive to age, sex and SES differences in the present sample.

The test is factorially complex and reflects general competency, a mixture of intellectual competency and cultural experience. There is little evidence to support its use as a measure of analytic functioning in this group of subjects at this age.

References

Coates, S. & Bromberg, P. M. The factorial structure of the WPPSI between the ages of 4 and 6 1/2. Unpublished report. State University of New York, Downstate Medical Center, 1972.

Cohen, J. The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. Journal of Consulting Psychology, 1959, 23, 285-299.

Goodenough, D. & Karp, S. A. Field dependence and intellectual functioning. Journal of Abnormal and Social Psychology, 1961, 63, 241-246.

Hollenbeck, G. P. & Kaufman, A. Factor structure of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI). A.P.A. Proceedings, 1972, 37-38.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

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

Witkin, H. A. Personal Communication, 1971.

Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R., & Karp, S. A. Psychological differentiation. New York: Wiley, 1962.

Witkin, H. A., Faterson, H. F., Goodenough, D. R., & Birnbaum, J. Cognitive patterning in mildly retarded boys. Child Development, 1966, 37, 301-316.

Witkin, H. A., Lewis, H. B., Hertzman, M., Machover, K., Meissner, P. B., & Wapner, S. Personality through perception. New York: Harper, 1954.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp. S. A. A manual for the embedded figures tests. Palo Alto: Consulting Psychologists Press, 1971.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

Technical Report 17

Preschool Embedded Figures Test

Diran Dermen
Judith A. Meissner

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Preschool Embedded Figures Test (PEFT)[1]

Background

The Preschool Embedded Figures Test (Coates, 1972) is a measure of field independence, or analytic functioning. It has its origins in the work of Witkin and his associates (Witkin, Dyk, Faterson, Goodenough, and Karp, 1962; Witkin, Oltman, Raskin, and Karp, 1971). The test, based partly on the Children's Embedded Figures Test (Karp & Konstadt, 1963; Witkin et al., 1971)[2], extends the investigation of the development of psychological differentiation downward to age three.

A brief but fairly complete discussion of the rationale for embedded figures tests, including extensive references, is given in A Manual for the Embedded Figures Tests (Witkin et al., 1971).

The embedded figures tests of field dependence-independence are measures of perceptual aspects of analytic functioning or, at a more general level, psychological differentiation. An extensive program of research over a period of more than twenty years has presented a fairly detailed picture of the more- versus less-differentiated person. Based on performance on embedded figures measures, the more differentiated (field independent) person has been shown to do well on other perceptual tasks and on a variety of intellectual tasks involving disembedding. Embedded figures performance has been shown to be related to several noncognitive areas such as social

---

[1]We are indebted to Susan Coates for providing an early draft of the manual for the PEFT as well as several unpublished reports and additional data bearing on the question of the validity of the PEFT and other measures of analytic functioning.

[2]The Children's Embedded Figures Test was originally published in 1963 by Cognitive Tests. It is now published by Consulting Psychologists Press.

behavior, nature of the body concept, nature of psychological defenses, form (though not extent) of psychopathology, patterns of autonomic reactivity, and different family and cultural experiences (see Witkin et al., 1971).

A general pattern of personality correlates of psychological differentiation has emerged from a number of studies among elementary school children on the CEFT and of adults on the Embedded Figures Test (Coates, 1972). The more differentiated children and adults have been found to have an active initiating attitude, to be self-assured, to have a developed sense of separate identity, to be task-oriented and to perceive their parents as supportive. Less differentiated individuals have been found to be more socially oriented, more suggestible, dependent in their social relations, likely to rely on others for their self-definition, and to perceive their parents as nonsupportive.

Factor analytic studies (Cohen, 1957, 1959; Goodenough & Karp, 1961; Karp, 1963) have found separate verbal comprehension, attention-concentration and analytic functioning factors. These studies, using the Wechsler scales, found the analytic functioning factor to be comprised of the perceptual field dependence-independence measures (when included in the battery), by Block Design and Object Assembly on the WISC and WAIS, and by Picture Completion on the WISC. Several studies (Witkin, personal communication, 1971; Witkin, Faterson, Goodenough & Birnbaum, 1966) have reported that among individuals classified as mildly retarded a substantial number were found whose verbal comprehension factor scores were quite low but who had near-average prorated IQs on the analytic functioning factor. Apparently the "retarded" label was applied as a function of the verbal comprehension level of these children and adults, independently of their level of analytic functioning. In view of such findings, it would seem judicious to attempt to measure analytic functioning independent of verbal comprehension.

The brief rationale and survey of implications of embedded figures performance presented has been based almost exlusively on studies employing the Embedded Figures Test with adults (Witkin, 1950; Witkin et al., 1962; Witkin, Lewis, Hertzman, Machover, Meissner & Wapner, 1954; Witkin et al., 1971) and with children older than nine using the Children's Embedded Figures Test (Karp & Konstadt, 1963; Witkin et al., 1971). To meet the need for a measure of differentiation for younger children, the Preschool Embedded Figures Test (PEFT) was developed (Coates, 1972). The pattern of personality correlates obtained from the PEFT is quite consistent with that cited for adults and older children. That is, the more differentiated children were found to be more creative and inventive, avid to learn about new things, more goal-directed, less distrustful of people and less self-defeating than were the less differentiated children (Coates, 1972). Other consistent findings have applied to one sex or the other, but not to both, and the supporting evidence appears to be clearer for girls than for boys.

Coates (personal communication, 1971) has reported data indicating that the Geometric Designs and the Block Design subtests of the WPPSI help mark factors interpretable as analytic functioning and loaded by the PEFT. In the Coates manual (1972), data are presented showing sizable correlations between WPPSI Block Design and the PEFT (correlation coefficients ranging from .55 to .67 for five samples ranging in size from 21 to 28). There is, then, some evidence for the validity of the PEFT in its relation to other cognitive measures. Unfortunately, correlations between the earlier embedded figures measures and the PEFT cannot be obtained, since these other embedded figures measures cannot be given successfully to children as young as those for whom the PEFT is intended.

At the time of the selection of the PEFT for the present study there was only one other preschool embedded figures measure available, Banta's Early Childhood Embedded Figures Test (Banta, 1970), which provided a good testing procedure but items that were not as good as those in the Coates test since they were either too easy or too difficult. Recently, as part of the Sesame Street evaluation (Ball & Bogatz, 1970), a multiple choice "Hidden Triangles Test" was developed based partly on the Coates PEFT. Unfortunately the test proved to be unreliable (Cronbach alpha reliability coefficient of .28, $N = 1017$); apparently most of the items were too difficult.

Task Description[3]

The PEFT contains 27, 8 1/2 x 11 inch, black and white drawings (3 practice drawings and 24 test drawings), in each of which is embedded a simple equilateral triangle. The child is presented a card on which is printed a small triangle and shown how to trace his fingers along the edges of the triangle. He is then shown, one at a time, three practice figures in which the triangle is embedded, asked to indicate the simple figure, and, having done so, to run his fingers along its sides. The child is given the three practice items a maximum of three times. If he fails to meet the criterion of two correct items on either the second or the third practice trial he is not administered the regular test items. In the test proper, the child is allowed two attempts per item in a maximum of 30 seconds.

[3]It should be mentioned that the PEFT manual used in the present study differs somewhat from that currently used by Coates and was changed slightly for the second year of the study. Differences are relatively minor; the test items are the same.

Administration Procedures

The PEFT is not an easy test to administer, at least at the younger age levels. It was considered by tester-trainers and testers alike to be one of the more difficult tests in the battery to administer. In the first year of the study, an attempt was made to administer the test to a total of 1616 children. Of this total, 140 children did not get past the practice series, and an approximately equal number of additional children were eliminated from the sample due to incomplete or otherwise unscorable records. In the second year of the study, a total of 830 children were able to perform successfully, but 72 others were unable to complete the practice series or produced unscorable records. Inability to pass the practice items appears to have been a function of the child's failure to understand the instructions, to grasp the requirements of tracing the hidden figure with the finger, or to general lack of ease in the testing situation. A contributing factor in the difficulty of test administration was the timing requirement, which added a further element to an already complex situation.

Scoring Procedures

Subjects' responses and response times were recorded by the testers, and the completed answer sheets were coded at the ETS Princeton office. Scores obtained were total number correct, total testing time (time from beginning of practice to end of test), average time to first response (whether or not correct), and average total time for correct responses. The two scores used for the purposes of this report were total number correct and average time to first response, since the total testing time and the average total time did not give enough additional information to be included in the final data analyses. The principal difficulty in the scoring-coding process was presented by ambiguous recording of responses by the testers, especially for the practice trials.

Score Characteristics

Score means, standard deviations, percentile distributions and coefficient alpha reliabilities for the Total Score and the Latency Score for Years 1 and 2 of the study are presented in Tables 1 and 2, respectively.

Table 1

Means, Standard Deviations, Percentile Distributions and Reliabilities for the Total Score* and the Latency Score in Year 1

Score	N	Mean	SD	Percentile 10	25	50	75	90	Coefficient Alpha Reliability
Total Number of items correct	1288	12.1	5.55	3.9	8.6	12.8	16.1	19.5	.86
Av. time for first response (seconds)	1287	6.6	2.95	3.1	4.5	6.3	8.3	10.5	.77

Note.--At the time of this testing, the subjects ranged from 42 to 59 months of age.

*Range = 0-24.

Table 2

Means, Standard Deviations, Percentile Distributions and Reliabilities for the Total Score* and the Latency Score in Year 2

Score	N	Mean	SD	Percentile 10	25	50	75	90	Coefficient Alpha Reliability
Total Number of items correct	830	14.8	4.48	9.1	11.9	15.4	18.2	20.8	.81
Av. time for first response (seconds)	831	5.1	2.19	2.4	3.5	4.9	6.4	8.1	.51

Note.--At the time of the Year 2 testing, the subjects ranged from 51 to 69 months of age.

*Range = 0-24.

The conventional score for the test is the total number correct. It can be seen from inspection of the two tables that subjects as a group correctly completed an average of 2.7 more items in Year 2 than they had in Year 1. The reliability coefficients of .86 in Year 1 and .81 in Year 2 compare favorably with reliabilities reported by Coates (1972). In Coates' standardization samples, the corrected odd-even reliability coefficients for the number correct score ranged from .74 to .91 (sample sizes ranging from 26 to 70). Data on stability is available in both the present study and in Coates' (1972) manual. In the present study the Year 1 - Year 2 correlation for the total correct score was .39 (N = 750) over an interval of about nine months. Coates reports test-retest correlations over a five-month interval for three small samples (N = 21-25) of .69 to .75. The Year 1 - Year 2 correlation for the latency score was only .17, suggesting that, of the two scores used on this test, the total correct score is considerably more stable. Coefficient alpha for Year 1 was .77; for Year 2 it was .51.

Table 3 presents the per cent of subjects passing each test item and the biserial correlations of each item with the total score for Years 1 and 2. The per cent passing ranged from 14.6 to 81.3% in Year 1 and from 15.6 to 95.1% in Year 2. The biserial correlations with total score ranged in Year 1 from .41 to .79 with only one falling below .50, and in Year 2 from .33 to .76 with again only one below .50. Thus, the test appears to have reasonably good psychometric properties. In exactly half of the 24 items there was a greater than 10% increase in subjects passing that item in Year 2, and in several of these items (such as items number 4 and 13, for example) there was a greater than 25% gain in subjects passing in the second year. It is interesting to note that the items in which there was

Table 3

Item Biserials -- Number Correct
Year 1 and Year 2 Data
(N = 1288 for Year 1 and 830 for Year 2)

Item Number*	Per Cent Passing		Biserial Correlation with Total Score	
	Year 1	Year 2	Year 1	Year 2
4	64.6	90.4	.56	.53
5	41.0	37.2	.56	.63
6	22.0	23.7	.52	.33
7	69.3	91.8	.53	.54
8	30.2	31.3	.55	.62
9	32.0	32.7	.55	.52
10	81.3	95.1	.64	.73
11	47.6	42.2	.60	.52
12	34.7	30.9	.54	.35
13	48.5	77.3	.74	.67
14	53.6	75.7	.71	.60
15	62.4	85.7	.66	.72
16	48.3	68.4	.79	.76
17	64.8	81.5	.69	.72
18	72.3	80.5	.68	.55
19	46.3	54.8	.71	.71
20	34.9	38.6	.41	.69
21	61.2	83.4	.70	.70
22	48.4	51.3	.61	.68
23	14.6	15.6	.52	.51
24	75.0	91.8	.59	.57
25	50.8	51.5	.64	.64
26	55.4	70.4	.69	.54
27	66.7	80.6	.68	.64

*Items 1, 2 and 3 were practice items only.

only a very slight increase in the subjects passing in Year 2 tended to be those items which originally showed the lowest number of subjects passing (items 6, 8, 9, 12, 20 and 23). It appears that the difficult items (40% or less passing) were no less difficult after 9 months of development. The majority of the biserial correlations did not vary greatly in magnitude between Year 1 and Year 2 (the only exceptions being items 6, 14, 18, 20 and 26).

Sample Performance

(a) Total Correct Score

Tables 4 and 5 present the means, standard deviations and percentile distributions for the total correct score for all subjects by three-month age intervals and sex for Years 1 and 2, respectively. Figures 1 and 2, a graphic form of the age data of Tables 4 and 5, clearly show the developmental trend present in these data.

For purposes of analysis, only data for longitudinal subjects were used (i.e., the subjects for whom test scores were available in both Years 1 and 2). The age data had to be subjected to separate analyses of variance in Years 1 and 2, since the testing schedule in Year 2 led to considerable shifting of subjects between the "younger" and "older" categories. Therefore, no Year 1 - Year 2 comparisons of the age data can be made. Since there was little or no shifting of subjects between SES or sex subgroups, these data could be subjected to repeated-measures analyses of variance performed on their combined Year 1 - Year 2 data, which yielded both overall and Year 1 vs. Year 2 F-values.

The age data were divided at the median in order to obtain an "older" and a "younger" age group. The difference between the means of the total

Table 4

Number Correct

Means, Standard Deviations and Percentile Distributions by Age and Sex Groups for Year 1 ($\underline{N}$ = 1288)

Group	N	Mean	SD	Percentiles 10	25	50	75	90
42-44 mo.	76	10.2	5.83	2.8	6.6	9.8	15.0	18.2
45-47 mo.	258	11.6	6.01	3.6	6.9	11.8	16.0	20.2
48-50 mo.	300	11.5	5.49	3.5	7.5	12.2	15.6	18.6
51-53 mo.	342	12.5	5.41	4.1	9.3	13.1	16.3	19.4
54-56 mo.	255	13.2	4.99	6.2	9.7	14.0	16.9	19.5
57-59 mo.	57	13.8	5.19	6.4	10.4	15.1	17.4	20.5
Boys	669	11.8	5.66	3.7	7.7	12.4	16.0	19.2
Girls	619	12.5	5.40	4.2	9.2	13.1	16.3	19.7
Total	1288	12.1	5.55	3.9	8.6	12.8	16.1	19.5

