DOCUMENT RESUME

ED 068 522 TM 001 883

AUTHOR Shipman, Virginia C.; And Others

TITLE Disadvantaged Children and Their First School Experiences: ETS-Head Start Longitudinal Study: Structure and Development of Cognitive Competencies and Styles Prior to School Entry.

INSTITUTION Educational Testing Service, Princeton, N.J.

SPONS AGENCY Office of Child Development (DHEW), Washington, D.C.

REPORT NO ETS-PR-71-19

PUB DATE Dec 71

NOTE 265p.

EDRS PRICE MF-$0.65 HC-$9.87

DESCRIPTORS *Cognitive Development; *Data Collection; *Disadvantaged Youth; *Educational Experience; Evaluation Methods; Item Analysis; *Longitudinal Studies; Preschool Children; Research Methodology

IDENTIFIERS *Project Head Start

ABSTRACT

In a continuing description of a Head Start longitudinal study, analyses are presented of the interrelationships among individual measures of the child's performances prior to school entry, accompanied by brief descriptions of the tasks and the scores used. Despite the size and extensiveness of the data base, the findings are considered tentative until further data is collected on socio-cultural determinants, developmental trends, and other interrelationships. This report describes the interrelationships among certain cognitive, perceptual, and personal-social behaviors of the children, age 4, in the first year of the study as assessed by the initial test battery. Chapters of the report include characteristics of the sample, methodology, results and discussion, and conclusions. Structural analyses of the Year 1 child test data yielded 1) a general ability dimension (i.e., information-processing skills) cutting across contents and operations sampled in the cognitive test battery, and 2) a stylistic response tempo dimension. Descriptions of each of the individual child measures are presented in the appendices, which comprise about half the report. (LH)

PR-71-19

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Structure and Development of Cognitive Competencies and Styles Prior to School Entry

Virginia C. Shipman

in collaboration with
John Barone
Albert Beaton
Walter Emmerich
William Ward

December 1971

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

ED 068522

DISADVANTAGED CHILDREN AND THEIR FIRST SCHOOL EXPERIENCES

ETS-Head Start Longitudinal Study

Structure and Development of Cognitive

Competencies and Styles Prior to School Entry

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 1971

TABLE OF CONTENTS

Preface

This is the sixth report describing the progress of the Longitudinal Study conducted under Contract OEO 4206 and Grants H-8256 and CG-8256. The first report (PR-68-4) discussed theoretical considerations and measurement strategies proposed for the study of disadvantaged children and their first school experiences. The second (PR-69-12) and third (PR-70-2) reports described operations during the first two years of the study. 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 fourth report (PR-70-20) gave a detailed description of the initial longitudinal sample in Portland, St. Louis and Trenton, prior to enrollment in school. It was based on the first analyses of 16 of the 33 instruments administered during 1969, including a parent interview and medical examination designed to elicit information about family and environmental characteristics. The fifth report (PR-71-20) dealt with the structure and development of personal-social behaviors in preschool settings in Portland, St. Louis, and Trenton.

The present report continues the description of the initial sample, incorporating data from Lee County and what is now known about the child's enrollment in Head Start or other preschool programs. The major focus of this report, however, is to present the first analyses of the interrelationships among individual measures of the child's performances prior to school entry, accompanied by brief descriptions of the tasks and the scores used. Despite the size and extensiveness of the data base, such findings must nevertheless be considered tentative; important clues to interpretability await the relating of these data to socio-cultural determinants, developmental trends, and to interrelationships that may become increasingly apparent with measurements in subsequent years.

ACKNOWLEDGMENTS

This report, a description of the initial child test findings, is the product not only of the current project staff (see Appendix D), but also of many other contributors at Educational Testing Service and the study sites.

For her able leadership during the difficult early years of the project, primary recognition is here acknowledged Scarvia B. Anderson. Also, the guidance and direction given to the study by the initial Steering Committee, Samuel J. Messick (Chairman), Albert E. Beaton, Walter Emmerich, Edmund W. Gordon, Winton H. Manning, Marshall P. Smith, Silvan S. Tompkins, and Melvin Tumin is due special recognition. Their questions, ideas, and constructive criticisms contributed greatly to both the form and substance of the study.

Special thanks are due the former Local Coordinators: Mrs. Lida Campbell, Lee County, Alabama; Mrs. Verna Shepherd, Portland, Oregon; Mr. Ronald Greeley and Mr. Bobby Westbrooks, St. Louis, Missouri; and Mr. Conrad McLean, 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 the tester trainers: Anne Bussis, Rosalea Courtney, Diran Dermen, Martha Friendly, Karla Goldman, Sandra Landes, Jean Orost, Masako Tanaka, Phyllis Ward, William Ward, and Patricia Warren, who spent many hours traveling to and from the test centers. Without their ability and willingness to function in a variety of roles, and most of all, without their patience and humor, the study could not have progressed through its initial trying stages. Gratitude must also be expressed for the monitoring and field consultation provided by ETS Regional Office staff: Junius Davis, Roderick Ironside, Chandra Mehotra, Daniel Norton, Santelia Knight, Robert Lambert and George Temp.

For measurement and analysis coordination during the initial year, particular thanks must go to John Barone, Ruth Ekstrom, Richard J. Martz and Victor Wichert; and thanks are due the women who painstakingly coded all the data presented here under the able supervision of Joan Tyson. The continuing support provided by the ETS administration, particularly the administrative staff of the Developmental Research Division, has been especially important for the study's progress.