Figure 1

Mean Score as a Function of Age for Year 1

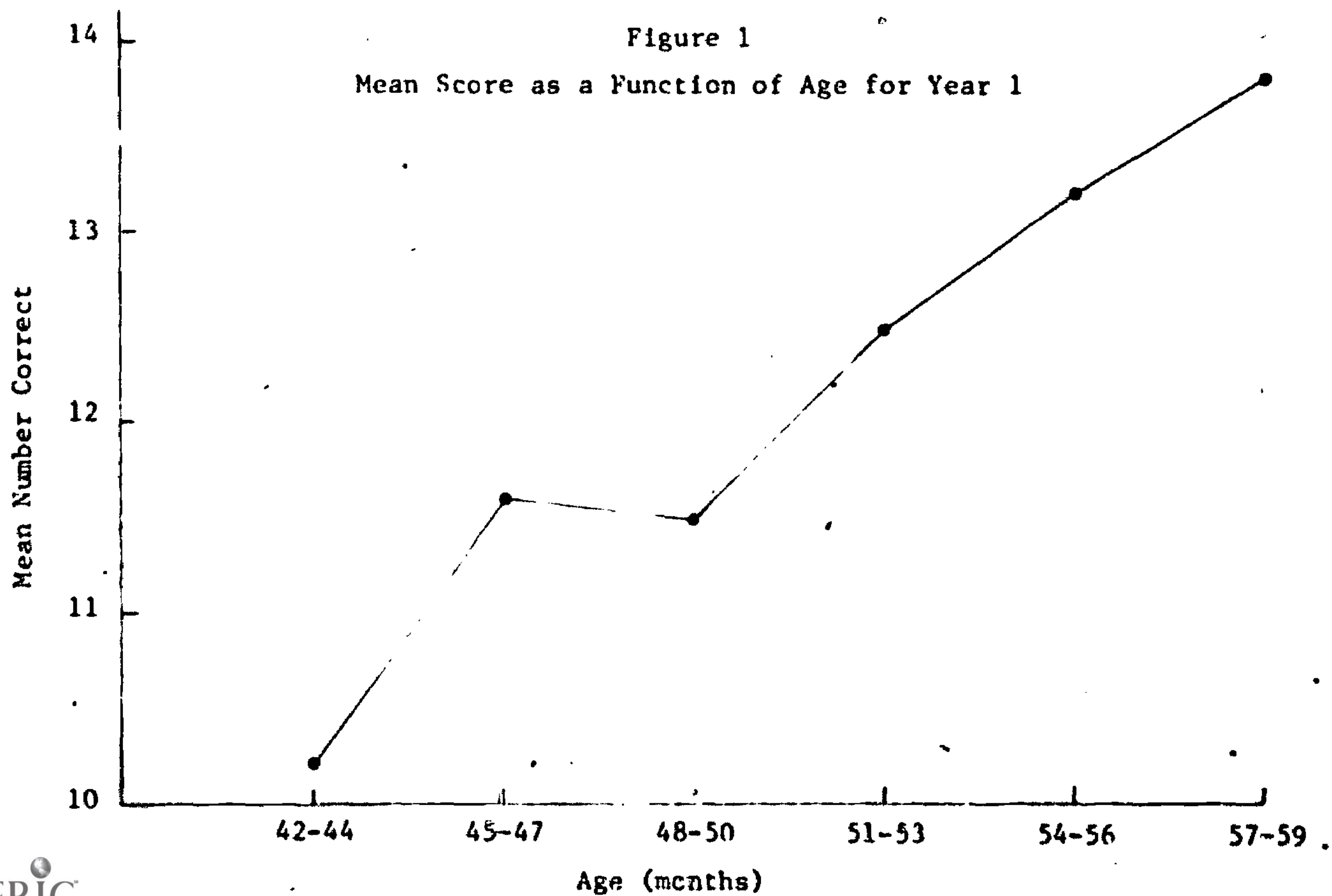


Table 5

Number Correct

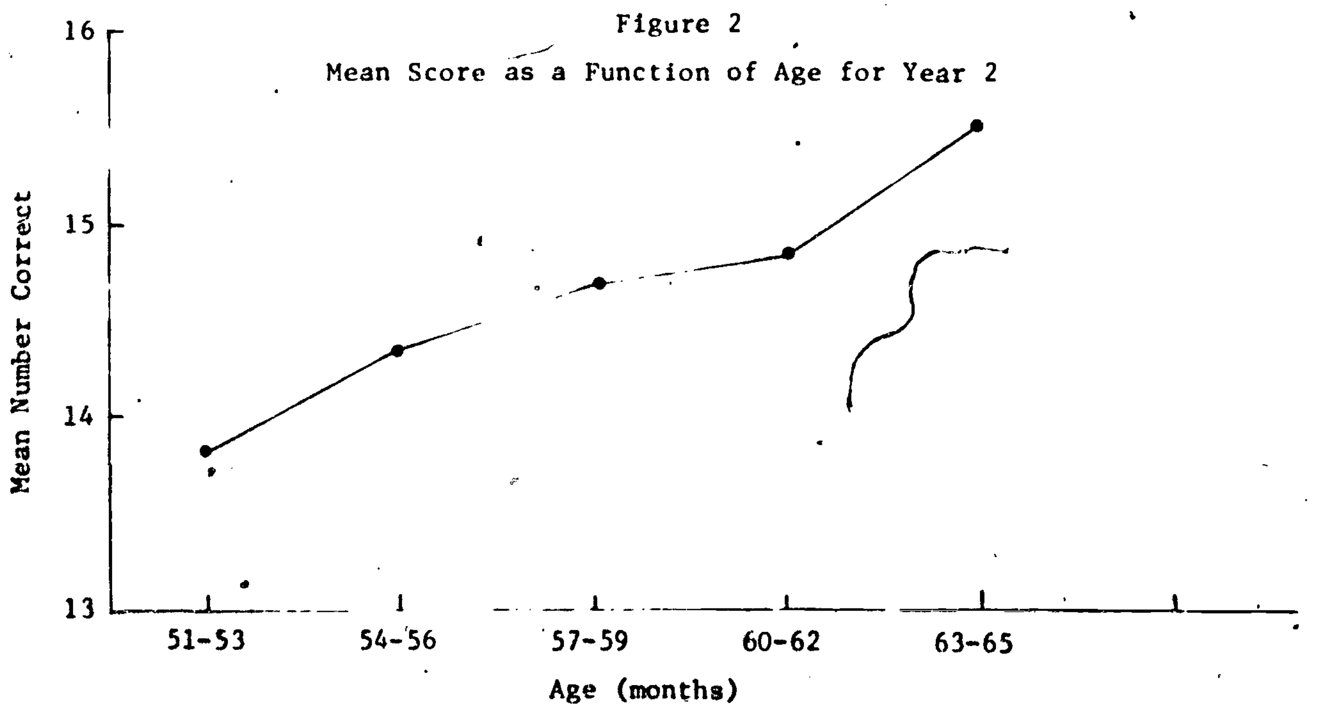
Means, Standard Deviations and Percentile Distributions by Age and Sex for Year 2 (N = 830)*

Group	N	Mean	SD	Percentiles 10	25	50	75	90
51-53 mo.	55	13.8	5.00	7.6	10.0	13.7	17.7	20.6
54-56 mo.	196	14.4	[illegible].57	8.5	11.0	15.2	18.2	20.3
57-59 mo.	178	14.7	4.53	9.0	11.6	15.3	18.0	20.6
60-62 mo.	240	14.8	4.54	9.2	12.2	15.4	18.2	21.2
63-65 mo.	152	15.5	3.82	10.2	13.1	15.7	18.6	20.9
66-69 mo.	9	14.4	5.94	1.7	12.9	14.7	18.3	21.0
Boys	426	14.2	4.67	8.3	11.1	15.0	17.8	20.5
Girls	404	15.3	4.21	9.7	12.8	15.7	18.6	21.0
Total	830	14.8	4.48	9.1	11.9	15.4	18.2	20.8

*In Year 2 of the Longitudinal Study, children from only three of the four study sites (Trenton, Portland, St. Louis) were given the PEFT. Hence, the drop in subjects from 1288 in Year 1 to 830 in Year 2.

Figure 2

Mean Score as a Function of Age for Year 2



correct score for these two groups (13.24 for the older subjects and 11.76 for the younger ones) was highly significant in Year 1 ($F = 18.60$, df = 1/980, $p < .001$) and also in Year 2 (15.82 for the older subjects and 14.75 for the younger ones; $F = 8.69$, df = 1/668, $p < .005$). Coates (1972) found significant age differences in performance with her three-, four-, and five-year-old subjects.

The repeated-measures analysis of variance summing scores across years found a significant overall difference between means for girls and boys on total number of items correct ($F = 8.37$, df = 1/599, $p < .005$). The difference in favor of girls at this age level was consistent with the results in the normative data on three-, four-, and five-year-olds reported by Coates (1972), although she also cites a study by Seitz (1971) in which significant differences were not found. In their normative data on the CEFT, Karp and Konstadt (1963) failed to find significant sex differences in performance in the age range 5 to 12. These apparently contradictory findings on sex differences on PEFT and CEFT total score might indicate that the construct validity of the task changes with age, as is suggested by the present findings.

As a rough measure of socioeconomic status, those children for whom longitudinal data were available were divided into three groups on the basis of their mother's educational level. Those subjects in the "high" group had mothers with more than 12 years of schooling, those in the "middle" group had mothers with 10 to 12 years of schooling, and the mothers of the "low" group had fewer than 10 years of schooling. The across-years analysis of variance was highly significant ($F = 11.61$, df = 2/599, $p < .001$), with the high SES group obtaining the highest mean scores and the low SES group the lowest. This finding is consistent with Beller's results as reported by Coates (1972). It should be noted that SES-related differences in performance have also been found on the CEFT (Witkin et al., 1971).

(b) Latency Score

Tables 6 and 7 present descriptive data on average time for first response (latency) scores by three-month age groups and by sex for all subjects for Years 1 and 2, respectively. These tables show that the latency values are somewhat greater in Year 1 than in Year 2.

The division of the longitudinal sample at the median into older and younger subjects revealed no significant mean differences in either Years 1 or 2. The repeated-measures analysis of variance showed sex differences in latency to be highly significant across years (F = 10.88, df = 1/599, $p < .001$), with the higher latencies obtained by the boys, but the difference in means was quite small in absolute terms. The latency score showed a marginally significant SES x Year interaction (F = 4.33, df = 2/599, $p < .05$), with

Table 6

Average Time for First Response (in seconds)

Means, Standard Deviations and Percentile Distributions
by Age and Sex for Year 1

				Percentiles				
Group	N	Mean	SD	10	25	50	75	90
42-44 mo.	76	6.9	3.54	3.4	4.7	6.4	8.2	10.8
45-47 mo.	258	6.8	3.10	3.2	4.5	6.2	8.6	11.0
48-50 mo.	299	6.7	2.98	3.1	4.4	6.2	8.2	11.0
51-53 mo.	342	6.5	2.93	3.0	4.4	6.2	8.5	10.3
54-56 mo.	255	6.4	2.67	3.0	4.5	6.3	8.0	9.6
57-59 mo.	57	6.5	2.67	3.2	4.4	6.2	8.2	9.9
Boys	669	6.8	3.08	3.2	4.6	6.4	8.7	10.8
Girls	618	6.4	2.80	2.9	4.3	6.1	8.0	10.0
Total	1287	6.6	2.95	3.1	4.5	6.3	8.3	10.5

Table 7

Average Time for First Response (in seconds)

Means, Standard Deviations and Percentile Distributions
by Age and Sex for Year 2

				Percentiles				
Group	N	Mean	SD	10	25	50	75	90
51-53 mo.	56	5.3	2.42	2.1	3.8	5.2	6.8	8.6
54-56 mo.	196	5.2	2.15	2.5	3.7	4.9	6.6	7.9
57-59 mo.	178	4.8	2.13	2.3	3.3	4.7	6.1	7.6
60-62 mo.	240	5.3	2.29	2.5	3.8	5.0	6.6	8.7
63-65 mo.	152	4.9	2.08	2.4	3.2	4.7	6.3	7.9
66-69 mo.	9	5.0	1.61	2.9	3.4	5.4	6.3	6.9
Boys	427	5.4	2.30	2.6	3.8	5.1	6.7	8.6
Girls	404	4.8	2.02	2.4	3.3	4.7	6.2	7.6
Total	831	5.1	2.19	2.4	3.5	4.9	6.4	8.1

the high SES group showing greater change within the two-year period than the middle or low SES subjects. In Year 1, high SES subjects took somewhat longer to give their first response than did the other two groups, but in Year 2 the reverse was true. This finding suggests a possible developmental sequence in the approach of young children to a task. The first stage might consist of a quick, impulsive response given by many children upon the presentation of a new task. With increasing experience and maturation, there may be more attention to complexities, and many subjects may react in a slower, more reflective manner. With increasing competency, reaction times would again become faster. If this series of stages is indeed the case, the high SES subjects, many of whom are likely to have had a wider variety of task-related experiences than have middle and lower SES children, may initially approach a task in a more reflective manner, as is suggested by the above.

ANOVA. By the second testing, middle and lower SES subjects may be just entering the reflective stage while high SES subjects, having become increasingly competent in the problem-solving skills involved, proceed without hesitation. Future analyses will indicate whether the lower SES groups in subsequent testing show the same pattern over time.

Validity

There are two sources of evidence bearing on the question of the validity of the PEFT, the first being the data from the present study, the other based on data from other studies. In the present study the evidence for PEFT validity is to be obtained from the correlations with other measures included and also from the factor loadings. Correlations of the score for number correct on the PEFT with these other measures ranged from zero to a high of .34 in Year 1 and .40 in Year 2. The highest correlations* in both Years 1 and 2 were with the Preschool Inventory (r = .32 and .36, respectively), Peabody Picture Vocabulary Test, Form A (.31 and .29), Seguin Form Board (negative correlations of -.34 and -.38, respectively, with fastest log placement time), Form Reproduction (.32 and .40), and the ETS Enumeration Test (.33 with total Enumeration score in Year 1 and .32 with the Counting items in Year 2). Of these measures, the Peabody and the Preschool Inventory are predominately verbal in nature, the others more perceptual or analytic. The highest factor loading of the PEFT total score was with a general intellectual functioning factor (Factor 1) in both Years 1 and 2 (Varimax loadings of .49 in both Years 1 and 2). All other factor loadings were considerably below .30 in both years (see Shipman, 1971, 1972). Thus not surprisingly, the factor analytic data are consistent with the correlational data.

In addition to the PEFT, the Picture Completion subtest of the WPPSI

*These correlations were based on data from the longitudinal subjects only.

was included in the study in Year 1 as a measure of analytic functioning. The correlation of Picture Completion with the PEFT was .29; a statistically significant but low correlation. The question can legitimately be raised as to whether the measures included in the study were such as to provide confirmative evidence for the validity of the PEFT. As mentioned in the introductory section of this report, in factor analytic studies WISC Picture Completion had helped define a factor identified as reflecting analytic functioning. Unfortunately, at the time the battery was compiled, there was no evidence bearing on the factorial structure of the WPPSI subtests. In a very recent unpublished study of the factorial structure of the WPPSI at each of the ages at which it was standardized (Coates and Bromberg, 1972), the factor loadings of the Picture Completion subtest were variable, loading on four differently interpreted factors at different age levels. Thus, WPPSI Picture Completion was a dubious choice as a marker of analytic functioning. Evidence that the PEFT is measuring something different from that which is being measured by the rest of the Longitudinal Study test battery is available in the observation that in several factor analyses of the battery, the PEFT had low communalities, i.e., had little variance in common with other variables in the analyses. The correlation between Year 1 and Year 2 total correct score was .39 and was only .17 for the Year 1 and Year 2 latency scores. Such low coefficients of stability are quite often found in testing young children and could be an indication of age changes due to growth of the ability in question. Perhaps the safest conclusion that can be reached from all these data, however, is that the question of the validity and reliability of the PEFT is still unanswered.

The above discussion has been limited primarily to the Total Correct score. The other score used in the present analysis was average time to first response (latency). There was a correlation of -.06 between the Total Correct and Latency score in Year 1 and -.19 in Year 2. Highest correlations of the Latency measure were with other response speed measures, specifically, response latency on the Sigel Sorting Task in Years 1 and 2 (.23 and .20, respectively), Matching Familiar Figures in Year 1 (r = .23) and Seguin log fastest time in Year 2 (r = .17). While the Latency score appeared to add very little to the interpretation of the test, it did contribute to a "latency" factor in both years (see Shipman, 1971, 1972) and was therefore of value within the context of the present study.

Summary

Although the PEFT was found to be appropriate and interesting for the older children tested in the Longitudinal Study, the testers found it to be quite difficult to hold the attention of many of the younger children (those approximately four years old and younger) throughout the administration of this test. The PEFT must be administered by carefully-trained, but not necessarily well-educated, adults. Particular care is required in orienting the child to the task during the practice trials.

The total score measure was found to be sensitive to age in both the Year 1 and Year 2 analyses of variance and to both sex and SES differences in a Year 1-Year 2 repeated-measures analysis of variance. The evidence from the latency measure was not very informative, and it is recommended that the timing of responses be dropped except where there is a particular need for a latency measure.

There was little direct evidence, either positive or negative, regarding the construct validity of the PEFT in the present study. Supportive, although not conclusive, evidence for the validity of the test is to be found elsewhere

(Coates, 1972). Evidence was apparent in the present study, however, for the changing meaning of the task from Year 1 to Year 2, the Year 1 correlations being highest with the more verbal measures of the Peabody and Preschool Inventory and the Year 2 correlations being highest with the more perceptual tasks of Seguin Form Board and Form Reproduction. Hopefully, data in subsequent years will provide evidence for the validity of the PEFT as a measure of analytic functioning as sufficient competence in general test-taking skills, verbal comprehension and form discrimination is attained and, perhaps, accounts for less of the PEFT performance.

References

Banta, T. J. Tests for the evaluation of early childhood education: The Cincinnati Autonomy Test Battery (CATB). In J. Hellmuth (Ed.) Cognitive Studies, Vol. 1, New York: Brunner/Mazel, 1970.

Ball, S. & Bogatz, G. A. The first year of Sesame Street: An evaluation. Educational Testing Service Project Report, PR-70-15, Princeton, N. J: Educational Testing Service, 1970.

Coates, S. Personal Communication, 1971.

Coates, S. Preschool Embedded Figures Test. Palo Alto: Consulting Psychologists Press, 1972.

Coates, S. & Bromberg, P. M. The factorial structure of the WPPSI between the ages of 4 and 6 1/2. Unpublished report, S.U.N.Y., Downstate Medical Center, 1972.

Cohen, J. The factorial structure of the WAIS between early adulthood and old age. Journal of Consulting Psychology, 1957, 21, 283-290.