The study has also benefited greatly from the constructive criticism and support provided by the Office of Child Development, particularly the Head Start Evaluation Research staff in Washington and the Head Start Research Advisory Committee. Appreciation for their counsel and understanding is here expressed to Dr. Lois-ellin Datta, Mr. Richard Orton and Dr. Edward Zigler of the Office of Child Development, to Dr. John McDavid, former Director of Head Start Research and Evaluation, and to Dr. Urie Bronfenbrenner, Dr. Boyd McCandless, Dr. Alfred Yankauer and the late Dr. Edward Suchman.

No report, of course, appears automatically on paper. It, too, is the product of many persons' efforts. I am grateful to Albert Beaton for his contribution to the detailed explanation and description of the sample in Chapter 2 and to John Barone for the description of his well-designed date processing system in Chapter 3. The following ETS research staff took major responsibility for describing and interpreting certain portions of the data found in Appendix B. They are:

Scarvia Anderson	Peabody Picture Vocabulary Test Preschool Inventory TAMA
Anne Bussis	ETS Matched Pictures Language Comprehension Task 1 Seguin Form Board Test
Edward Chittenden	Enumeration I Spontaneous Numerical Correspondence
Diran Dermen	Picture Completion (WPPSI) Preschool Embedded Figures Test
Walter Emmerich	Boy-Girl Identity Task
Carolyn Massad	Children's Auditory Discrimination Inventory Massad Mimicry Test I
Masako Tanaka	ETS Story Sequence, Part I
William Ward	Matching Familiar Figures Test Motor Inhibition Test Open Field Test

The writer also wishes to express appreciation to William Meyer and Irving Sigel of the External Advisory Committee for their aid in interpreting the test findings. For providing the daily coordination with analysis, research, and production staff, in addition to assistance with proofing and editing, special thanks go to Susan Simosko. Appreciation is also expressed to Thelma Benton and June Daly for their care and speed in typing the several drafts and final copy. I am also grateful to Lynn Gilbert, David Lindstrom, James Towery, Ihor Wynnyckyj for their ability and willingness to perform a variety of demanding functions.

Special gratitude goes to Walter Emmerich, Robert Linn, and William Ward for reviewing the various sections of the report, providing thoughtful and constructive criticism throughout, and for their support and counsel throughout the study.

Deepest gratitude, however, goes to the children and their families who praticipated in the study. We are grateful for their belief in us and in the purposes of the study. This report discusses the child test findings from the first year of the study. To all those children who, while pointing at pictures, talking, laughing, walking around the room, turning pages, putting blocks in holes, squirming, were helping us to understand the delightful complexity that is the young child, go our deepest thanks. Together we hope to provide insights that will bring about meaningful educational change.

Virginia C. Shipman

Princeton, New Jersey
December 31, 1971

INTRODUCTION

The Longitudinal Study of Disadvantaged Children and Their First School Experiences was initiated in the spring of 1967 as a cooperative venture of the Head Start Research Office (Office of Economic Opportunity) and Educational Testing Service. The study brings together the concerns of the psychologist, sociologist, and educator as it seeks answers to the questions: what are the components of early education that are associated with the cognitive, personal, and social development of disadvantaged children; what are the environmental and background variables that moderate these associations; and how do these moderators produce their influence?

The specific age range chosen for study was the critical developmental span of approximately 4 through 8 years of age--or from two years prior to entrance into the 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 1,800 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 on these children and their families, communities and schools is planned to continue through spring of 1974. Of particular interest as the study progresses is identification of differential growth patterns that may be associated with certain characteristics of Head Start and Follow Through programs and their interaction with characteristics of the child and his family.

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 in the ordinary way by the parents involved, after the study was underway. Thus, given a lack of control in assigning children to "treatments" or programs, the 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, one can also assess the comparability of beginning grade school experiences for both Head Start and non-Head Start youngsters--e.g., the degree to which primary grade curricula are congruent with and capitalize on what the child has learned in preschool. Finally, a longitudinal design 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 techniques of assessing young children and their environments. It offers the possibility to:

a. Determine the cognitive, personal, social and physical characteristics of "disadvantaged" children prior to any formal preschool experience, and to relate these characteristics to home and community variables;

b. Determine the differential characteristics of families that do and do not send their children to Head Start;

c. Identify the characteristics of preschool and primary grade programs in the study sites and to determine the relationships among these characteristics within and between the educational levels involved;

d. Determine the cognitive, social and personal outcomes in children that seem to be associated with various aspects of compensatory preschool experience, and to study the permanence of such effects through the first three primary grades;

e. Determine the relationship of Head Start to family and community characteristics and attitudes;

f. Relate particular characteristics of children and their growth patterns to particular characteristics of families and educational programs;

g. Determine relationships among physical, personal, social, and cognitive characteristics of children in each of the years of the study;

h. Describe changes in the structures of cognitive abilities and personal-social characteristics of these children over the crucial developmental period of the study;

i. Develop much needed and, it is hoped, generally useful techniques for the assessment of some of the individual and environmental characteristics under consideration.

The initial study report (ETS, PR-68-4) 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 school programs. Selection of these measures followed certain inherent assumptions about what we felt was necessary to accomplish the goals of the study. Whenever possible, multiple sources of information about a particular phenomenon were proposed (e.g., verbal behavior was seen as a function of the stimulus materials, the communicator-communicant relationship, and the purpose of the act--to inform, seek help, express emotion). We emphasized process rather than static variables, especially those process variables involving parent-child and teacher-child interactions, such as modes of information-processing and reinforcement strategies. Implicit throughout was our belief that only for the intermediate purpose of structural analysis and measure derivation within domains could one separate cognitive-perceptual and social-personal domains or study the child without taking his environment into account.

The present report describes the interrelationships among certain cognitive, perceptual and personal-social behaviors of the children in the first year of the study as assessed by the initial test battery. The questions asked of the data were: To what extent are these indices of the functioning of the 4-year-old describable in terms of differential processes? How do cognitive styles and competencies interact? Within the particular age period represented, are differential results obtained by age, sex, social status or general ability level of the child, 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.

The report consists of five chapters, of which this introduction is the first. Chapter 2, Characteristics of the Sample, provides tables and statistics which indicate both the composition of the sample and the degree to which we were successful in unconfounding its major independent variables. Chapter 3, Methodology, presents a brief discussion of how the data were gathered as well as a statement about the methods of analysis (such as coding, validity checks, computer procedures, etc.). Chapter 4, Results and Discussion, presents the findings from the various structural analyses of the test data, including comparisons by major subject classifications. Chapter 5, Conclusions, summarizes and discusses the general results of the analysis to date and presents a statement of plans for further analysis. Brief descriptions of each of the individual child measures are presented in the appendices.

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. 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.

CHAPTER 2--CHARACTERISTICS OF THE SAMPLE

Introduction

Chapter 2 describes the initial sample, the basis for selection of sites, and certain demographic characteristics (i.e., parents' occupational and educational level, race, and the study child's sex and later attendance in Head Start or other preschool programs) that emerged from the nonrandom selection of children and their families. We had anticipated disproportionate numbers of children in the above categories because of the basic design of the study. And though this disproportion is a necessary characteristic of the sample, it does complicate interpretation of general means because the groups defined through a simple classification on a single variable will not have equal numbers of children in important related classifications. Thus, a major purpose of this chapter is to point out some of the disproportionalities in terms of single and multiple classifications and to caution the reader against unwarranted interpretations of the results reported later in Chapter 4.

The information is essentially the same as reported earlier in Progress Report 70-20, except that numbers have been updated on the basis of the most recent information from school records, and preschool information for the Lee County sample has been added.

Since the reader may find our necessarily detailed accounting somewhat burdensome, we have tried to lighten his labors by first presenting the following summary of major findings:

The attempt to gather data on children in the four selected sites was, in general, successful. At least partial data were obtained for a total of 1875 children, 99.6% of the 1882 children originally expected from these four communities (ETS, PR-68-4). However, the distribution of children from site to site was different from our expectations, since we had expected St. Louis and Trenton to be our large sites (and we were least successful in enrolling subjects there), but found more children than we had anticipated in Lee County and Portland (and we were most successful in enrolling subjects there). The other problems were the slightly older ages at testing time of the St. Louis sample, because we had extended their test-period (although the ages of the children are actually in the appropriate rante) and the impossibility of collecting full data on all subjects.

There are, of course, a number of disproportionalities in the various classifications of importance. There are almost one and three-quarters times as many blacks as whites, more boys than girls, more children who did attend preschool programs, and various interactional differences such as different proportions of blacks and whites attending Head Start. These disproportionalities make the interpretation of general means quite difficult, 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. The sample, then, dictates our caution in interpreting such measures.

Such differences in the numbers of children in various classifications is a necessary part, in some ways a desirable 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. As recompense for the disproportionality, however, we have a measure, albeit crude, of the naturally occurring interrelationship among the classificatory variables at various sites.

The disproportionalities, in any case, do not prevent statistical estimation of effects that would be expected if the sample were proportional. Disproportionality does affect the power of tests to reject alternate hypotheses, but we feel this diminution is not of primary importance.

Some of the salient facts about the sample are these:

1. The number of subjects at different sites varies, with Lee County and Portland together constituting about 60% of the sample.
2. The sample is 62% black.
3. Boys make up 53% of the sample. For the four sites they make up 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, 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 whites are, generally, better educated 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--the 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 Selection of Sites

The sites were selected from areas where there is an opportunity for children to attend Head Start, and thus from areas with a substantial proportion of the population below the poverty level. Considerations of cost and feasibility of the study determined that four communities could participate, and these were selected according to the following major criteria:

1. Program. To be considered, a school system had to serve children who had an opportunity to attend a year-long Head Start program. To increase the variety of preschool-primary grade experiences, we preferred school systems with Follow Through programs and tried for at least one without a kindergarten.
2. National spread. Urban-rural variation, population stability, and representation from different sections of the country were all considered vital criteria.
3. Sufficient number of students. A community was considered eligible if it had a sufficient number of children in school and in the Head Start program. We attempted to obtain a reasonable racial mix and also took into account factors that might significantly change the area's characteristics during the life of the study.

4. Opportunity to follow. Bussing of children to schools outside their home districts and high mobility reduced the chance of a city being selected.
5. Cooperation. The study would, of course, be impossible without the cooperation of the community, including its school officials and community leaders. Areas whose continued support was doubted were disqualified.