Cohen, J. The factorial structure of the WISC at ages 7-6, 10-6 and 13-6. Journal of Consulting Psychology, 1959, 23, 285-299.

Goodenough, D. & Karp, S.A. Field dependence and intellectual functioning. Journal of Abnormal and Social Psychology, 1961, 63, 241-246.

Karp, S. A. Field dependence and overcoming embeddedness. Journal of Consulting Psychology, 1963, 27, 294-302.

Karp, S. A. & Konstadt, N. L. Children's Embedded Figures Test. New York: Cognitive Tests, 1963.

Seitz, E. The relationship between cognitive abilities and impulse control in Project Head Start children. Unpublished doctoral dissertation, New York University, 1971.

Shipman, V. C. 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, PR-71-19, December, 1971.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, PR-72-18, September, 1972.

Witkin, H. A. Individual differences in ease of perception of embedded figures. Journal of Personality, 1950, 19, 1-15.

Witkin, H. A., Dyk, R., Faterson, H. F., Goodenough, D. R. & Karp, S. A. Psychological differentiation. New York: Wiley, 1962.

Witkin, H. A., Faterson, H. F., Goodenough, D. R. & Birnbaum, J. Cognitive patterning in mildly retarded boys. Child Development, 1966, 37, 301-316.

Witkin, H. A., Lewis, H. B., Hertzman, M., Machover, K., Meissner, P. & Wapner, S. Personality through perception. New York: Harper, 1954.

Witkin, H. A., Oltman, P. K., Raskin, E. & Karp, S. A. A manual for the embedded figures tests. Palo Alto: Consulting Psychologists Press, 1971.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

Technical Report 18

Preschool Inventory

Lynn E. Gilbert
Virginia C. Shipman

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Preschool Inventory

Background

The single task in the Longitudinal Study battery most clearly associated with general cognitive development, is the Preschool Inventory, which taps a range of verbal, quantitative, and perceptual-motor skills. Developed by Bettye Caldwell during the early years of the Head Start program, it was used initially in the summer program of 1965 and has been widely administered in Project Head Start ever since. As stated in the 1970 Handbook:

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

The original form of the Preschool Inventory consisted of 161 items, and a shortened version of 85 items was made available in 1967. In 1969, a Revised 64-item Inventory was distributed, which was the version used in the present study. More recently, the Stanford Research Institute created a 32-item and then a 29-item version for the Head Start Planned Variation Study.

Statistical information on the standardization sample for the 1970 Revised Edition (64-items) is contained in the Handbook (Cooperative Tests and Services, 1970). Summary data are provided in Table 1.

Ernhart, Jordan and Spaner (1972) reported a 1970 study by Miller and Dyer in which a six-month test-retest stability coefficient of .79 was obtained. The Huron Institute (Walker, Bane & Bryk, 1973) examined data from the Head Start Planned Variation evaluation study and reported a KR-20 reliability of .92 for 1674 children in the 1969 sample, and .92

for 2591 children in the 1970 sample. The score distributions obtained from the 1970 sample are presented in Table 2.

Table 1

Distribution of Preschool Inventory Scores*
from the 1970 Stand rdization Sample

Age Group	N	Mean	SD	KR-20
3-0 to 3-11	158	25.6	9.8	.88
4-0 to 4-5	528	30.0	10.1	.88
4-6 to 4-11	438	33.9	10.5	.86
5-0 to 5-5	259	38.4	10.1	.89
5-6 to 6-5	148	42.4	11.0	.92

*Range = 0-64.

Table 2

Distribution of Preschool Inventory Scores*
from the Fall 1970 Head Start Planned Variation Sample

Age Group	N	Mean	SD
42-44 mo.	8	20.5	11.53
45-47 mo.	63	24.7	10.79
48-50 mo.	204	27.1	9.96
51-53 mo.	316	30.2	10.91
54-56 mo.	341	33.2	11.60
57-59 mo.	348	35.6	11.05
60-62 mo.	270	38.9	11.47
63-65 mo.	228	40.0	11.12
66-68 mo.	180	41.0	10.90
69-71 mo.	176	43.2	10.92
Total	2134	35.2	12.22

*Range = 0-64.

The Research Triangle Institute (1972) reported results on the Inventory with the 1968-69 Head Start national evaluation sample, using a subsample of 1162 children ranging in age from 2-7 to 6-0. Mean scores were slightly higher than those reported in the Handbook. Age and SES were found to have a significant relationship to the Preschool Inventory score ($p < .05$), while sex was not a significant variable. No significant relationships were found between age, sex, or SES and the amount of raw score gain between pre- and posttest scores over an approximate 6-month period.

The relationship of family variables to the Preschool Inventory was also examined by the Research Triangle Institute. Mother's employment status, level of aspiration, and frequency of reading to her child all had significant effects ($p < .01$) on the pretest score. Ernhart et al. (1972) found mean total score to increase with level of maternal education in a sample of 188 white children of varying socioeconomic status. In that study, correlations with SES and with maternal education were .44 and .47, respectively. For 97 black children, the correlations were .26 and .18.

Because the Inventory was developed as a general achievement measure, particularly for use with Head Start children, it has been related by various investigators to such classroom variables as amount of teacher control, level of parent participation, and type of curriculum (Adkins, 1970; Curtis & Klock, 1970; Hervey & Story, 1967; Research Triangle Institute, 1972; Systems Development Corporation, 1972). The Research Triangle Institute found a mean gain score of 9.4 from pre- to posttest for the 1968-69 national evaluation sample, a gain greater than expected for the approximate six months difference in age means.

Task Description

The 1970 Revised Edition of the Preschool Inventory contains 64 items classified in the Inventory Manual under four major headings: Personal-Social Responsiveness (18 items, e.g., "How old are you?", "Raise your hand."); Associative Vocabulary (12 items, e.g., "What does a dentist do?"); Concept Activation--Numerical (15 items, e.g., "How many wheels does a car have?"); Concept Activation--Sensory (19 items, e.g., "Which is heavier, a brick or a shoe?"). The majority (about 60%) of the items require an oral response from the child, while the rest require him to follow directions such as "Wiggle," "Point to the middle checker," "Color the circle yellow."

This is not a difficult test to administer, although training is required in the manipulation of test materials and the appropriate use of probes. The form reproduction items presented a problem in that children were able to see through the paper and trace the figures; to prevent tracing, E inserted a sheet of dark paper underneath these items.

Testing time is approximately 20 minutes with three- to five-year-old children.

Scoring

Testers recorded children's answers verbatim for the verbal items and described their performance on nonverbal ones. ETS staff at Princeton scored each item as follows: correct, incorrect, child said "don't know," child refused to answer, or indeterminate.

Since the four copying items were also part of the Form Reproduction test administered in the study, an adjusted total score based on the 60 remaining items was the measure used in structural analyses. The Inventory Handbook (1970) advises against the determination of subset scores, and factor analyses of the present data did not support their use.

Score Properties

The alpha coefficient of reliability for 1467 subjects in Year 1 was .92. In Year 2, with N = 1311, alpha was .93. This compared favorably with reliabilities for the 1970 standardization sample, which ranged from .88 to .92.

The correlation between Year 1 and Year 2 scores was .66, one of the highest stability coefficients obtained for this sample with the Longitudinal Study Year 1-Year 2 test battery.

Sample Performance

For a total of 1474 children in the four sites combined in Year 1, the mean total correct (for the full 64 items) was 27.9 and the standard deviation was 11.91. The 1311 children in Year 2 obtained a mean score of 38.1, with a standard deviation of 12.32. Tables 3 and 4 present the distribution data for age and sex subgroups for Years 1 and 2, respectively. Longitudinal Study data are similar to results for the 1970 standardization sample and for the 1970 Head Start Planned Variation sample reported earlier.

Age x sex x SES analyses of variance were performed on the longitudinal sample, those children tested both in Year 1 and Year 2. The index of SES was mother's education: below 10th grade; between grades 10 to 12; and above 12th grade. In a repeated-measures ANOVA, both sex and SES were significant beyond the .001 level ($F = 20.06$, $df = 1/1080$ for sex and $F = 143.81$, $df = 2/1080$ for SES) when the data were combined across years. Girls obtained significantly higher scores, as did those children whose mothers had completed more years of schooling. SES was also significant ($F = 8.09$, $df = 2/1080$, $p < .001$) when the within-group differences from

Table 3

Distributions of Total Score* by Age and Sex, Year 1

Group	N	Mean	SD	Percentiles 10	25	50	75	90
42-44 mo.	89	22.3	11.44	8.2	13.4	20.7	28.9	38.0
45-47 mo.	317	25.0	10.85	11.2	17.3	23.0	31.4	40.1
48-50 mo.	348	26.4	11.45	11.7	17.7	24.5	33.3	42.4
51-53 mo.	392	29.0	11.55	13.6	20.4	28.0	35.8	44.1
54-56 mo.	270	32.1	12.04	17.0	22.3	31.2	40.8	47.2
57-59 mo.	58	35.3	12.63	17.8	25.6	34.2	44.5	50.6
Boys	780	26.8	11.89	11.7	18.0	25.0	34.2	43.2
Girls	694	29.1	11.83	13.6	20.1	28.2	36.8	45.3
Total	1474	27.9	11.91	12.4	19.0	26.6	35.6	44.4

*Range = 0-64.

Table 4

Distributions of Total Score* by Age and Sex, Year 2

Group	N	Mean	SD	Percentiles 10	25	50	75	90
51-53 mo.	82	35.2	11.16	20.9	26.7	35.7	44.2	49.5
54-56 mo.	309	35.8	12.34	19.6	26.8	37.0	45.2	52.5
57-59 mo.	306	37.3	11.96	21.1	27.6	38.6	46.3	53.0
60-62 mo.	351	39.0	11.72	23.2	30.2	40.6	48.2	54.0
63-65 mo.	247	41.6	13.15	23.0	33.5	43.0	53.2	57.5
66-69 mo.	16	42.1	10.44	29.1	36.8	40.0	49.6	57.3
Boys	697	36.5	12.64	19.5	26.9	37.6	46.3	53.2
Girls	614	40.0	11.69	24.2	31.7	41.1	49.0	55.5
Total	1311	38.1	12.32	21.4	28.8	39.5	47.8	54.5

*Range = 0-64.

Year 1 to Year 2 were examined. The middle SES group made the largest gain, followed by the high and low groups, respectively. Further examination of the inter-relationship between SES and preschool experience may help clarify this finding.

To determine age effects, an ANOVA was performed for each year separately, using a median split of younger and older children. Both in Year 1 and Year 2 the older children scored significantly higher (F = 68.86, df = 1/1107, in Year 1 and F = 19.86, df = 1/1127, in Year 2; $ps < .001$). Furthermore, the repeated-measures ANOVA showed a highly significant increase in total score from Year 1 to Year 2 (F = 1566.19, df = 1/1084, $p < .001$).

Relationships with Other Measures

The data support consideration of the Preschool Inventory as a general achievement measure. Correlations* with other cognitive-perceptual tasks were among the highest obtained. Table 5 presents data for tasks whose correlation in either year was above .40.

Factor analyses of the child test data for Year 1 and Year 2 yielded a first component best defined as general information-processing skills or "g." The Preschool Inventory had the highest loading of any measure on this factor in both years (see Shipman, 1971, 1972).

Summary

Both the internal analyses and the correlational data suggest that the Preschool Inventory is a reliable and valid measure of cognitive abilities among preschool children. Its reliabilities and correlations with other tasks were generally the highest in the Longitudinal Study battery.

The large difference among SES groups indicates the importance of

---

*These correlations are based on data from the longitudinal subjects only.

Table 5

Correlations of the Preschool Inventory with Selected Measures for the Longitudinal Sample

Task	Year 1 (N = 399-760)	Year 2 (N = 491-800)
Peabody A: receptive language score	.58	.66
TAMA: Total Correct	.53	.70
Children's Auditory Discrimination Inventory: Total Correct	.51*	not available
Form Reproduction Score	.51	.54
Hess and Shipman Toy Sorting Task: Total Score	.50*	not given
Peabody B: productive language score	.50	.63
Hess and Shipman Eight-Block Sorting Task: Total Score	.47	.53
Seguin Form Board: Log Fastest Time to Correct Placement	-.44	-.52
Matching Familiar Figures: Errors	-.42	-.54
Children's Auditory Discrimination Inventory: nonsense words score	.41	.52
Sigel Object Categorization Test: Total grouping responses	.39	.50
ETS Story Sequence I: Total Score	.39	.43
ETS Matched Pictures: Total Score	.35	.59
Johns Hopkins Perceptual Test: Total Score	.32	.48
Massad Mimicry: Real Words, final sounds	.31	.53
ETS Enumeration I: Total Correct	.30	
ETS Enumeration II: Items 1-4 (Counting)		.58

Note.--r.01 = .148 for N = 300.

*Correlation is based on total Year 1 sample since a longitudinal sample was not available.

experience as an influence on general knowledge and skills, and substantiates the claim of the test developers that it is "by no means culture free." Again not surprisingly, older children, with the opportunity for further experience and development, scored significantly higher than younger children.

Significantly higher performance of girls on the Preschool Inventory may reflect differential instruction in the home or greater cooperation and interest in the task. As observed in the Year 1 mother-child interaction situations, girls appeared more attentive to the mother; and preliminary findings from the parent interview revealed a small but consistent trend for mothers of girls to be more involved in school-relevant activities (e.g., reading to the child). These results may also reflect differential verbal interaction with the mother (both in amount and elaboration), as has been suggested in previous research (Goldberg, Godfrey & Lewis, 1967; Halverson & Waldrop, 1970; Hess, Shipman, Brophy & Bear, 1968; Moss, 1967).

References

Adkins, D. C. Hawaii Head Start evaluation 1968-1969. Hawaii: Educational Research and Development Center, University of Hawaii, January 1970.

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

Curtis, H. A. & Klock, J. A. Florida pre-kindergarten migrant compensatory program: An evaluation. Tallahassee, Fla.: Florida State University, September 1970.

Ernhart, C. B., Jordan, T. E. & Spaner, S. D. Validity and robustness of selected preschool cognitive measures. Unpublished paper, University of Missouri at St. Louis, 1972.

Goldberg, S., Godfrey, L. & Lewis, M. Play behavior in the year-old infant: Early sex differences. Paper presented at the meeting of the Society for Research in Child Development, New York, March 1967.

Halverson, C. F., Jr. & Waldrop, M. F. Maternal behavior toward own and other preschool children: The problem of "ownness." Child Development, 1970, 41, 839-845.

Hervey, S. D. & Story, N. L. Annual report: The program of evaluation of the Head Start Evaluation and Research Center, Michigan State University. East Lansing, Mich.: Michigan State University, College of Home Economics, 1967.

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.

Moss, H. A. Sex, age and state as determinants of mother-infant interaction. Merrill-Palmer Quarterly, 1967, 13, 19-36.

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

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Systems Development Corporation. Effects of different Head Start program approaches on children of different characteristics: Report on analysis of data from 1968-1969 national evaluation. Santa Monica, Calif.: S.D.C., 1972.

Walker, D. R., Bane, M. J. & Bryk, T. The quality of the Head Start Planned Variation data. Cambridge, Mass.: Huron Institute, 1973, in press.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

Technical Report 19

Risk-Taking 2

Karla S. Goldman
Virginia C. Shipman

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

While there is considerable research on risk-taking behavior for adults, adolescents and older children (Kogan & Wallach, 1967), little or none exists on preschoolers. Similarly, locus of control has been investigated primarily with adults and adolescents (Rotter, 1966). It has been hypothesized that individuals who feel a sense of competence and effectance (White, 1960) and believe they are capable of manipulating their environments and receiving consequences from that manipulation are likely to be willing to take reasonable risks. In locus of control theory (Rotter, 1954, 1966), the individual's belief that his actions either are (internal locus) or are not (external locus) capable of producing consequences in the environment is believed to be an important motivational construct for predicting performance in school and in other cognitive tasks (Lewis & Goldberg, 1969). By studying risk-taking behavior one hopes to learn about the development of locus of control and sense of competence.

In their 1963 study, Battle and Rotter developed a pictorial projective test for children and reported that external locus of control (LOC) was more characteristic of lower-class black children (6th grade) than of middle-class black children or lower-class white children. Milgram (1970), using the Battle-Rotter test to study LOC and level of aspiration in a sample of economically disadvantaged 6-year-olds, predicted that disadvantaged children would demonstrate more external LOC than advantaged children, and within the disadvantaged groups, blacks more than whites. The absence of any racial differences within the disadvantaged group appeared to indicate that the factors responsible for these phenomena are common to disadvantaged children regardless of race.