As an added condition, we decided that one participating community should be relatively near Princeton, thus making possible a close interaction between ETS staff and a local site.

The selection procedure began by examining a list of the 30 school systems having Follow Through programs at the time. The list was scrutinized carefully in terms of the other criteria and several systems were selected for further investigation. Members of the ETS staff visited the respective sites for additional information, including evidence of willingness to engage in a relatively long-term study. Since the Follow Through program was nonexistent in any Southern rural school system which met all our criteria, additional lists of Southern communities had to be reviewed as well. After an extensive period of information-gathering and the preparation of a list of eligible pairs of cities to guide our selection, the following study sites were finally chosen:

a. Lee County, Alabama. Lee County is mainly a Southern rural area. There are two small cities, Auburn and Opelika, within the county, but outside the city limits the area is distinctly rural and poor. Auburn is dominated by its university which is a major employer in that city. Opelika has a few small factories and serves as the county seat. The population is approximately 33% black (OEO, 1970).

b. Portland, Oregon.* Portland is a medium-size city on the West Coast. Its population is fairly stable, having risen from 373,000 in 1960 to 375,000 in 1970. About 6% are black. Unlike the populations of other large cities, Portland whites have not fled to suburbia. The population is better educated than in many other parts of the country, and poverty in Portland is not as intense as in our other sites.

c. St. Louis, Missouri.** St. Louis is a central city, with declining population amid quickly growing suburbs. The city's population dropped from 750,000 in 1960 to 607,000 in 1970. As the white population moved out of the city, the non-white population increased from approximately 29% in 1960 to 43% in 1965; it is believed to be nearly 50% in 1970. Largely industrial, the city is also a trading center.

d. Trenton, New Jersey.** Trenton is a small city on the Eastern seaboard. The city's population dropped slightly from 114,000 in 1960 to 102,000 in 1970. The non-white population was estimated to be 35%-38% of the total population in 1968. The city is industrial and also serves as the state capital.

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, the schools in the target districts are located near Head Start centers. It is in these school districts that the Longitudinal Sample is expected to be enrolled when they reach third grade in the fall of 1973. In each school

---

*The statistics reported are based on 1970 U. S. Bureau of Census figures supplied by Opinion Research Corporation, Princeton, N. J.

**The statistics reported are based on 1970 U. S. Bureau of Census figures supplied by local city officials.

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 speaking a foreign language, and those with severe physical handicaps. The 1969 sample was identified through a canvass of each neighborhood of the school districts and an enumeration of the resident children.

The Basic Sample

The number of children on whom information has been collected is shown in Table 2-1. These are the children who fit all the qualifications for membership in the sample and about whom we have collected at least one piece of information in the 1969 testing program. In some cases the data available for the children included are incomplete.

There are some fairly substantial differences in sample size by site; Lee County and Portland have over 500 cases, whereas Trenton and St. Louis have under 400. Consequently, there is a need for caution in interpreting statistics computed over all subjects since any factors associated with site are disproportionately represented.

Racial composition: Racial composition varies strikingly from site to site. The basic numbers are shown in Table 2-2. Table 2-3 shows these same figures as percentages of the children in a community. We see that the total sample is 62.5% black and 36.4% white, with a few (1.0%) classified as "Other" (i.e., Puerto Rican, American Indian). The proportion of blacks varies sharply from site to site with as many as 77.8% of the Trenton sample being black, and only 47% in Lee County. Therefore, general comparisons from site to site will inevitably require consideration of racial differences.

Table 2-1

Number of Subjects in Each Site

Site	N	%
Lee County	593	31.6
Portland	542	28.9
St. Louis	353	18.9
Trenton	387	20.6
TOTAL	1875	100.0

Table 2-2

Racial Composition in Sites

Site	Black	White	Other	Total
Lee County	279	312	2	593
Portland	350	180	12	542
St. Louis	243	109	1	353
Trenton	301	82	4	387
TOTAL	1173	683	19	1875

Table 2-3

Racial Composition in Sites by Percentages

Site	Black	White	Other	Total
Lee County	47.0	52.6	0.4	100.0
Portland	64.6	33.2	2.2	100.0
St. Louis	68.8	30.9	0.3	100.0
Trenton	77.8	21.2	1.0	100.0
TOTAL	62.5	36.4	1.0	100.0

Sex differences: As one might expect, there are small differences in the numbers of boys and girls from site to site. Summary statistics are in Table 2-4 and are expressed in percentages in Table 2-5. The percentage of boys and of

Table 2-4

Number of Children in Each Site, Classified by Sex

	Boys	Girls	Total
Lee County	323	270	593
Portland	292	250	542
St. Louis	180	173	353
Trenton	199	188	387
TOTAL	994	881	1875

girls is about equal in Trenton and St. Louis, but there is a disproportionately large number of boys in both Lee County and Portland. The result is that the total sample is 53% boys and 47% girls. This difference is sufficient to warrant care in making general comparisons of Lee County and Portland with Trenton and St. Louis, but it does not appear as serious as the confounding on some of the other variables.

Table 2-5

Percentage of Children in Each Site, Classified by Sex

	Boys	Girls	Total
Lee County	54.5	45.5	100.0
Portland	53.9	46.1	100.0
St. Louis	51.0	49.0	100.0
Trenton	51.4	48.6	100.0
TOTAL	53.0	46.9	100.0

Preschool attendance: The sample statistics for attendance in Head Start and other preschool programs are shown in Table 2-6 and the percentages are shown in Table 2-7. It should be noted that Head Start was not available to Lee County children until their kindergarten year.

Table 2-6

Number Attending Head Start and Other Preschool Programs, Classified by Site

	HS	PS	No Known	TOTAL
Lee County	247	113	233	593
Portland	219	74	249	542
St. Louis	133	12	208	353
Trenton	125	55	207	387
TOTAL	724	254	897	1875

Table 2-7

Percentages Attending Head Start and Other Preschool Programs, Classified by Site

	HS	PS	No Known	TOTAL
Lee County	41.7	19.1	39.3	100.0
Portland	40.4	13.7	45.9	100.0
St. Louis	37.7	3.4	58.9	100.0
Trenton	32.3	14.2	53.5	100.0
TOTAL	38.6	13.5	47.8	100.0

The children are divided into three groups. The first group consists of children who attended Head Start during 1969-70 in Portland, Trenton, and St. Louis and during 1970-71 in Lee County. Information was taken from Head Start registers in the communities, and the number given is the minimum number of Head Start children. The second group, other preschool (PS), consists of children who are known to have attended other preschool or nursery programs during 1969-70 in Portland, Trenton, and St. Louis and during 1970-71 in Lee County, so this too is a minimum number. Children who were not on Head Start or other preschool lists are in the "no known" category; it is likely that many of these children attended neither Head Start nor other preschool programs, but this category also includes children who may have moved out of the community and were enrolled in Head Start elsewhere or those who were enrolled in Head Start out of the general area. As the children in the "no known" category are followed up, they may be reassigned to the Head Start or other preschool categories.

Across the three urban sites 38.6% of the children attended Head Start. In Lee County 41.7% attended Head Start. However, we note that the number of children in the Head Start category at the individual site runs from 32.3% to 41.7% and the number in the preschool category runs from 3.4% to 14.2%. As indicated later, there are substantial interactions between race and Head Start attendance which vary from site to site; this may, perhaps, make Head Start children incomparable to other children at the different sites.

Cross-Classification by Major Variables

The following section contains tables displaying all cross-classifications of the major variables: site, race, sex, and Head Start attendance for Portland, St. Louis, Trenton, and Lee County.

Complete cross-classification: Table 2-8 contains a complete cross-classification by the four major variables. Although there are a substantial number of void cells, there are none in the areas of particular interest. Void cells occur only in the cells representing "other preschool programs" and in the "other" racial category. It is therefore possible to estimate a mean value for each cell of black or white children by Head Start or by known preschool program for any measured variable, although the means for the largest cell (Lee County's black males in the Head Start category) will be much better estimated than for the smallest cells (e.g., St. Louis's one white female and Lee County's one black male in the Preschool category).

Race by sex classification: Since there are often differences in performance level of boys and girls, we now ask whether there is the same percentage of black boys as white boys and black girls as white girls. The percentages are shown in Table 2-9.

Overall, the boys are a substantial majority in the black sample and a slight majority in the white. This relationship is not consistent over sites. In Trenton, the proportion of boys is slightly over 50% for both black and white; in Portland, a large percentage (58%) of the blacks are boys, whereas only 46.1% of the whites are boys; in St. Louis the sample of blacks is about 50% male, whereas the white sample is 52.3% male. In Lee County the proportion of boys is over 50% for both black and white. These differences again dictate caution in interpreting general means, for otherwise Portland would have a special advantage on variables where white girls excelled.

The "other" race category varies widely, but the cell sizes are too small to interpret.

Table 2-8

Subjects Classified by Site, Preschool Program, Race, and Sex

		HS				PS				NO KNOWN				TOTAL			
		B	W	O	T	B	W	O	T	B	W	O	T	B	W	O	T
Lee County	M	138	5	0	143	1	57	0	58	22	100	0	122	161	162	0	323
	F	94	10	0	104	7	46	2	55	17	94	0	111	118	150	2	270
	T	232	15	0	247	8	103	2	113	39	194	0	233	279	312	2	593
Portland	M	97	16	2	115	25	15	0	40	81	52	4	137	203	83	6	292
	F	83	19	2	104	18	16	0	34	46	62	4	112	147	97	6	250
	T	180	35	4	219	43	31	0	74	127	114	8	249	350	180	12	542
St. Louis	M	51	21	1	73	3	0	0	3	68	36	0	104	122	57	1	180
	F	41	19	0	60	8	1	0	9	72	32	0	104	121	52	0	173
	T	92	40	1	133	11	1	0	12	140	68	0	208	243	109	1	353
Trenton	M	62	4	0	66	23	4	0	27	68	35	3	106	153	43	3	199
	F	57	2	0	59	23	5	0	28	68	32	1	101	148	39	1	188
	T	119	6	0	125	46	9	0	55	136	67	4	207	301	82	4	387

Table 2-8 cont.