Although evidence on sex differences in risk-taking at young ages is inconclusive, it tends to indicate that boys may be more inclined towards risk-taking than girls. Slovic (1966), using large volunteer samples of children ages 6 to 16 (N = 1047) at a county fair, studied the influence of age and sex upon children's performance on a task designed to assess risk-taking. The task required the child, seated before a panel of ten switches, to pull one switch at a time for reward or loss. The game ended when the child either chose to stop and collect his winnings or pulled the loss switch. The results revealed no sex differences in risk-taking for children 6 to 10. Significant or near-significant sex differences were found in children 11 to 16, with boys taking more risks. Kass (1964) reported significant sex differences for children aged 6, 8 and 10 placed in a repetitive "pay-to-play" gambling game. The 6- to 10-year-old boys manifested greater risk-taking by choosing the higher risk games. This study, however, had a small number (seven) of Ss in each age and sex category.

Kogan and Wallach (1967) state that Slovic's results must be taken seriously, given the advantages of combining a natural field situation with experimental control. Nonetheless, several other factors must be considered, such as subject sampling bias reflected in the field setting employed, the effect of spectators, and public performing which may maximally pressure boys to be daring, adventurous and bold. Also relevant is the research finding that children in the primary grades have well-defined conceptions of "boy-traits" and "girl traits," including the notion that the typical boy is more daring than the typical girl, and that boldness is positively correlated with popularity for boys, but negatively so for girls (Tuddenham, 1951, 1952). As Slovic (1966) points out, "It would be quite surprising if these social pressures did not result in an increased sex difference in risk-taking propensity [p. 170]." The popular notion that males have a greater tolerance than females for risk is discussed by Kogan and Wallach (1967). In summarizing some of their earlier work on risk-taking and decision-making (in college students), they assert that each sex appears willing to tolerate high-risk levels in pursuit of sex-appropriate goals.

In the absence of an adequate measure of risk-taking or LOC for preschoolers, Risk-Taking 2 was devised for this study to relate risk-taking behavior to LOC measures administered in subsequent years, and to other variables for this age group.

Task Description and Administration

Risk-Taking 2 investigates behaviors of young children who are presented with a choice without prior knowledge of the outcome. The child is asked to choose between a certainty, a toy placed in front of him, and an uncertainty, a paper bag which E had previously explained might contain five toys or none at all. In Year 1 small plastic cars were used for boys and small paper parasols for girls; to simplify procedures, in Year 2 various colored balloons, which were found to be generally liked by both boys and girls, were used. If a child chose the certain (i.e., visible) toy he was shown that the paper bag contained five of those toys and he was administered a second trial with a different bag. The second trial was administered only to Ss who initially chose the certainty. On trial 2 the child was again informed that the bag either would be empty or would contain five toys. The paper bag always contained the five items, and even if the child did not choose the bag, he was given the toys upon completion of the task.

Administration of the task is relatively simple and quick. The tester [illegible]t make certain, however, that the child does not touch the bag since he would then be able to determine its contents, that items are center-positioned for the child and that she (he) does not inadvertently give cues favoring one response.

Scoring

Responses for each trial were scored either 0 for choosing the toy (certainty outcome) or 1 for choosing the paper bag containing the unknown quantity. Refusals and multiple choices were also noted. Derived scores used in structural analyses indicated degree of risk-taking: 0 = not choosing the bag on either trial; 1 = choosing the bag on trial 2; 2 = choosing the bag on trial 1.

Sample Performance

In this sample 60.7% of the children selected the uncertain outcome on the first trial in Year 1 and 59.4% did so in Year 2. That is, they elected to take a risk and chose the paper bag rather than the visible single toy on the first trial. The remaining Ss (approximately 40%) were given a second trial. On trial 2, an additional 18.4% chose the bag while 20.9% still chose the certainty outcome in Year 1; in Year 2 these remaining Ss split equally between the two options (20.3%). Thus, after two trials approximately 79% of the children during this age period were willing to choose an uncertain outcome. Distributions by sex and three-month age groups are presented in Table 1.

There was no linear relationship with age for either trial in either year. On trial 1 in Year 1, however, boys were more likely to choose the uncertain outcome; on the second trial approximately equal proportions of boys and girls took the risk and chose the bag. For the total task in Year 1, 66.8% of boys and 53.7% of girls received the maximum derived score

Table 1

Percentage of Children Choosing the Uncertain Outcome on 1st, 2nd or No Trials in Years 1 and 2

Year 1

Group	N	0	2nd Trial	1st Trial
42-44 mo.	86	14.0	12.8	73.3
45-47 mo.	309	26.2	18.4	55.3
48-50 mo.	34[illegible]	20.9	20.3	58.5
51-53 mo.	375	19.7	19.5	60.8
54-56 mo.	266	18.4	15.4	66.2
57-59 mo.	58	22.4	22.4	55.2
Boys	763	15.1	18.1	66.8
Girls	671	27.6	18.8	53.7
Total	1434	20.9	18.4	60.7

Year 2

Group	N	0	2nd Trial	1st Trial
51-53 mo.	66	24.2	21.2	54.5
54-56 mo.	217	21.2	18.4	60.4
57-59 mo.	195	20.0	21.5	58.5
60-62 mo.	238	16.8	21.8	61.3
63-65 mo.	156	21.8	17.9	60.3
66-69 mo.	9	44.4	33.3	22.2
Boys	464	20.5	20.0	59.5
Girls	417	20.1	20.6	59.2
Total	881	20.3	20.3	59.4

of 2, while 15.1% of boys and 27.6% of girls received the minimum of zero, having chosen the single, visible toy (certainty) on each trial. In Year 2 59.5% of boys and 59.2% of girls received the maximum score, and 20.5% of boys and 21.1% of girls received the minimum. Derived scores for Years 1 and 2 correlated .07.

A repeated-measures analysis of variance (sex x age x SES) was performed on the longitudinal sample (N = 668). Using mother's schooling (below 10th grade, 10th-12th grade, above 12th grade) as an index, no significant SES difference was found. Age, too, was not significant, though sex was. Combining data across years, boys obtained significantly higher scores (F = 10.49, df = 1/663, $p < .001$). There was, however, a significant sex by year interaction (F = 12.54, df = 1/663, $p < .001$), with girls showing a greater increase in risk-taking from Year 1 to Year 2.

Risk-Taking 2 did not correlate* significantly with any other measure in the Year 1 or Year 2 test battery. This may be due to the paucity of indices of risk-taking or other personal-social behaviors, such as LOC, being tapped by these measures. The Mischel Technique, included in the study to assess the ability to delay gratification, correlated only -.01 in Year 1 and .04 in Year 2 with Risk-Taking. In Risk-Taking 2 the risk (uncertainty) is <u>what</u> the child will get, whereas in Mischel it is not what the child will get but <u>when</u> he will get it. (The latter appears to be complicated by the concept of trust. For a further discussion of the Mischel Task the reader is referred to the Mischel Technique Technical Report [Lindstrom & Shipman, 1972].) However, in addition to the task specificity of Risk-Taking 2 in the present battery, there was virtually no consistency in performance across years (r = .07).

---

*These correlations are based on data for the longitudinal subjects only; however, essentially zero correlations were also found for the total Year 1 sample.

Like other stylistic measures included, situational determinants appeared to exert a more pervasive influence on the child's behavior.

Summary

It would appear that children have little understanding of "risk" in this situation, since (a) most preferred the uncertain outcome, (b) this choice had no relationship to other cognitive skills, and (c) responses were independent of age. Moreover, as was found with other variables defining task-specific factors in Year 2, there was little or no correlation with performance in Year 1. This lack of stability apparently results from transient consistencies in the specific testing situation.

The Risk-Taking 2 task appears to provide only limited differentiation for children at this age in that almost 80% of the children chose an uncertain outcome over two trials. It is possible that there is too little "risk" involved in the task and that too little is at stake to really discriminate risk-taking considerations. Other characteristics which may affect task results are degree of previous exposure to risk, characteristics of the risk-taking agent, previous success or failure and value of the object. Further understanding of this task should be provided in later years as measures of LOC and other personal-social behaviors (e.g., those observed in the preschool setting [Emmerich, 1971]) are analyzed.

References

Battle, E. S. & Rotter, J. B. Children's feelings of personal control as related to social class and ethnic group. Journal of Personality, 1963, 31, 482-490.

Emmerich, W. Disadvantaged children and their first school experiences: Structure and development of personal-social behaviors in preschool settings. Princeton, N. J.: Educational Testing Service, 1971, PR-71-20. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Kass, N. Risk in decision-making as a function of age, sex, and probability preference. Child Development, 1964, 35, 577-582.

Kogan, N. & Wallach, M. Risk-taking as a function of the situation, the person, and the group. In T. M. Newcomb (Ed.) New directions in psychology III. New York: Holt, Reinhart & Winston, 1967. Pp. 113-278.

Lewis, M. & Goldberg, S. Perceptual-cognitive development in infancy: A generalized expectancy model as a function of the mother-infant interaction. Merrill-Palmer Quarterly of Behavior & Development, 1969, 15 (1), 81-100.

Lindstrom, D. R. & Shipman, V. C. Mischel Technique technical report No. 19. In V. C. Shipman (Ed.) Disadvantaged children and their first school experiences: Technical reports. Princeton, N. J.: Educational Testing Service, 1972, PR-72-27, Pp. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Milgram, N. Level of aspiration and locus of control in disadvantaged children. Psychological Reports, 1970, 27, 343-350

Rotter, J. B. Social learning & clinical psychology, Englewood Cliffs, N. J.: Prentice-Hall, 1954.

Rotter, J. B. Generalized expectancies for internal versus external locus of reinforcement. Psychological Monographs, 1966, 80, No. 1 (Whole No. 609).

Slovic, P. Risk-taking in children: Age and sex differences. Child Development, 1966, 37, 169-176.

Tuddenham, R. D. Studies in reputation: III. Correlates of popularity among elementary-school children. Journal of Educational Psychology, 1951, 42, 257-276.

Tuddenham, R. D. Studies in reputation: I. Sex and grade differences in school children's evaluation of their peers. II. The diagnosis of social adjustment. Psychological Monographs, 195`, 66, No. 333.

White, R. W. Competence and the psychosexual stages of development. In M. R. Jones (Ed.), Nebraska symposium on motivation. Lincoln, Nebraska: University of Nebraska Press, 1960. Pp. 97-141.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

Technical Report 20

Seguin Form Board

David R. Lindstrom
Virginia C. Shipman

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Seguin Form Board

The Seguin Form Board Test has been employed to assess visual discrimination and matching, and eye-hand coordination. The relation between these abilities and intellectual development is suggested by incorporation of the Seguin into traditional tests of intelligence such as the Merrill-Palmer Scale (Stutsman, 1931) and the Arthur Point Scale of Performance (Arthur, 1943). While these studies provided normative data primarily for middle-class Ss, the Longitudinal Study affords exploration of performance in a sample comprised of Ss mainly from low-income backgrounds.

Relatively few research studies are available which report data for preschool-aged Ss using the Seguin. Gordon and Hyman (1970) reported scores for a sample of Head Start Ss comparable to earlier norms. They suggested that the Seguin was more a measure of visual-motor integration and motor speed than visual discrimination, since it related to form reproduction tasks but not to simple visual discrimination tasks. O'Piela (1968) and Van DeRiet and Van DeRiet (1969) reported significant performance gains for preschool-aged Ss in compensatory education programs which emphasized visual-motor skills. In the absence of controls, however, it cannot be assumed that these increases were due to the programs, as the abilities involved in Seguin performance show rapid growth during the preschool years.

Stott and Ball (1965) factor analyzed several intelligence tests for young children and reported that Seguin time and error scores defined factors differently at age levels 30, 36, 48 and 54 months. At younger age levels, the Seguin and other tasks formed factors terpreted as ability to comprehend verbal directions and skillful manipulation of objects. At the older age levels, factor structures were interpreted as abilities requiring visual foresight and formatting, and spatial judgments and relations. They emphasized the relationship of these factors to cognition rather than to speed and psychomotor ability. Because factor loadings for the same scale differed among age levels, the test may be tapping different abilities at different age levels.

Task Description and Administration

The test materials consist of ten differently shaped wooden blocks (circle, star, triangle, etc.) and a large form board with recesses corresponding to these shapes. The board is placed in front of the child and the blocks are stacked in three prearranged piles on the far side of the board (the child must reach over the board to obtain the blocks). The child is instructed to "see how fast you can put the shapes back where they belong." Three trials are given, each with similar instructions, and S is instructed to go faster on each trial. A trial is terminated when (a) all blocks are placed correctly; (b) a three-minute time period elapses; or (c) the child indicates he is finished, even though his placements are incorrect or incomplete and he has been encouraged to continue.

Administering the Seguin is a fairly simple task but testers require much practice. They must learn to stack the blocks in the proper order quickly and to use the stopwatch accurately, as trial time is a major variable. Further, they must be thoroughly aware of what constitutes an error (any definite attempt to put a block into the wrong recess). Testers also must be careful how they urge the child to continue when he omits some blocks and accept the child's termination of a trial when he says that he is finished.

Scoring

Time (in seconds) and number of errors were obtained for each trial.

Scores used in subsequent analyses were (1) quickest time to correct placement (log 10 transformation), and (2) number of errors for this trial. Given the moderately high correlation between time and errors, the time score only was used in structural analyses.

Score Properties

Since coefficient alpha was not appropriate, lower bound estimates of test-retest reliabilities were obtained by examination of correlations between trials for time and errors. Intercorrelations among time and error scores for Years 1 and 2 total samples are presented in Tables 1 and 2, respectively. As expected, time and errors were somewhat more consistent for adjacent trials. Year 2 correlations were slightly lower than Year 1 correlations, and error scores intercorrelated less than time scores. The correlation between Year 1 and Year 2 quickest-time-to-correct-placement was .55, and for errors, .25. The correlation between time and error scores was .59 in Year 1 and .44 in Year 2.

Table 1

Intercorrelations Among Trials for Time to Placement

	Trial 1	Trial 2	Trial 3
Trial 1	--	.64	.60
Trial 2	.67	--	.72
Trial 3	.62	.74	--

Note.--Values to the right of and above the diagonal represent Year 2 data (N = 828-844).

Values to the left of and below the diagonal represent Year 1 data (N = 973-1044).

All correlations are significant beyond the .001 level of confidence ($r_{.001}$ = .148 for N = 500).

Table 2

Intercorrelations Among Trials of Error Scores

	Trial 1	Trial 2	Trial 3
Trial 1	--	.61	.51
Trial 2	.63	--	.58
Trial 3	.57	.66	--

Note.--Values to the right of and above the diagonal represent Year 2 data (N = 829-844).

Values to the left of and below the diagonal represent Year 1 data (N = 974-1085).

All correlations are significant beyond the .001 level of confidence ($r_{.001}$ = .148 for N = 500).

Sample Performance

Tables 3 and 4 present mean quickest time scores by sex and three-month age interval groups for Years 1 and 2, respectively. A significant decrease in response time from Year 1 to Year 2 was obtained (F = 1067.63, df = 1/531, $p < .001$). Analysis of variance using a median-split for age indicated a highly significant difference in both Year 1 (F = 80.77, df = 1/850, $p < .001$) and in Year 2 (F = 34.60, df = 1/707, $p < .001$) favoring the older group. The actual magnitude of the mean differences between the groups in both years was very small, however. Repeated-measures analysis of variance (sex x age x SES) for the longitudinal sample (i.e., those Ss tested in both years) indicated a statistically significant but quite small sex difference favoring girls (F = 5.82, df = 1/527, $p < .02$) for data combined over years. The mean time taken for girls was slightly less than the mean for boys. Contrasting scores of Ss classified by mother's education--12 years or more, 10-12 years, below 10 years--resulted in a significant SES difference favoring the upper education groups (F = 12.68, df = 2/527, $p < .001$). Mean time to quickest

Table 3

Means and Standard Deviations for Quickest Time to Correct Solution for Year 1 (Log 10 transformation)

Group	N	Mean	SD
42-44 mo.	55	1.74	.18
45-47 mo	226	1.69	.19
48-50 mo	255	1.67	.18
51-53 mo	309	1.59	.17
54-56 mo.	229	1.56	.16
57-59 mo.	55	1.55	.18
Boys	583	1.64	.19
Girls	546	1.63	.19
Total	1129	1.63	.19

Table 4

Means and Standard Deviations for Quickest Time to Correct Solution for Year 2 (Log 10 transformation)

Group	N	Mean	SD
51-53 mo.	70	1.52	0.15
54-56 mo.	203	1.47	0.14
57-59 mo.	203	1.47	0.15
60-62 mo.	239	1.42	0.13
63-65 mo.	148	1.41	0.12
66-69 mo.	13	1.48	0.14
Boys	457	1.47	0.15
Girls	419	1.43	0.13
Total	876	1.45	0.14

solution decreased as amount of mother's education increased. Within-trial performance showed similar trends to the data presented in Table 3; across trials, time also decreased in accord with tester instructions to go faster on each trial, indicating the child's comprehension of the task.