		HS				PS				NO KNOWN				TOTAL			
		B	W	O	T	B	W	O	T	B	W	O	T	B	W	O	T
TOTAL	M	348	46	3	397	56	76	0	128	239	223	7	469	639	345	10	994
	F	275	50	2	327	56	68	2	126	203	220	5	428	534	338	9	881
	T	623	96	5	724	108	144	2	254	442	443	12	897	1173	683	19	1875

Table 2-9

Percentages of Boys and Girls by Race and Sex

		Boys	Girls	N
Lee County	Black	57.7	42.3	279
	White	51.9	48.1	312
	Other	0.0	100.0	2
	TOTAL	54.5	45.5	593
Portland	Black	58.0	42.0	350
	White	46.1	53.9	180
	Other	50.0	50.0	12
	TOTAL	53.9	46.1	542
St. Louis	Black	50.2	49.8	243
	White	52.3	47.7	109
	Other	100.0	0.0	1
	TOTAL	51.0	49.0	353
Trenton	Black	50.8	49.2	301
	White	52.4	47.6	82
	Other	75.0	25.0	4
	TOTAL	51.4	48.6	387
TOTAL	Black	54.5	45.5	1173
	White	50.5	49.5	683
	Other	52.6	47.4	19
	TOTAL	53.0	47.0	1875

Race by preschool attendance classification: Table 2-10 presents the basic statistics, classified by race, for the number of children who attended Head Start or other preschool programs or were not known to have attended a preschool program. The information is separated by site. Table 2-11 contains the information in percentage form.

We first note that there are 96 white children who attended Head Start. This is about 5% of the total sample or about 14% of the whites in the sample. On the other hand, a much larger percentage (53%) of blacks in the sample attended Head Start. This racial difference is especially marked in Lee County and in Trenton where only fifteen out of 312 and six of 82 whites attended Head Start. Thus, we must consider Head Start in Lee County and Trenton essentially a black program. In Portland and in St. Louis there are, respectively, 35 and 40 white children in Head Start. This sample is substantial enough to work with for some purposes in both sites; it is a relatively large proportion in St. Louis and relatively close to what would be expected from the marginals.

All in all, it is necessary to be very careful in making overall comparisons of Head Start children with non-Head Start children, since race is disproportionately represented among these groupings.

Sex by preschool attendance classification: Table 2-12 shows the percentage of children who attended Head Start, other preschool programs, or neither. This table is classified by sex. Overall, 39.9% of the boys and 37.1% of the girls attended Head Start. There is not a consistent pattern over the four sites. In Lee County, St. Louis, and Trenton a larger percentage of boys attended, whereas in Portland a larger percentage of girls attended Head Start. In all cases the differences in proportions are slight.

Table 2-10

Number Attending Preschool Programs, Classified by Race and Site

		Black	White	Other	Total
	HS	232	15	0	247
Lee County	PS	8	103	2	113
	No Known	39	194	0	233
	Total	279	312	2	593
	HS	180	35	4	219
Portland	PS	43	31	0	74
	No Known	127	114	8	249
	Total	350	180	12	542
	HS	92	40	1	133
St. Louis	PS	11	1	0	12
	No Known	140	68	0	208
	Total	243	109	1	353
	HS	119	6	0	125
Trenton	PS	46	9	0	55
	No Known	136	67	4	207
	Total	301	82	4	387
	HS	623	96	5	724
TOTAL	PS	108	144	2	254
	No Known	442	443	12	897
	Total	1173	683	19	1875

Table 2-11

Percentages Attending Preschool Programs, Classified by Race and Site

		Black	White	Other	Total
Lee County	HS	39.1	2.5	0.0	41.7
	PS	1.3	17.4	0.3	19.1
	No Known	6.6	32.7	0.0	39.3
	Total	47.0	52.6	0.3	100.0
Portland	HS	33.2	6.5	0.7	40.4
	PS	7.9	5.7	0.0	13.7
	No Known	23.4	21.0	1.5	45.9
	Total	64.6	33.2	2.2	100.0
St. Louis	HS	26.1	11.3	0.3	37.7
	PS	3.1	0.3	0.0	3.4
	No Known	39.7	19.3	0.0	58.9
	Total	68.8	30.9	0.3	100.0
Trenton	HS	30.7	1.6	0.0	32.3
	PS	11.9	2.3	0.0	14.2
	No Known	35.1	17.3	1.0	53.5
	Total	77.8	21.2	1.0	100.0
TOTAL	HS	33.2	5.1	0.3	38.6
	PS	5.8	7.7	0.1	13.5
	No Known	23.6	23.6	0.6	47.8
	Total	62.6	36.4	1.0	100.0

Table 2-12

Percentages of Boys and Girls Attending a Preschool Program, Classified by Site

		% in HS	% in PS	% in No Known	Number
	Boys	44.3	18.0	37.8	323
Lee County	Girls	38.5	20.4	41.1	270
	TOTAL	41.7	19.1	39.3	593
	Boys	39.4	13.7	46.9	292
Portland	Girls	41.6	13.6	44.8	250
	TOTAL	40.4	13.7	45.9	542
	Boys	40.6	1.7	57.8	180
St. Louis	Girls	34.7	5.2	60.1	173
	TOTAL	37.7	3.4	58.9	353
	Boys	33.2	13.6	53.3	199
Trenton	Girls	31.4	14.9	53.7	188
	TOTAL	32.3	14.2	53.5	387
	Boys	39.9	12.9	47.2	994
TOTAL	Girls	37.1	14.3	48.6	881
	TOTAL	38.6	13.5	47.8	1875

Eligibility by preschool attendance classification: Table 2-13 shows the number of children who attended Head Start, other preschool programs, or no known preschool program, classified according to their family's eligibility under the Head Start income guidelines for varying size households. Eligibility data were obtained as part of the interview with the mother or the maternal surrogate at the testing center the spring of the Head Start year (Year 2 for Portland, St. Louis and Trenton; Year 3 for Lee County). Table 2-14 presents these same data expressed in percentages. When the respondent was unable or unwilling to provide income information, eligibility was coded as indeterminate. Missing from these tables are those initial study families who were not able to be interviewed during the Head Start year.