Examination of error scores indicated a decrease across trials and a significant decrease from Year 1 to Year 2 ($\underline{F}$ = 79.88, $\underline{df}$ = 1/531, $\underline{p}$ < .001). Distribution data for errors made during the trial with the fastest time solution are presented in Tables 5 and 6 for Years 1 and 2, respectively. Analysis of variance for age using a median-split indicated a significant difference in Year 1 ($\underline{F}$ = 13.40, $\underline{df}$ = 1/850, $\underline{p}$ < .001) in favor of the older subjects, but not in Year 2. Repeated-measures analysis of variance (sex x age x SES) indicated no sex differences but a significant SES difference ($\underline{F}$ = 3.91, $\underline{df}$ = 2/527, $\underline{p}$ < .02) overall. The mean number of errors decreased as the amount of mother's schooling increased.

Table 5

Means and Standard Deviations for Errors Made During Trial with Quickest Solution Time for Year 1

Group	N	Mean	SD
42-44 mo.	71	2.84	3.58
45-47 mo.	272	2.72	3.71
48-50 mo.	285	2.69	3.54
51-53 mo.	326	2.14	3.40
54-56 mo.	255	2.18	2.72
57-59 mo.	68	2.52	3.26
Boys	668	2.52	3.42
Girls	609	2.38	3.34
Total	1277	2.45	3.38

Table 6

Means and Standard Deviations for Errors Made During Trial with Quickest Solution Time for Year 2

Group	N	Mean	SD
51-53 mo.	70	1.51	1.97
54-56 mo.	203	1.18	1.61
57-59 mo.	203	1.36	1.94
60-62 mo.	239	1.34	2.01
63-65 mo.	148	1.12	1.88
66-69 mo.	13	2.46	5.17
Boys	457	1.44	2.02
Girls	459	1.16	1.90
Total	876	1.30	1.97

Relationship with Other Measures

Structural analysis of Year 1 child test data indicated that Seguin time and error scores loaded on a factor interpreted as "g" or general information-processing skills (Shipman, 1971). This finding was consistent for the total sample and for the longitudinal sample. A similar finding for the time score was also reported for Year 2 data (in extension analyses, the error score was found to correlate significantly with this "g" factor) (Shipman, 1972). This "g" factor was defined by scores on general information, verbal, classification, perceptual and visual-motor measures.

Correlations of Seguin scores with correctness and speed of responding on other measures in Years 1 and 2 are presented in Table 7. The correlations with other time scores indicate little commonality of response; however, the nature of the time scores may account for this finding. Latency to response on the Sigel, Matching Familiar Figures and Preschool Embedded Figures Test involves S's processing of information in a problem-solving situation before responding, while the Seguin time score, being based on time to completion and performance, involves, in addition, perceptual-motor dexterity which these other tests do not. The Seguin time score correlated highest with scores from purported visual analysis tasks (e.g., Picture Completion,

Table 7

Correlations* of Seguin Time Scores and Selected Test Variables

	Seguin Time Year 1 (N = 454-611)	(N = 989-1104)	Seguin Time Year 2 (N = 709-780)
Motor Inhibition: Average Time, Trial 2, for the Walking and Drawing subtests	-.24	(-.34)	-.18
Form Reproduction: Total Score	-.43	(-.47)	-.50
Picture Completion: Total Correct	--	(-.42)	--**
Sigel: Grouping Responses	-.26	(-.34)	-.34
Sigel: Average Time to Response (Log 10)	-.02	(.08)	.06
Johns Hopkins Perceptual Test: Total Correct	-.38	(-.41)	-.36
Matching Familiar Figures: Mean Log (X + 1) of Response Times	.07	(.12)	-.06
Matching Familiar Figures: Mean Errors per Valid Item	.43	(.45)	.50
Preschool Embedded Figures Test: Total Correct	-.34	(-.34)	-.38
Preschool Embedded Figures Test: Average Time to First Response	.12	(.04)	.17
Seguin: Number of Errors (for Trial with Fastest Time) - Year 1	.55	(.59)	--
Seguin: Number of Errors (for Trial with Fastest Time) - Year 2	--	--	.44

*Correlations for Year 1 and Year 2 are based on data from the three-site longitudinal sample; those correlations listed in parentheses for Year 1 are for the total Year 1 sample ($r_{.001}$ = .148 for N = 500).

**Not administered in Year 2.

Matching Familiar Figures) indicating moderate consistency of performance across tasks requiring perceptual discriminations. The correlations in Table 7, and the diversity of tasks included on the "g" factor (e.g., general achievement, verbal ability, auditory discrimination, general task comprehension) reflect the many task components in addition to speed associated with performance on the Seguin.

Summary

The data reported here are similar to those reported by Stott and Ball (1965) and emphasize that cognitive-perceptual abilities are involved in Seguin performance. Sex differences, although small, indicated that girls performed better than boys, and faster response times and fewer errors were found to be associated with age in both years. SES differences were found for time and error scores suggesting experiential factors served to improve scores. In absolute terms, however, these differences were small, reflecting the general finding that perceptual development is less influenced by socio-cultural variations than is language development. Children's comprehension of the task demands appears evident, as across-trials time scores decreased in accord with testers' instructions to go faster; error scores also decreased across trials which may indicate presence of a practice effect.

References

Arthur, G. A point scale of performance: Volume 1, clinical manual (revised). New York: The Commonwealth Fund, 1943.

Gordon, G. & Hyman, I. The measurement of perceptual-motor abilities of Head Start children. Princeton, N.J.: Educational Testing Service, 1970, RB-70-11.

O'Piela, J. Evaluation of the preschool child and parent education project through the use of Elementary and Secondary Education Act, Title I, Funds. Detroit Public Schools, Research and Development Department, Program Evaluation Section, February, 1968.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education and Welfare.

Stott, L. H. & Ball, R. S. Infant and preschool mental tests: Review and evaluation. Monographs of the Society for Research in Child Development, 1965, 30 Serial No. 101.

Stutsman, R. Guide for administering the Merrill-Palmer Scale of Mental Tests. In Lewis M. Terman (Ed.), Measurement and adjustment series. Part II. Mental Measurement of Preschool Children. New York: Harcourt, Brace, & World, Inc., 1931 and 1948. Pp. 139-262. Reprinted separately.

Van DeRiet, V, & Van DeRiet, H. A sequential approach to early childhood and elementary education, Phase I Grant Report. Prepared under Grant CG-822A/O, Office of Economic Opportunity, December, 1969.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

A LONGITUDINAL STUDY

Technical Report 21

Sigel Object Categorization Test

David R. Lindstrom
Virginia C. Shipman

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Sigel Object Categorization Test

The Sigel Object Categorization Test (SOCT) is a method for studying classification abilities in young children. The criteria children employ in classifying and organizing various stimuli into meaningful groups are indicative of both cognitive ability and style. Kagan, Moss and Sigel (1963) define cognitive style as "stable individual preferences in the mode of perceptual organization and conceptual categorization of the external environment [p. 74]." Examination of the young child's categorization ability aids in the study of the growth of logical thought and yields information relevant to his perception and differentiation of the environment. It has been suggested that stimulus differentiation increases with age and is stable within the individual at any one age level (cf. Wohlwill, 1960). As the child matures a broadening of classification behaviors is expected, with growth from single, unidimensional organization to multiple and polydimensional categorization. The SOCT provides information indicative of the individual's transformation of concrete instances into symbolic thought. "This transformation is essential to the development of conceptual thought and it is thereby a critical developmental requirement as a prerequisite for adequate functioning in our world of symbols [Sigel, Anderson & Shapiro, 1965, p. 1]." Differences in classification ability between three- and two-dimensional stimuli in young children has led to work in representational thought and formulation of the distancing hypothesis (Sigel, 1968, 1970). Thus, as a measure of categorization ability, the SOCT can be used to suggest levels of cognitive complexity and ability. In a longitudinal study, the SOCT can be used to investigate cognitive growth and concomitant development of processes associated with much growth.

Scoring of responses is based on two aspects: the verbal level which includes the number of Grouping responses (those in which a meaningful relationship between all items grouped has been established) and number of Non-grouping responses (those in which a noninclusive and idiosyncratic reason is offered for a grouping). Together, the Grouping and Non-grouping responses make up the general category "scorable responses." Also scored on the verbal level are Global responses (responses which are vague and can be used for any grouping) and Nonscorable responses (responses in which a rationale is not given, the reason is not clear enough to score, or no objects have been grouped). The second aspect scored is the type of classification, which is scored when S has grouped objects and represents S's basic rationale for groupings. Scores for type of classification fall into three main categories (see Sigel & Olmsted, no date, for further elaboration of all scoring categories):

Descriptive responses represent groupings based on objective attributes of the stimuli (e.g., form, color, structural characteristics) and reflect differentiation of components of the total stimulus. Descriptive-analytic (i.e., part-whole) responses have been interpreted as indicating reflective and analytic styles (Kagan, Rosman, Day, Albert & Phillips, 1964).

Relational-contextual responses are based on functional or thematic relationships between the stimuli.

Categorical-inferential responses are those based on inferences and class membership concepts wherein every member of the grouping is representative of the same class of inference.

Several studies have found that grouping ability increased with age in preschool and primary grade children and that rationales varied as function of age (Hess, Shipman, Brophy & Bear, 1969; Kagan et al., 1964; Meyer, 1971; Sigel, 1964; Sigel, Jarman & Hanesian, 1967; Sigel & Olmsted, 1970; White, 1971). Changes in rationales over age periods indicated an increasing tendency with age for "children to go beyond the perceptually dominant stimulus characteristics and to analyze, reflect over and use alternative dimensions [Katz, 1971, p. 745]."

Sex differences in classification ability have not been consistent, apparently due to the interaction of sex with socioeconomic status in producing differences. For studies which do not report SES, Davis (1971) and White (1971) found no sex differences for Ss ranging from kindergarten to college. Meyer (1971) reported no sex differences in a middle-class preschool-aged sample. Sigel, Anderson and Shapiro (1965) found middle-class preschool females gave more scorable responses than middle-class boys and all lower-class Ss. Content of responses was found to vary with social class in the same study and was also reported by Sigel and Olmsted (1967) and Sigel and McBane (1966). Middle-class Ss produced more grouping and descriptive-analytic responses. Using an earlier two-dimensional version of the test, Hess et al. (1969) found classification responses and style related to maternal language style and varied as a function of sex and social status.

Intellectual and personality variables have been studied in relation to classification behaviors in order to explore further the cognitive abilities of young children. Sigel and Olmsted (1967) found descriptive responses were related positively to achievement striving, cautiousness and independence for young boys; while for girls, descriptive responses were related negatively to cautiousness and positively to achievement striving. Sigel et al. (1965) reported no difference in mean number of scorable responses as a function of IQ for lower-class preschool-aged Ss, but for middle-class Ss there was a difference. Meyer (1971) reported correlations of .43 and .41 between Raven's Colored Progressive Matrices and number of scorable responses given to three- and two-dimensional stimuli on a pretest with preschool-aged Ss and slightly lower correlations for a posttest. Hess et al. (1968) reported categorical-inferential responses correlated significantly with Stanford-Binet IQ for working-class boys; descriptive responses related significantly to IQ for middle- and working-class Ss combined (boys and girls), for middle- and working-class boys combined and for working-class boys alone.

No studies are available which report reliability of the twelve item object version of the SOCT for preschool-aged Ss; some studies do report reliabilities combining responses from the two- and three-dimensional versions. Sigel and Olmsted (no date) reported a test-retest coefficient of .69 for grouping responses and .71 for scorable responses combining objects and pictures over a six-month interval. Meyer (1971) reported responses to pictures and objects correlated .59 on a pretest and .56 on a posttest (nine months later). Kagan et al. (1964) reported stability of analytic responses on the Conceptual Style Test for elementary school children over a twelve-month period ranged from .43 to .70 with reliabilities being higher for females. Split-half reliability was reported as .94.

Task Description and Administration

Two types of stimuli comprise the test: one set involves 12 relatively familiar three-dimensional objects; the second involves colored photographs of these objects. For either stimulus type, two sorting conditions can be used: the active sort requires S to select objects/pictures to go with one

E has selected; the passive sort uses groups that E has created, and S is requested to tell why the objects in the group are the "same" or "alike." Given the reported difficulty of the two-dimensional set of objects and the passive condition, Year 1 and Year 2 data collection used objects and the active condition (cf. Sigel, Anderson & Shapiro, 1965).

The objects include a book of matches with a blue cover, four multi-colored alphabet blocks glued together, a white spoon, a yellow pencil, a red, white and blue top, a black and brown pipe, a yellow cup, a white-covered 3" x 5" notebook, a blue ball, a white cigarette, a closed box of crayons and a metal bottle opener. The objects can be related in any number of ways, and given the multiple characteristics of the stimuli, the child is faced with the task of seeking commonalities in the face of differences (Sigel & McBane, 1966).

After an initial identification of each object by the child, E selects a different object on each of twelve trials and requests the child to put with it "the ones that are the same or like" the stimulus. Three sets of instructions are available for use: the first uses "same or like;" the second uses "ones that belong with;" the third uses "ones that go with" the stimulus. If the child does not respond to the first instruction, E must use the second; if still no response is evident, E uses the third. Thus, three trials are possible for each item. After the child has selected objects to go with the stimulus, E asks for the rationale for the sort. Latency to first object choice after completion of instructions also is recorded. The test is discontinued if the child piles all objects together with no scorable rationale on four successive trials.

This task requires considerable training. All responses are recorded verbatim; standard probes are used for vague and undifferentiated responses.

E must also be sure to record manual (i.e., pointing) responses, especially as this is often S's indication whether all or only some of the objects are involved in the sort. E must also be sure to indicate the total array of objects to S on each trial to help ensure S pays attention to all stimuli. If response latency is to be recorded, additional training in accurate use of a stopwatch is necessary.

Scoring

Responses are scored using the categories defined previously. The verbal level scores are Grouping, Non-grouping, Nonscorable, or Global and reflect the appropriateness of the child's verbalization to the objects chosen. The type of classification scores represent the child's basic rationale for the grouping and fall into three main styles: descriptive, relational-contextual, and categorical-inferential. Latency to first object choice was also used as an additional index of information processing. Also scored for this study was the adequacy of the child's initial identification of the 12 objects (scored 1-4 as correct label, appropriate label, functional description or incorrect), and the number of different classification categories used.

Scores selected for inclusion in structural analyses of Years 1 and 2 data were (1) total number of grouping responses, (2) average latency to first response (with a log transformation), and (3) sum of correct labels given to the objects. Because the frequencies of individual grouping rationales were very low, these scores were not analyzed separately.

Special Scoring Considerations

Scoring for the SOCT is a relatively difficult task. For the Longitudinal Study, ETS Princeton staff were trained by senior research staff and frequent checks on scoring were made. Further training was done as necessary while scoring was in process. After all protocols had been double scored, they were reviewed by two senior staff members for accuracy, inconsistencies and systematic errors. Agreement was found to be very high. With the approval of the test author, refinements were made to the original scoring system especially in the case of nonscorables (e.g., where a child made sorts but either did not respond, repeated the tester, piled all objects, etc.). These refinements were made in order to provide more differentiated scoring information and to yield information relevant to understanding of nonscorable behaviors.

Score Properties

For the Year 1 and Year 2 scores, estimated reliabilities (coefficient alpha) and correlations across years were as follows:

Response	Alpha - Year 1	Alpha - Year 2	Year 1-Year 2 Correlation
Total Grouping	.91	.90	.23
Latency	.77	.71	.06
Total Correct Labels (only those scored 1)	.62	.44	.26

The grouping and latency scores showed high internal consistency in both years, but there was a substantial decrease in the alpha for the correct label score as a ceiling effect became evident. Across-years correlations were low for all three scores.

Table 1 presents intercorrelations* among grouping, latency and correct label scores. The relationship between correct label and grouping responses suggests that the child's ability to create groups was not highly dependent upon producing a correct label for the objects. Similarly, latency to first object choice had little relationship with ability to group.

---

*Correlations are based on total samples for both years.

Table 1

Intercorrelations Among Selected SOCT Scores for Years 1 and 2

Score	1	2	3
1. Grouping	--	-.11(-.08)	.16 (.17)
2. Latency	-.10 (.11)	--	-.02(-.04)
3. Correct Label	.25 (.10)	-.19(-.03)	--

Note.--Values to the right of the diagonal represent Year 2 data (N = 805). Values to the left of the diagonal represent Year 1 data (N = 1090). Values in parentheses are for the longitudinal sample (N = 558 in Year 1 and 722 in Year 2).

Sample Performance

Tables 2 and 3 present score distributions for grouping responses by age and sex for all Ss tested in Years 1 and 2. Low mean scores for grouping are apparent. In conjunction with this the mean score for sorting rationales was also low; when rationales were given they most often were based on manifest stimulus characteristics, e.g., form and color.

Table 2

Means and Standard Deviations for Sigel Grouping Responses for Year 1

Group	N	Mean	SD
42-44 mo.	59	2.56	3.19
45-47 mo.	226	2.50	3.38
48-50 mo.	249	2.95	3.71
51-53 mo.	282	3.71	3.82
54-56 mo.	230	4.00	4.03
57-59 mo.	44	4.30	4.50
Boys	565	3.02	3.70
Girls	525	3.62	3.88
Total	1090	3.31	3.80

Note.--Range = 0-12.

Table 3

Means and Standard Deviations for Sigel Grouping Responses for Year 2

Group	N	Mean	SD
51-53 mo.	63	4.02	3.84
54-56 mo.	183	4.56	4.04
57-59 mo.	185	4.50	4.08
60-62 mo.	225	5.32	3.96
63-65 mo.	138	5.40	4.17
66-69 mo.	11	3.73	4.43
Boys	412	4.74	4.02
Girls	393	4.96	4.10
Total	805	4.85	4.06

Note.--Range = 0-12.

For Ss tested in both years, analysis of variance using a median-split on age indicated significant differences in grouping responses both in Year 1 ($F = 30.03$, $df = 1/828$, $p < .001$) and in Year 2 ($F = 6.86$, $df = 1/655$, $p < .01$) favoring the older group. For the total longitudinal sample, a significant increase in number of grouping responses from Year 1 to Year 2 was also obtained ($F = 144.71$, $df = 1/468$, $p < .001$). Repeated-measures analysis of variance for the longitudinal sample (i.e., those tested in both years) indicated that sex was a marginally significant variable in both years ($F = 4.43$, $df = 1/463$, $p < .04$) favoring girls. To examine SES effects, Ss were classified by mother's education--above 12 years of schooling, 10-12 years, below 10 years. A significant difference was found when data were combined across years ($F = 18.56$, $df = 2/463$, $p < .001$). The mean number of grouping responses increased as mother's education increased. A significant year x SES difference was also obtained ($F = 7.22$, $df = 2/463$, $p < .001$). This difference indicated that the largest gain in grouping score from Year 1 to Year 2 was made by the middle group; the next largest gain was in the high group, and the low group made the smallest gain.

Average time to first object choice was recorded on the SOCT as an indicator of cognitive style and rate of information processing. Means and standard deviations are presented in Tables 4 and 5 by sex and three-month age intervals. When data were analyzed by age, sex, and mother's education, no significant differences were found. Overall, response times were very short in both years.

Table 4

Means and Standard Deviations for Average Time to First Response for Year 1 (Log 10 transformation)

Group	N	Mean	SD
42-44 mo.	64	.90	.29
45-47 mo.	229	.88	.30
48-50 mo.	245	.92	.34
51-53 mo.	279	.90	.32
54-56 mo.	231	.86	.28
57-59 mo.	54	.88	.28
Boys	572	.90	.30
Girls	530	.87	.30
Total	1102	.89	.30

Data for correct object identification are presented in Tables 6 and 7 by sex and three-month age intervals. Except for the toy top, the objects were readily identified by most children. Almost all children gave appropriate labels or functional descriptions if they did not provide a correct label. Analysis of variance for age (median-split) using longitudinal _S_s indicated a highly significant difference in Year 1 ($F = 15.13$, $df = 1/829$, $p < .001$) and in Year 2 ($F = 5.41$, $df = 1/655$, $p < .02$) favoring the older

Table 5

Means and Standard Deviations for Average Time to First Response for Year 2 (Log 10 transformation)

Group	N	Mean	SD
51-53 mo.	63	0.81	0.23
54-56 mo.	183	0.77	0.20
57-59 mo.	185	0.79	0.20
60-62 mo.	225	0.78	0.22
63-65 mo.	138	0.79	0.18
66-69 mo.	11	0.88	0.21
Boys	412	0.78	0.21
Girls	393	0.79	0.20
Total	805	0.79	0.20

Table 6

Means and Standard Deviations for Correct Identification of Objects for Year 1 (coded 1 only)

Group	N	Mean	SD
42-44 mo.	59	8.61	1.61
45-47 mo.	226	8.31	1.93
48-50 mo.	250	8.64	1.66
51-53 mo.	282	9.03	1.81
54-56 mo.	230	8.97	1.65
57-59 mo.	44	9.32	1.20
Boys	565	8.81	1.70
Girls	526	8.73	1.83
Total	1091	8.77	1.76

Note.--Range = 0-12.

Table 7

Means and Standard Deviations for Correct Identification of Objects for Year 2 (coded 1 only)

Group	N	Mean	SD
51-53 mo.	63	8.67	1.88
54-56 mo.	183	8.79	1.56
57-59 mo.	185	8.97	1.30
60-62 mo.	225	9.26	1.15
63-65 mo.	138	9.07	1.33
66-69 mo.	11	9.27	0.91
Boys	412	9.07	1.32
Girls	393	8.94	1.46
Total	805	9.01	1.39

Note.--Range = 0-12.

subjects. No significant differences were found when data were analyzed by sex and mother's education.

Relationship with Other Measures

Structural analysis of the child test data indicated that in both years grouping responses loaded onto a first factor which was interpreted as "g" or a general information-processing skills factor (see Shipman, 1971, 1972 for a detailed description of these analyses). This first factor was defined by scores on general achievement, and more specific verbal, perceptual, auditory, and perceptual-motor tasks. Results were similar when data were analyzed for both total and longitudinal samples, and for six- and 13-factor solutions. Correlations between grouping scores and scores defining this factor were moderate. Highest correlations were with general knowledge (Preschool Inventory and TAMA) and vocabulary (PPVT, Forms A and B) tests. Classification skills as measured by the SOCT and the Toy-Sorting and Eight-Block Sorting tasks tended to have moderate intercorrelations (range of .31 to .37), and correlations between perceptual analysis scores and grouping responses were moderate (range of .20 to .41). Thus, variables which tended to be associated with grouping ability were scores on other tasks entailing classification ability, verbal production, visual discrimination and general achievement. These relationships increased in magnitude in Year 2, especially for the general knowledge tasks.

In Year 1, the latency score loaded onto a factor which was defined by latency scores from the Preschool Embedded Figures Test (PEFT) and Matching Familiar Figures (MFF). The Sigel latency score was correlated .23 and .46 with each of these respectively. This factor was not related to the "g" factor and was interpreted as a response tempo dimension. In the Year 2 structural analyses the Sigel and Preschool Embedded Figures Test latency scores defined a factor, but the Matching Familiar Figures latency score formed a separate factor. The correlation of the Sigel latency score with PEFT latency was .20, and with MFF latency, .09. "This split of the latency scores may reflect an emerging differentiation of perceptual speed and cognitive style factors [Shipman, 1972, p. 62]."

The score for correct labeling loaded onto the "g" factor in Year 1 for a six-factor solution (Shipman, 1971) based on the total sample. When a 13-factor solution was used for both Years 1 and 2, based on data from longitudinal Ss only, the meaning of this score became more specific. The object identification score loaded onto a specific factor, correlated with the first factor, which was defined by itself in Year 1 but also by Children's Auditory Discrimination Inventory (real words) in Year 2. The correlation between these scores in Year 2 was .17. Thus, vocabulary and the child's familiarity with the meaning of [illegible] as a separate factor tapped, in part, by the object identification score on the SOCT.

Summary

The low mean scores for total grouping responses indicate the task was fairly difficult for these subjects during this age period. Although internal consistency was moderate, response stability across years was low. The ability to recognize and articulate commonalities across stimulus events and order them into logical classes and groupings is one which appears to increase with age as indicated by an increase in grouping responses in the Year 2 data. Within-year age differences also were evident in both years. This suggests that ability to group objects and to provide meaningful rationales for them is one sensitive to both experience and developmental level. Although girls consistently obtained higher scores, these differences were negligible.

Despite the fact that most children were able to recognize and correctly label most of the objects, few were able to give reasons for their groupings. When rationales were given, they often did not relate to discernible characteristics of the stimuli. This emphasizes the cognitive use of language in this task, especially in terms of the child's ability to verbalize his discriminations.

Several Year 1 and 2 measures showed significant SES differences consonant with past research; such differences were found for the SOCT, general knowledge tasks, and tasks which involved the child's language and classification abilities. It was suggested that these abilities may be sensitive to environmental manipulation with Ss in the higher SES groups receiving input most likely to enhance these abilities (cf. Shipman, 197[illegible]).

Although the highest SES group (mother's education over 12 years) had the largest mean grouping score in both years, data from the present study indicate that the middle group (mother's education 10-12 years) made the largest gain in score from Year 1 to Year 2.

The analyses performed indicate that the latency score from the SOCT measured response speed in the present study. Because the number of sorting rationales was so low, it was not possible to examine the relationship between latency and type of rationale. The latency score loaded onto a factor with latency scores from other measures, and was not correlated with the "g" factor. In future years, analysis of the latency score in relation to sorting rationales (e.g., analytic responses) may indicate cognitive-stylistic factors associated with latency.

Although only a small percentage of the sample was able to give scorable responses for their sorts or in fact to make appropriate sorts, the data here can be viewed as "baseline data" for these Ss. Future analyses of responses from the SOCT administered in subsequent years of the study will permit investigation of changes in classification ability associated with environmental influences and experience (e.g., preschool experience). Growth in classification ability will be investigated by examination of changes in frequency of nonverbal sorts and nonscorable responses, and changes in rationales provided, e.g., from color to form responses. With an increase in the number of scorable responses, classification scores also will be used as developmental markers to study the development of logical and analytic thought.

Although the SOCT provides a wealth of data, it is a difficult task to use in large-scale evaluations. Administration of the task requires considerable training and examiners must be extremely sensitive to young children's verbal behaviors. Examiners must also try to obtain as clear a response as possible without challenging the child's own grouping rationales. Scoring of the task also requires considerable training and time, and a thorough understanding of the scoring categories.

References

Davis, A. J. Cognitive style: Methodological and developmental considerations. Child Development, 1971, 42, 1447-1459.

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.

Kagan, J., Moss, H. A. & Sigel, I. E. Psychological significance of styles of conceptualization. In J. C. Wright & J. Kagan (Eds.), Basic cognitive processes in children. Monographs of the Society for Research in Child Development, 1963, 28 (2, Serial No. 86), 73-124.

Kagan, J., Rosman, B. L., Day, D., Albert, J. & Phillips, W. Information processing in the child: Significance of analytic and reflective attitudes. Psychological Monographs, 1964, 78 (1, Whole No. 578).

Katz, J. M. Reflection-impulsivity and color-form sorting. Child Development, 1971, 42, 745-754.

Meyer, W. J. The development of representational competence. Paper presented at the Convention of the International Society for the Study of Behavioral Development, Nijmegen, The Netherlands, July 1971.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Sigel, I. E. The child's attainment of concepts, in M. Hoffman & L. Hoffman (Eds.), Review of Child Development Research, Vol. I, New York: Russell Sage Foundation, 1964.

Sigel, I. E. The distancing hypothesis: A hypothesis crucial to the development of representational competence. Paper presented at the meeting of the American Psychological Association, San Francisco, September 1968.

Sigel, I. E. The distancing hypothesis: A neo-Piagetian view of the acquisition of representational thought. Paper presented at the meeting of the American Psychological Association, Miami Beach, Florida, September 1970.

Sigel, I. E., Anderson, L. M. & Shapiro, H. Categorization behavior of lower- and middle-class Negro preschool children: Differences in dealing with representation of familiar objects. Paper presented at the meeting of the Society for Research in Child Development, Minneapolis, 1965.

Sigel, I. E., Jarman, R. & Hanesian, H. Styles of categorization and their intellectual and personality correlates in children. Human Development, 1967, 10, 1-17.

Sigel, I. E. & McBane, B. Cognitive competence and level of symbolization among five-year-old children. Paper presented at the meeting of the American Psychological Association, New York, 1966.

Sigel, I. E. & Olmsted, P. P. Analysis of the Object Categorization Test and the Picture Categorization Test for preschool children. The Merrill-Palmer Institute, no date (mimeo).

Sigel, I. E. & Olmsted, P. Styles of categorization among lower-class kindergarten children. Paper presented at the meeting of the American Educational Research Association, New York, 1967.

Sigel, I. E. & Olmsted, P. Modification of cognitive skills among lower-class black children: A follow-up training study. In J. Hellmuth (Ed.), The disadvantaged child, Vol. 3, New York: Bruner, Mazel, 1970.

White, K. M. Conceptual style and conceptual ability in kindergarten through eighth grade. Child Development, 1971, 42, 1652-1656.

Wohlwill, J. F. Developmental studies of perception. Psychological Bulletin, 1960, 57, 249-288.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

Technical Report 22

Spontaneous Numerical Correspondence Test

Judith A. Meissner
Edward A. Chittenden
Virginia C. Shipman

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Spontaneous Numerical Correspondence

Background

Instruments derived from Piaget's research on concepts of quantity have commonly dealt with the problem of conservation of number (e.g., Sigel & Hooper, 1968). These instruments, some standardized (Asher, Feldhusen, Gruen, Kane, McDaniel, Stephens & Wheatley, 1971; Goldschmid & Bentler, 1968), generally are applicable within the age range of 5 to 9, since they are designed to assess the transition from preoperational thinking to logical understandings of quantity. By contrast, the present instrument is intended to assess changes within the preoperational period itself, as it attempts to measure the articulateness of the perceptual response to a "number" task. As a secondary purpose, the instrument represents an attempt to construct a measure that could be repeated, intact, over several years, eventually becoming combined with measures of number conservation.

Procedures in which the tester attempts to establish in the child's mind one-to-one correspondence between two rows of objects are commonly used as a prelude to tests of conservation of number. The understanding of correspondence itself is not the focus of assessment in such methods, but the present instrument is primarily concerned with just such understandings. Although the contemporary experimental literature on number conservation is very extensive, relatively little has been done on studying numerical correspondence as a measure of quantitative thinking. The most pertinent reference for the present task remains Piaget's original study (1952), in which children were presented with geometric and random figures made with counters, the child being required to place out the same number of counters as each figure contained. Three qualitatively different stages of performance were found on this correspondence task. The first stage, occurring generally at age four and younger, was that of a nonquantitative Global Comparison in which the child reproduced the configuration of the model but paid no attention to the number of counters comprising each figure. Stage two, Intuitive Correspondence, was a transitional stage in which the child was often able to reproduce both the configuration and number of counters of the actual figure present, but could not make this same correspondence when the dimensions of the figure were changed slightly. Only in stage three, that of Operational Correspondence which began to appear at about age seven, did true conservation of the number of counters finally occur. The other relevant study is that of Rothenberg and Courtney (1969) who, although primarily concerned with number conservation, also included one numerical correspondence item (i.e., reproduction of the tester's row of five blocks). Of the 44 subjects aged 2-5 to 4-4, only 37% produced a row equal in number to the tester's model. According to Rothenberg and Courtney (1969), "These results indicate that the majority of subjects from 2 1/2- to 4-years old are not able to correctly reproduce a simple linear configuration and would be considered to be in Stage 1 (global comparison), as described by Piaget [1952, p. 495]." The majority of a group of 210 older subjects (ranging in age from 4-3 to 5-0) were in stage two, being able to reproduce correctly the existing configuration, but unable to understand that this correspondence also involved one-to-one equivalence of the figures regardless of their rearrangement.

From these previous studies it was decided that numerical correspondence would be an appropriate task with which to study the development of preconservation of number.

Task Description and Administration

The task is an adaptation of the procedure described by Piaget (1952). In the present version, the tester sets out an array of blue ceramic tiles (1" x 1") and, providing the child with his own box of 15 blue and 15 red tiles, asks him to "take out just as many" or "put out the same number." The task is repeated four times: twice with seven tiles, once with eight and once with ten tiles. In three of the presentations (test items 1, 2 and 4), the tester arranges the tiles in a straight line; in one presentation (test item 3), the tiles are set in a designated "random" arrangement.

The test requires nonverbal responses from the subject, takes only about five minutes and is easy to administer. The great majority of children appeared to enjoy the task.

Scoring Procedures

For each of the four items, the tester recorded the number and color of tiles put out by the child. A graphic record depicting the configuration of the child's arrangement was also made. From this information, scores were derived which indicated the extent to which the child's response matched the array set out by the tester according to (a) arrangement of configuration and (b) number.

(a) Configuration matching. For each item, the configuration of tiles by the child was coded into one of four categories:

1. Straight line. These were chains or lines of tile which were relatively straight and ran roughly parallel to the tester's array.
2. Pattern. These consisted of chains which clearly changed directionality (i.e., were not straight) or rows or columns which appeared to represent a pattern or design.

3. Restricted. These were responses consisting of only one or two tiles.

4. Random. No evidence of either straight line or pattern was apparent in these responses.

Each response was scored as "matching" if the child's configuration, as coded, was of the same type as the configuration set up by the tester. The range for possible matching scores was zero to four. A maximum score of four consisted of "straight" configurations on the three items (1, 2 and 4) where the tester arranged the tiles in a straight line and a tester-matched "random" configuration on item 3, the item in which the tiles were randomly arranged by the tester.