Seventy-five percent of the families who were eligible for Head Start did send their children to Head Start. The percent attending varied from around 58% in Trenton to nearly 89% in Lee County. This estimate is reduced to the extent that children in the no-known-preschool attendance category also attended Head Start and those in the indeterminate eligibility category were actually eligible. A review of the interviews revealed that many of the household heads in Head Start families with no income information provided held jobs that appeared unlikely to provide wages above the guidelines. About a third of the ineligible children also attended Head Start. The proportion of those ineligible who attended Head Start varied from 25% in Trenton to fully 61.3% in St. Louis. Thus there was socioeconomic diversity in the programs sampled in the study and ineligible children were not completely segregated from their more advantaged neighbors. In looking at ineligible Head Start attended percentages the reader is cautioned to remember that the families were in many different programs, and ineligible families may, therefore, be a

Table 2-13

Number of Children Attending Preschool, Classified by Eligibility and Site.

		HS	PS	NK	Total
Lee County	Eligible	152	4	15	171
	Ineligible	62	94	41	197
	Indeterminate	16	4	3	23
	TOTAL	230	102	59	391
Portland	Eligible	99	8	41	148
	Ineligible	91	50	101	242
	Indeterminate	11	4	16	31
	TOTAL	201	62	158	421
St. Louis	Eligible	88	2	20	110
	Ineligible	19	3	9	31
	Indeterminate	32	6	8	46
	TOTAL	139	11	37	187
Trenton	Eligible	67	5	43	115
	Ineligible	31	34	59	124
	Indeterminate	18	6	12	36
	TOTAL	116	45	114	275
TOTAL	Eligible	406	19	119	544
	Ineligible	203	181	210	594
	Indeterminate	77	20	39	136
	TOTAL	686	220	368	1274

Table 2-14

Percentage of Children Attending Preschool, Classified by Eligibility and Site

		HS	PS	NK	Total
Lee County	Eligible	88.9	2.3	8.8	171
	Ineligible	31.5	47.7	20.8	197
	Indeterminate	69.6	17.4	13.0	23
Portland	Eligible	66.9	5.4	27.7	148
	Ineligible	37.6	20.7	41.7	242
	Indeterminate	35.5	12.9	51.6	31
St. Louis	Eligible	80.0	1.8	18.2	110
	Ineligible	61.3	9.7	29.0	31
	Indeterminate	69.6	13.0	17.4	46
Trenton	Eligible	58.2	4.3	37.4	115
	Ineligible	25.0	27.4	47.6	124
	Indeterminate	50.0	16.7	33.3	36
TOTAL	Eligible	74.6	3.5	21.9	544
	Ineligible	34.2	30.5	35.4	594
	Indeterminate	56.6	14.7	28.7	136

smaller percentage of a particular program's enrollment. Moreover, income data were obtained in the spring of the Head Start year, whereas enrollment was in the fall. Given the greater instability of job opportunities for the poor, the line between "eligible" and "ineligible" for many of the families in this study may be fine indeed.

Socioeconomic Variables

We have selected for description in this report four variables that are components of socioeconomic status. They are mother's and father's education and mother's and father's occupation. A more fine-grained description of socioeconomic indices will be presented in the next report. We have chosen to present the mother's variables first since these are available for a substantially larger sample.

Mother's education: Data are available for mothers of 1752 of the 1875 children in the four sites. The index of mother's education used as a variable here is highest grade attended. Mean values for the different sites are shown in Table 2-15.

Mothers of children in the Portland sample have the highest average grade attended -- 11.58 -- or a half year under high school graduation. The Lee County average is 10.89, the Trenton sample 10.58 grades, and the St. Louis sample is lowest with an average of 9.59 grades. These averages and the numbers on which they are based are cross-classified by race and preschool attendance in Table 2-16.

First, we note that the mothers of children who go to other preschool programs are in all cases more highly educated than mothers of either Head Start children or of those with no known preschool program. This holds for both races and over all sites. Overall, the mothers of these children average nearly three years more education than Head Start mothers.

In general, the mothers of the white children have approximately a year and a half more education than the mothers of black children, but this pattern is not consistent throughout the sites. In Lee County, Trenton, and Portland the white mothers are better educated, but in St. Louis the mothers of the black children have, on the average, over a year more education. This change in relationship must be considered in site-to-site comparisons.

Mothers of the Head Start children have about a year less schooling than the mothers of the children in the no-known-preschool category. The difference is found to varying degrees for both races and within all of the different sites.

From the observed variation in mother's education, then, we see that the more educated mothers tend to send their children to other preschool programs and that the less well educated, both black and white, tend to send their children to Head Start. The whites in the sample are on the average slightly more educated than the blacks, except in St. Louis where the blacks are more educated.