(b) Total number of items correct. The second type of scoring procedure ignored configuration and took into account accuracy in the number of tiles put out by the child. A pass/fail method of scoring was used, with one point given for each item only if the child placed out the correct number of tiles on that item. Thus, a perfect score of four required placing out 7, 10, 8 and 7 tiles on items 1 through 4, respectively.

Score Properties

(a) Configuration Matching. Coefficient alphas of .56 in Year 1 and .46 in Year 2 indicate that the configuration measure has moderate internal consistency. A Year 1-Year 2 correlation of .23 was obtained, indicating relatively low stability of the configuration score across years.

(b) Total Number Correct. The coefficient alphas of .61 in Year 1 and .66 in Year 2 show that the total score also has a moderate internal consistency, but somewhat higher than the configuration score. The correlation of .24 obtained between Year 1 and Year 2 revealed a low stability across years, which again may be an indication that this numerical skill is undergoing a developmental change during this time period. The Configuration Matching and the Total Correct scores correlated only .29 with each other in Year 1 and .28 in Year 2, indicating that they were measuring somewhat different abilities.

Sample Performance

(a) Configuration Matching. The Year 1 data presented in Table 1 show that the sample generally was responsive to configuration inasmuch as about 80% constructed "straight line" arrangements on items 1, 2 and 4, and the majority (56%) shifted to the matching arrangement of "random" on item 3. Table 2, which is an age and sex breakdown of Table 1's "straight line" responses of items 1, 2 and 4 and "random" responses of item 3, shows that configuration matching improved with age. With the exception of item 3, scores began to approach ceiling with the older subjects.

Table 1

Percentage Frequencies of Types of Configuration for the Four Test Items in Year 1

		Response Configuration			
Item (configuration)	N	Straight Line	Pattern	Random	Restricted
1 (straight)	1291	80.6	10.5	6.2	2.6
2 (straight)	1287	80.1	12.4	5.4	2.0
3 (random)	1278	22.6	19.7	55.8	1.9
4 (straight)	1269	77.1	13.6	7.7	1.5

Table 2

Percentage Matching Configuration on Four Items by Age and Sex in Year 1

Item	42-44	45-47	48-50	51-53	54-56	57-59	Boys	Girls	N
	Age (months)								
1	71.9	72.0	80.4	84.6	85.2	87.9	78.6	82.9	1291
2	76.6	73.8	79.0	80.2	87.1	87.9	77.6	82.9	1287
3	53.1	53.0	52.7	56.0	60.6	65.5	55.8	55.8	1278
4	66.7	70.1	78.7	76.0	83.7	89.5	75.7	78.5	1269

Tables 3 and 4 give the distributions of the total configuration score for all subjects by sex and by three-month age breakdowns, for Years 1 and 2, respectively. In Year 2, only three of the four study sites (Trenton, Portland and St. Louis) were administered the test, hence the drop in total number of subjects from 1280 in Year 1 to 871 in Year 2. It can be seen in Table 4 that by Year 2, subjects were close to ceiling on the Total Configuration score, obtaining an average of 3.39 out of a total of four possible points, with a standard deviation of .87 points. Larger differences in performance were obtained in Year 1, with a somewhat lower mean of 2.93. The age data in these two tables show a gradual increase in Configuration scores with increasing age (with the exception of the slight reversal of the 54-56 month subjects in Year 2), but it must be emphasized that the absolute score differences are very small. In both years, girls obtained slightly higher mean scores than did boys.

In addition to age and sex breakdowns, the data were also split into a three-way grouping of subjects by mother's educational level (more than 12 years, 10-12 years, and less than 10 years of schooling). This grouping provided a rough index of socioeconomic status.

Table 3

Distribution of Total Configuration Score*
by Age and Sex in Year 1

Group	N	Mean	SD
42-44 mo.	66	2.67	1.19
45-47 mo.	272	2.68	1.19
48-50 mo.	293	2.92	1.07
51-53 mo.	333	2.98	1.11
54-56 mo.	257	3.15	1.05
57-59 mo.	59	3.29	0.89
Boys	673	2.88	1.17
Girls	607	2.99	1.04
Total	1280	2.93	1.11

*The range of scores is from 0 to 4.

Table 4

Distribution of Total Configuration Score*
by Age and Sex in Year 2

Group	N	Mean	SD
51-53 mo.	70	3.31	0.84
54-56 mo.	203	3.24	1.01
57-59 mo.	201	3.34	0.90
60-62 mo.	241	3.47	0.79
63-65 mo.	145	3.55	0.74
66-69 mo.	11	3.73	0.47
Boys	454	3.32	0.92
Girls	417	3.46	0.81
Total	871	3.39	0.87

*The range of scores is from 0 to 4.

For purposes of analysis, only the longitudinal subjects of the study were used (those subjects for whom scores were available in both Years 1 and 2). The age data were divided at the median in order to obtain an "older" and a "younger" group. Older children obtained a significantly higher mean Configuration score in both Year 1 ($\underline{F}$ = 19.43, $\underline{df}$ = 1/956, $\underline{p} < .001$) and in Year 2 ($\underline{F}$ = 10.28, $\underline{df}$ = 1/703, $\underline{p} < .005$). A repeated-measures analysis of variance revealed a Year 1-Year 2 increase in Configuration score which was highly significant ($\underline{F}$ = 117.04, $\underline{df}$ = 1/614, $\underline{p} < .001$) Significant age differences have been found in several other studies of preschool children's number concepts (Rothenberg & Courtney, 1969; Siegel, 1971).

This same repeated-measures analysis of variance revealed a significant sex difference across years favoring girls on configuration matching ($\underline{F}$ = 8.80, $\underline{df}$ = 1/610, $\underline{p} < .005$). Preschool sex differences in spatial and configurational ability generally have not been found (Maccoby, 1966), but since there has been no configuration task found in the literature exactly like the present measure (and neither Piaget nor Rothenberg & Courtney made comparisons on configuration matching ability by sex), a direct comparison with other research cannot be made. These results may reflect sex differences in compliance, with girls more sensitive and responsive to modeling adult behaviors at this age. The repeated-measures ANOVA on the SES data (i.e., mother's educational level) was only marginally significant when the data were combined across years ($\underline{F}$ = 3.46, $\underline{df}$ = 2/610, $\underline{p} < .05$), and the mean difference in absolute terms was quite small. In studies employing a wider range of SES levels, large differences on Piagetian tasks in favor of middle-class subjects have been found (Almy, Chittenden & Miller, 1966; Hess, Shipman, Brophy & Bear, 1969; Rothenberg & Courtney, 1969; Siegel, 1971), adding confirmation to the findings from the present study.

(b) Total Number Correct. Tables 5 and 6 present the percentile distributions and the mean number of items correct by age and by sex for Year 1 and Year 2 data, respectively. These tables show that the Spontaneous Correspondence task was a difficult one for these subjects. In Year 1, only 2% obtained a perfect score of four items correct, and the majority of children (59.4%) received no credit on any item. In the second year, only 5.4% completed all four items correctly, and somewhat fewer than half the subjects (42.3%) had no items correct.

The analyses of variance performed separately by year for longitudinal subjects only showed that age differences in total score were significant in both Year 1 ($F = 16.50$, $df = 1/958$, $p < .001$) and Year 2 ($F = 8.10$, $df = 1/703$, $p < .005$) in favor of the older subjects. The total score increase from Year 1 to Year 2 was highly significant ($F = 79.00$, $df = 1/616$, $p < .001$).

The across-years repeated-measures ANOVA revealed no significant sex differences, but a highly significant SES difference favoring children whose mothers had more schooling ($F = 11.29$, $df = 2/612$, $p < .001$). The SES, sex and age differences are all similar to those generally found in the literature (Almy et al., 1966; Hess et al., 1969; Maccoby, 1966; Rothenberg & Courtney, 1969; Siegel, 1971).

Relationship with Other Measures

(a) Total Configuration Matching. Year 1 correlations* of the Total Configuration Matching score with other Longitudinal Study measures were generally low, the range being from zero to .31. Highest correlations were

*These correlations are based on data from the longitudinal subjects only.

Table 5

Distribution of Total Number Correct* by Age and Sex for Year 1

Group	N	Mean	SD	Percent Response 0	1	2	3	4
42-44 mo.	64	0.56	0.96	64.1	25.0	4.7	3.1	3.1
45-47 mo.	267	0.49	0.83	66.3	24.0	5.2	3.4	1.1
48-50 mo.	301	0.56	0.86	63.5	21.6	11.0	3.3	0.7
51-53 mo.	335	0.74	1.01	55.5	25.4	10.1	7.5	1.5
54-56 mo.	251	0.93	1.24	53.8	21.1	8.8	11.2	5.2
57-59 mo.	56	1.14	1.30	48.2	16.1	10.7	23.2	1.8
Boys	665	0.72	1.04	57.7	24.1	8.9	6.9	2.4
Girls	609	0.66	1.00	61.2	21.7	8.7	6.7	1.6
Total	1274	0.69	1.02	59.4	22.9	8.8	6.8	2.0

*Range = 0-4.

Table 6

Distribution of Total Number Correct* by Age and Sex for Year 2

Group	N	Mean	SD	Percent Response 0	1	2	3	4
51-53 mo.	70	0.91	1.10	48.6	25.7	12.9	11.4	1.4
54-56 mo.	203	1.00	1.10	47.8	23.2	13.8	12.8	3.0
57-59 mo.	201	1.02	1.21	46.3	25.9	11.4	11.9	4.5
60-62 mo.	241	1.28	1.30	37.8	25.3	14.5	15.8	6.6
63-65 mo.	145	1.52	1.38	33.8	19.3	17.9	19.3	9.7
66-69 mo.	11	1.36	1.36	36.4	18.2	27.3	9.1	9.1
Boys	454	1.07	1.23	45.2	25.1	12.6	12.3	4.8
Girls	417	1.28	1.29	39.1	22.5	16.1	16.3	6.0
Total	871	1.17	1.26	42.3	23.9	14.2	14.2	5.4

*Range = 0-4.

with total score on the Form Reproduction Task (r = .31), total score on the Preschool Inventory (r = .28), fastest time on the Seguin Form Board (-.22), total score on Peabody Picture Vocabulary Test (PPVT) Form B (.20), and total score on the ETS Enumeration Task (.20). Year 2 correlations were of the same magnitude, ranging from .02 to .30. In general, the same tasks that correlated most highly with the Configuration score in Year 1 did so in Year 2. The Preschool Inventory total score and Form Reproduction total score each correlated .27, Seguin log fastest time correlated -.30, Peabody Form A correlated .23, and Peabody Form B correlated .20. Of these tasks listed, the Peabody and Preschool Inventory are highly verbal in nature, but the Form Reproduction and Seguin Form Board tap perceptual-motor skills.

(b) Total Number Correct. Year 1 correlations of the Total Correct score with other measures ranged from .01 to .30, with the highest correlations being with performance on the Preschool Inventory (r = .30), Form Reproduction (.30), TAMA (.27), Peabody Form B (.23), ETS Enumeration (.22) and Peabody Form A (.21). Correlations with Year 2 measures ranged from .02 to .42. The Form Reproduction total score correlated most highly (.42), followed by the Preschool Inventory (.36), Johns Hopkins Perceptual Test (.32), Seguin log fastest time (-.35), Preschool Embedded Figures Test total score (.32), Peabody Form A (.32) and TAMA (.31). Here again the same group of tasks correlated most highly with this score in both Years 1 and 2, and as was the case with the configuration matching score, these tasks were mainly verbal or perceptual in nature.

In contrast to the majority of tests in the Longitudinal Study battery, Spontaneous Numerical Correspondence did not always load highly on the general intellectual functioning factor (Shipman, 1971, 1972). Rather, both Total Correct and Configuration Matching scores helped to define a separate factor in the structural analyses for the total sample. When data for the longitudinal sample only were factor analyzed, in Year 1 the scores loaded .36 and .34, respectively, with the first factor. In Year 2, the loading of the total correct score on general competency had increased substantially (.50). The Configuration Matching score, however, defined with smiling in the picture for the Brown Self-Concept Referents Test a separate orthogonal factor. Its correlation with the "g" factor for a 13-factor Promax solution was .05.

This test was intended to be a measure of the child's understanding of quantity in the preoperational period. The question of validity for tasks derived from Piaget's theory is difficult to resolve (Laurendeau & Pinard, 1962). Several indications are present that the Spontaneous Numerical Correspondence task is measuring a skill that has both a cognitive and perceptual aspect and develops both qualitatively and quantitatively with age, in line with Piaget's (1952) analysis. Both the correlational and factor analytic data show that a general ability component is present in this task, and the correlational data also suggest the presence of a spatial or perceptual component. The relatively low magnitude of the correlations of this task with other measures in the test battery would seem to indicate that the test is measuring something quite different from that of the other tasks.

The correlations of the two subscores on this numerical task with the scores on the other quantitative measure, ETS Enumeration, were somewhat lower both in Year 1 (.22 with Total Correct and .20 with Configuration Matching) and in Year 2 (a range of .14 of Configuration Matching with Enumeration Same Number and Order items to .30 of Total Score with

Enumeration II Counting) than were the correlations with other tests in the battery such as Preschool Inventory and Form Reproduction. Researchers such as Wohlwill (1960), Piaget (1952) and Dodwell (1960) have found the operation of counting and other early quantitative tasks not to be closely related to number conservation in children. This present test, which measures pre-conservation skills, has also found a relatively weak relationship between the skills tapped by Enumeration and the pre-conservation stages of number found in Spontaneous Numerical Correspondence.

Although data were not formally collected on the incidence and frequency of Piaget's three pre-operational "substages," inspection of individual protocols revealed that all three substages occurred with this age sample. There were some children who responded to the tester's arrangement in neither number nor configuration. Secondly, there were many children who took pains to match the configuration but who nevertheless were not very close in matching number, and finally there were children who clearly matched configuration with also a very close matching in number. Piaget interprets these three patterns as reflecting substages in the development of understanding of number that remain essentially perceptual in character. In his terms, these substages represent development from "global" to more "articulated" matching responses. Taken as a whole, these data give evidence for the connection of the developmental theory to the test.

Summary

The Spontaneous Numerical Correspondence test was included in the Longitudinal Study test battery as a Piagetian-based measure of the preoperational child's understanding of the concept of quantity. The test is relatively simple to administer, and was apparently enjoyed by most subjects. The two scores were sensitive to age differences in Years 1 and 2 and to SES differences

when analyzing data combined over both years. Sex differences were found only on the Configuration Matching score, as analyzed across years.

Correlations with other tests in the battery were quite low. Noteworthy was the relatively low correlation with the other task purporting to tap quantitative skills, ETS Enumeration. This lack of relationship between quantitative tests given to young children has also been found by other researchers, who have concluded that the development of the understanding of number might not be a unitary ability. Instead, if these measures are each tapping a part of a discrete sequence of number abilities, those at different locations in the sequence would most likely be uncorrelated. Differences in task requirements for performance of the two tests may also have contributed to the low correlation between them. ETS Enumeration requires a pointing or counting response, whereas Spontaneous Numerical Correspondence is comprised of a more complex sequence of responding which includes attending to the tester's request and the tester's array of tiles, selecting tiles from a box, arranging them and finally checking the configuration and number set out with that which the tester has set out.

Although direct evidence on the test's validity is lacking at present, the evidence for the test's theoretical soundness with the data on reliability indicate that the instrument, as presently constructed, can serve as a Piaget-based measure of the preoperational understanding of number. The Total Correct score and the Configuration score appear to be equally good indicators of performance, and their intercorrelation of .29 would suggest that they are measuring somewhat different aspects of ability.

Several researchers have recently used the technique of scalogram analysis to study the development of a set of differing numerical abilities of young children (D'Mello & Willemsen, 1969; Siegel, 1971; Wohlwill, 1960). Although it was not used to analyze the present data, evidence from the study indicates that such an analysis within preoperational tasks such as Spontaneous Numerical Correspondence and Enumeration might be extremely valuable.

References

Almy, M., Chittenden, E. & Miller, P. Young children's thinking. New York: Teachers College Press, 1966.

Asher, J. W., Feldhusen, J. F., Gruen, G. E., Kane, R. E., McDaniel, E., Stephens, M. I. & Wheatley, G. H. The development of new measures of cognitive variables in elementary school children, Phase I and II. Final Report, U. S. Office of Education Contract Number OEC-C-70-4952, 1971.

D'Mello, S. & Willemsen, E. The development of the number concept: A scalogram analysis. Child Development, 1969, 40, 681-688.

Dodwell, P. C. Children's understanding of number and related concepts. Canadian Journal of Psychology, 1960, 14, 191-205.

Goldschmid, M. L. & Bentler, P. M. The dimensions and measurement of conservation. Child Development, 1968, 39, 777-802.

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

Laurendeau, M. & Pinard, A. Causal thinking in the child. New York: International Universities Press, 1962.

Maccoby, E. E. Sex differences in intellectual functioning. In E. E. Maccoby (Ed.) The development of sex differences. Stanford, Ca.: Stanford University Press, 1966. Pp. 25-55.

Piaget, J. The child's conception of number. New York: Humanities, 1952.

Rothenberg, B.B. & Courtney, R. G. Conservation of number in very young children. Developmental Psychology, 1969, 1, 493-502.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Siegel, L. S. The sequence of development of certain number concepts in preschool children. Developmental Psychology, 1971, 5, 357-361.

Sigel, I. E. & Hooper, F. H. (Ed.) Logical thinking in children. New York: Holt, Rinehart and Winston, Inc., 1968.

Wohlwill, J. F. A study of the development of the number concept by scalogram analysis. Journal of Genetic Psychology, 1960, 97, 345-377.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipr , Editor

A LONGITUDINAL STUDY

Technical Report 23

TAMA General Knowledge Test

Lynn E. Gilbert
Virginia C. Shipman

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

TAMA General Knowledge Test*

Purpose

The TAMA was devised as a nonverbal test of general knowledge specifically for the Longitudinal Study and was included for the following overlapping reasons; (a) General knowledge is one of the most prevalent operational definitions of "intelligence." (b) Some general knowledge is important as a base for acquiring other general knowledge. (c) Inclusion of general knowledge measures in the study allows a more comprehensive assessment of explicit school goals than a measurement strategy limited to academic skills and social development (see ETS, PR-68-4, p. C-64 f.).

Various general nonverbal tests have been devised for young children, primarily to determine school readiness. Most, however, focus more on reading and quantitative skills than on a broad fund of information (Boehm, 1969; Cochran & Shannon, 1969). One measure which includes items similar to the TAMA is the Sprigle School Readiness Screening Test, devised in 1965 "to measure the extent to which the child has developed basic skills and abilities needed to negotiate a school program [Sprigle, 1965, p. 1]." While some of the items require an oral response, about half of the test involves a choice from among several pictorial alternatives. A separate subscore for general knowledge, however, is not available.

Task Description

The TAMA General Knowledge Test requires the child to point to the correct picture among three alternatives in response to a question from E.

---

*TAMA was derived from the last names of the team responsible for developing the test: Masako Tanaka, Scarvia Anderson, Carolyn Massad and Dolores Ahrens.

The two practice items and 25 test items are printed on bound cards; most are in black and white, but a few are in color to make necessary distinctions among alternatives (e.g., American flag, traffic light). The correct response is assigned randomly to each of the three positions.

For example, S is shown pictures of a glass, a ball, and a book and asked, "Which of these will break if you drop it on the floor?", or three figures and asked, "Which one is Humpty Dumpty?" Two different forms, with ten items in common, were used in Years 1 and 2. The response mode for the TAMA contrasts with that for the Cooperative Preschool Inventory and the Information subtest of the Wechsler Preschool and Primary Scale of Intelligence, both of which require the child to make an oral response.

The items can be classified in the following categories: social environment, physical environment, health and safety, practical arts, consumer behavior, sports and games, literature, and TV and comics.

This is an easy test to administer and takes approximately 10 minutes.

Scoring

Each item was scored as follows: correct, incorrect, refusal, or indeterminate. Total score was the number of correct choices made and was the score used in the structural analyses of Year 2 child test data. Data from this task were not included in the Year 1 structural analyses, but rather were included in extension analyses, because of a substantially smaller sample size. Prior to keypunching, a portion of the Year 1 data were misplaced and, as yet, have not been recovered.

Score Properties

In Year 1, with $N = 628$, the alpha coefficient of reliability was .65. Although the items differ in content, only five of the 25 biserial correlations fell below .30. Sixteen exceeded .40, and nine were in the range .50 to .64. Alpha for Year 2 (N = 908)* was .64, with r-biserials ranging between .15 and .65. Three items had biserial correlations below .30 and eight exceeded .50. The correlation between Year 1 and Year 2 scores was .46, indicating moderate consistency in performance despite differences in item content across years.

Sample Performance

For the available subsample of 629 cases across the four sites in Year 1, the mean total correct score was 13.8 with a standard deviation of 3.90. In Year 2, for a sample of 908 children in three sites, the mean score was 14.2 (SD = 3.72). Tables 1 and 2 summarize scores and percentile distributions by age and sex subgroups for Years 1 and 2, respectively.

Age by sex by SES** analyses of variance for each year were performed separately on the longitudinal sample; that is, the children for whom data were available for Year 1 and Year 2. In both years SES was significant (in Year 1 $F = 30.64$, $df = 2/413$, $p < .001$; in Year 2 $F = 48.42$, $df = 2/730$, $p < .001$), with children of low SES obtaining the lowest mean scores and high SES children the highest. Using a median age split, the older children scored significantly higher in both Year 1 ($F = 14.32$, $df = 1/413$, $p < .001$) and Year 2 ($F = 20.95$, $df = 1/730$, $p < .001$). No significant sex differences were found.

---

*The TAMA was administered in three sites only in Year 2. It was not included in the short battery given in Alabama.

**Mother's education was used as the index of SES. Low SES children had mothers whose educational level was below the 10th grade; mothers of middle SES children completed grades 10 to 12; and mothers of high SES children attended school beyond the 12th grade.

Table 1

Means, Standard Deviations and Percentile Distributions of Total Score* by Age and Sex, Year 1

Group	N	Mean	SD	Percentiles 10	25	50	75	90
42-44 mo.	47	12.7	3.50	8.4	10.4	12.2	15.0	17.6
45-47 mo.	121	13.2	3.87	8.7	10.6	12.6	16.2	18.4
48-50 mo.	156	13.0	3.89	8.5	10.1	12.6	15.2	18.4
51-53 mo.	164	14.7	3.85	9.9	11.9	14.5	17.4	19.9
54-56 mo.	114	14.5	3.91	9.8	11.9	14.1	17.6	19.5
57-59 mo.	27	15.1	3.26	10.9	12.9	15.1	17.2	19.6
Boys	332	14.0	3.86	9.3	11.0	13.6	16.8	19.2
Girls	297	13.7	3.95	8.8	10.9	13.3	16.4	19.2
Total	629	13.8	3.90	9.0	11.0	13.5	16.6	19.2

*Range = 0-25.

Table 2

Means, Standard Deviations and Percentile Distributions of Total Score* by Age and Sex, Year 2

Group	N	Mean	SD	Percentiles 10	25	50	75	90
51-53 mo.	76	13.2	3.10	9.5	11.1	13.0	15.5	17.4
54-56 mo.	216	13.5	3.91	8.6	10.8	12.8	16.2	19.0
57-59 mo.	209	13.8	3.65	9.0	11.8	13.2	16.3	18.5
60-62 mo.	252	14.7	3.58	10.0	12.4	14.8	17.[illegible]	19.2
63-65 mo.	146	15.3	3.72	10.6	12.6	15.3	17.6	20.7
66-69 mo.	9	14.0	3.64	9.2	10.9	14.6	15.5	19.5
Boys	477	14.1	3.70	9.3	12.0	13.8	16.9	18.7
Girls	431	14.3	3.74	9.1	11.9	14.0	17.0	19.4
Total	908	14.2	3.72	9.2	11.9	13.9	16.9	19.1

*Range = 0-25.

Relationship with Other Measures

Correlations* of the TAMA with other measures were moderate to high over a wide range of cognitive-perceptual tasks. Table 3 presents the correlations which reached .40 for either study year.

Table 3

Selected Correlations of the TAMA General Knowledge Test in Year 1 and Year 2

Measure	Year 1 r	Year 1 N	Year 2 r	Year 2 N
Preschool Inventory: Total Score	.53	399	.70	800
Peabody A: Receptive Language Score	.52	374	.63	789
Peabody B: Productive Language Score	.42	92	.60	779
Form Reproduction Score	.37	400	.47	783
Matching Familiar Figures: Errors	-.36	384	-.45	790
Hess & Shipman Eight-Block Sorting Task: Total Score	.34	379	.48	718
Children's Auditory Discrimination Inventory: Nonsense Words	.32	397	.45	787
Seguin: Log Fastest Time to Correct Placement	-.28	313	-.41	777
Johns Hopkins Perceptual Test: Total Score	.26	382	.42	796
Massad Mimicry: Real Words Score	.19	268	.45	491
ETS Enumeration I: Total Score	.19	381		
ETS Enumeration II: Counting Items			.46	701
ETS Matched Pictures: Total Score	.12	383	.47	798

Note.—$r_{.01}$ = .267 for N = 90.
$r_{.01}$ = .148 for N = 300.
$r_{.01}$ = .115 for N = 500.

Factor analyses of the Year 1 and Year 2 child test data revealed a first factor best defined as general information-processing skills, or "g." The TAMA General Knowledge Test which because of a reduced sample size was

*All correlations are based on data from the longitudinal subjects only.

included in extension analyses only, was found to correlate significantly with this "g" factor (see Shipman, 1971, 1972 for a detailed presentation of these results).

Summary

As a general knowledge test, the TAMA correlated with a wide range of cognitive-perceptual measures and correlated highly with a factor representing general information-processing skills. Its internal reliability, convergent validity, ease of administration, and reduced emphasis on productive skills suggest its usefulness as a supplementary index in test batteries aimed at tapping cognitive abilities. The significant SES and age effects reflect the expected maturational and experiential influences upon the child's increased assimilation of knowledge from his environment.

References

Boehm, Ann E. Boehm Test of Basic Concepts: 1969 Edition. New York: Psychological Corporation, 1969.

Cochran, E. V. & Shannon, J. L. Assessment Program of Early Learning Levels. Orange, California: Edcodyne, 1969.

Educational Testing Service. Disadvantaged children and their first school experiences: Theoretical considerations and measurement strategies. Princeton, N. J.: ETS, December 1968, PR-68-4. Prepared under OEO Contract 4206 and OEO Grant C6-8256.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Sprigle, H. A. Sprigle School Readiness Screening Test. Jacksonville, Florida: Psychological Clinic and Research Center, 1965.

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Technical Report Series

Virginia C. Shipman, Editor

Technical Report 24

Vigor 2

Virginia C. Shipman
David R. Lindstrom

Report under

Grant Number H-8256

Prepared for: Project Head Start
Office of Child Development
U. S. Department of Health,
Education, and Welfare

December 1972

Vigor 2

Measures of the child's vigor were included in the first two years of the study in an attempt to assess immediate energy level. As suggested by research in health and nutritional status as correlates of school performance and learning, vigor level contributes to persistence, motivation and to performance on cognitive tasks (Birch & Gussow, 1970). Thus, a child's vigor may directly affect his performance on specific tasks; it may also indicate his involvement in the testing situation. Moreover, the low-vigor child might be regarded by a tester or teacher as poorly motivated while the highly energetic child might be labeled as aggressive.

Several measures of vigor are available for older Ss, mainly for use in the physical education area. As their age appropriateness was dubious, it was necessary to devise measures for this study. Two vigor measures were included in the test batteries: in one the child was asked to run as fast as he could on a 12 foot long runway (Vigor 1); in the second the number of crank turns the child made in 15 seconds was counted (Vigor 2). Results for the running task were found to be confounded by space limitations, children's fears of running into walls, etc., and thus were not included in further analyses of test data. Vigor 2, then, was the measure used in subsequent analyses.

Task Description and Administration

A large wooden crank was mounted on a stand, and the child was asked to "turn the crank as fast as you can until I say 'stop'." Two trials of 15 seconds each were presented. To establish that the child understood the task and was able to turn the crank, the child was given a practice trial before testing began.

Administering the Vigor 2 task is relatively easy. Testers, however, must time accurately while simultaneously attending to the number of turns made by the child. They must also be sure to remind the child to continue if he stops during the time period and to monitor that he keeps his hand on the crank handle.

Scoring

The score was the number of turns completed for each 15-second trial.

Score Characteristics

The correlation between trials 1 and 2 in Years 1 and 2 was .76 and .74, respectively. The score used for analysis was the mean number of turns across trials; score reliability (coefficient alpha) was .86 for Year 1 and .85 for Year 2. The correlation across years was .52. Thus, Ss showed moderately high consistency in performance on this task within and across years.

Sample Performance

Tables 1 and 2 present the data by three-month age intervals, sex and total group for Years 1 and 2, respectively. A clear increase in number of turns as a function of age is evident in the Year 1 data, but less so in Year 2, as ceiling effects become apparent. For the longitudinal sample (i.e., Ss tested in both years) analysis of variance performed separately by year revealed age (using a median-split) to be highly significant in both Year 1 ($F = 147.93$, $df = 1/1106$, $p < .001$) and Year 2 ($F = 45.94$, $df = 1/714$, $p < .001$), with the older group obtaining a higher score. For the older and younger Ss Year 1 means were 12.55 and 9.90, respectively; Year 2 means were 14.07 and 12.08.

Repeated-measures analysis of variance for longitudinal Ss indicated a

Table 1

Mean Number of Crank Turns by Age and Sex for Year 1

Group	N	Mean	SD
42-44 mo.	84	8.87	2.49
45-47 mo.	316	9.87	2.81
48-50 mo.	348	10.53	3.15
51-53 mo.	372	11.97	3.33
54-56 mo.	277	12.67	3.33
57-59 mo.	61	13.33	3.31
Boys	776	11.45	3.43
Girls	689	10.89	3.27
Total	1465	11.19	3.37

Table 2

Mean Number of Crank Turns by Age and Sex for Year 2

Group	N	Mean	SD
51-53 mo.	73	11.74	3.20
54-56 mo.	203	12.34	3.28
57-59 mo.	204	12.93	3.38
60-62 mo.	245	14.59	3.41
63-65 mo.	148	13.76	3.54
66-69 mo.	11	13.36	4.23
Boy	463	13.75	3.74
Girl	421	12.81	3.18
Total	884	13.30	3.52

significant increase (F = 342.67, df = 1/695, $p < .001$) in the number of crank turns from Year 1 to Year 2. There was also a significant sex difference across years favoring males (F = 16.08, df = 1/691, $p < .001$). When Ss were grouped according to mother's educational level as a rough index of SES (more than 12 years of schooling, 10-12 years, less than 10 years) no significant effects were obtained.

Relationship with Other Scores

Vigor 2 had low correlations with other test scores in both years. For the longitudinal sample, the highest correlations in Years 1 and 2 were with the Preschool Inventory (r = .32 and .27, respectively), Peabody Picture Vocabulary Test, Form A (r = .25 and .29) and Form B (r = .15 and .29), and Seguin Form Board, Fastest Time to Correct Placement (r = -.29 and -.30). All of these measures had high loadings on a general competency factor (see Shipman, 1972, for a detailed description of these results).

When Year 1 child test data were factor analyzed, Vigor 2 loaded onto this factor which was interpreted as general information-processing skills or "g," being primarily defined by scores from verbal, perceptual discrimination and general knowledge tasks (Shipman, 1971). This factor was also interpreted in terms of "non-cognitive" aspects of the child's test performance, such as "ease and willingness to relate and assert oneself in the testing situation, attention, persistence, and task orientation. A common component is the ability to understand and follow directions [Shipman, 1971, pp. 75-76]." Thus, the Vigor 2 score appeared to be determined by more than immediate energy level. Given its low communality estimate in Year 1 and the lack of other similar measures, the nature of these other factors is unclear. The significant age differences obtained in both years suggests that physical coordination may also have affected the score. The fact that age correlated higher with the Vigor 2 score in Year 1 (r = .34) than in Year 2 (r = .25) suggests that at younger age levels physical coordination and comprehension of task demands affected this score more than at older levels.

Factor analyses of Year 2 child test data for the longitudinal sample also indicated a "g" factor (Shipman, 1972). However, in Year 2 Vigor 2 did not load onto this general competency factor; rather it defined another factor with smiling in the photograph taken for the Brown Self-Concept Test. These findings suggest some generalization of personal-social behaviors across tasks, including willingness to please the examiner, cooperation and attention, and also differential task meaning during this age period.

Summary

From the correlational and factor analytic data it appears that the Vigor 2 task measured more than immediate energy level; however, the nature of these other components is not clear at this time. (There were no obviously similar measures, and it had low communality with other measures in the test batteries.) The data indicate that cognitive and motivational variables were associated with the Vigor 2 score. Stable sex differences across years suggest the possibility that sex-typed behaviors (e.g., assertiveness) were also being tapped. The significant age effects obtained suggest that physical coordination was a factor in performance for this age sample. However, the extent to which this measure also taps vigor, persistence and/or willingness to please the examiner is unknown. "Its lack of loading on the first (general competency) factor in Year 2 and its correlation with smiling suggests that for this age sample differences in coordination and task comprehension may be less influential than the child's orientation to the social context of testing [Shipman, 1972, p. 63]."

References

Birch, H. G. & Gussow, J. D. Disadvantaged children: Health, nutrition and school failure. New York: Harcourt, Brace & World, 1970.

Shipman, V. C. 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, PR-71-19. Prepared under Grant H-8256, Department of Health, Education, and Welfare.

Shipman, V. C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N. J.: Educational Testing Service, 1972, PR-72-18. Prepared under Grant H-8256, Department of Health, Education, and Welfare.