Table 2-15

Mother's Education Classified by Site

	N	Mean	S.D.
Lee County	584	10.89	3.05
Portland	520	11.58	2.23
St. Louis	287	9.68	2.33
Trenton	361	10.58	2.09
Total	1752	10.83	2.60

Table 2-16

Average Highest Grade Attended by Mother: Classified by Site, Race and Child's Preschool Attendance

		Head Start		No Known		Preschool		Total	
		(N)	Mean	(N)	Mean	(N)	Mean	(N)	Mean
	White	15	9.53	193	11.93	101	13.68	309	12.41
Lee County	Black	230	9.21	37	8.59	8	10.63	275	9.17
	TOTAL	245	9.23	230	11.43	109	13.46	584	10.89
	White	33	11.73	111	12.14	30	13.57	174	12.30
Portland	Black	177	11.08	127	10.91	42	12.69	346	11.21
	TOTAL	210	11.18	238	11.48	72	13.06	520	11.58
	White	27	7.59	64	9.41	1	8.00	92	8.86
St. Louis	Black	68	9.84	119	10.13	8	10.88	195	10.06
	TOTAL	95	9.20	183	9.88	9	10.56	287	9.68
	White	4	8.50	67	10.87	9	13.89	80	11.09
Trenton	Black	110	10.12	131	10.40	40	11.45	281	10.44
	TOTAL	114	10.06	198	10.56	49	11.90	361	10.58
	White	79	9.73	435	11.46	141	13.63	655	11.72
TOTAL	Black	585	10.02	414	10.32	98	11.87	1097	10.30
	TOTAL	664	9.99	849	10.91	239	12.91	1752	10.83

Father's education: The information on father's education was available for 1340 of the 1875 children. The proportion of fathers for which this information is available is markedly different for blacks and whites; in the white sample, information was available for 94% as many fathers as mothers, whereas in the black sample, data were available for only 66% as many. As with mother's education, the measure of education is the highest grade attended. The mean values for different sites are shown in Table 2-17.

The average father has reached a significantly higher grade than the average mother in Lee County, a slightly higher grade in Portland and in Trenton and approximately the same grade in St. Louis.

The average highest grade attended by fathers is shown in Table 2-18, cross-classified by preschool attendance, race, and site. The overall pattern is largely the same as for mother's education.

We see that the children who attend other preschool programs have fathers who have attained a higher grade in school than either the fathers of the Head Start children or those in the no-known-preschool category. This holds true for both black and white students. The white fathers on the average have attained a higher grade than black fathers, except in St. Louis. The white fathers average a striking 5 1/2 years more education in Lee County.

Table 2-17

Father's Education Classified by Site

	N	Mean	S.D.
Lee County	489	11.67	4.75
Portland	398	11.74	2.78
St. Louis	209	9.65	2.36
Trenton	244	10.30	2.72
TOTAL	1340	11.13	3.68

Table 2-18

Average Highest Grade Attended by Father: Classified by Site, Race and Child's Preschool Attendance

		Head Start		No Known		Preschool		Total	
		(N)	Mean	(N)	Mean	(N)	Mean	(N)	Mean
Lee County	White	13	10.15	189	12.87	99	16.32	301	13.89
	Black	159	8.04	23	8.39	6	9.50	188	8.13
	TOTAL	172	8.20	212	12.39	105	15.93	489	11.67
Portland	White	28	12.89	102	12.44	25	13.88	155	12.75
	Black	109	10.84	98	10.88	36	12.47	243	11.10
	TOTAL	137	11.26	200	11.67	61	13.05	398	11.74
St. Louis	White	24	9.08	59	9.98	1	6.00	84	8.98
	Black	43	9.42	75	10.37	7	11.29	125	10.10
	TOTAL	67	9.30	134	9.76	8	10.63	209	9.65
Trenton	White	4	10.50	61	10.84	8	15.25	73	11.30
	Black	59	9.27	87	10.05	25	10.72	171	9.88
	TOTAL	63	9.35	148	10.37	33	11.82	244	10.30
TOTAL	White	69	10.91	411	11.91	133	15.72	613	12.67
	Black	370	9.22	283	10.29	74	11.53	727	9.87
	TOTAL	439	9.49	694	11.25	207	14.22	1340	11.13

Mother's occupation: Mother's occupation is coded as the three-digit code used by the Census Bureau; however, for the purposes of this report, only the first digit will be reported. An eleventh group was added to the 10 groups used by the Census Bureau to accommodate the unemployed. The coding used was:

01 Professionals

02 Farm Owners and Managers

03 Managers and Proprietors

04 Clerical and Kindred Workers

05 Sales Workers

06 Craftsmen, Foremen, Kindred Workers

07 Operatives and Kindred Workers

08 Service Workers (including private household workers)

09 Farm Laborers and Foremen

10 Laborers, Except Farm and Mine

11 Unemployed

For purposes of simplicity, we have grouped categories 1 through 5 under the general title "white collar" and categories 6 through 10 under the general category "blue collar." This rough categorization is useful for descriptive purposes; full information on the 11-category code for race x sex x site x preschool attendance is presented in Progress Report 70-20, Appendix A, and will be updated for the next report which will include a detailed report of the interview findings.

Table 2-19 summarizes the analyses of basic white-collar/blue-collar data in each site by race and by category of preschool attendance. Note that some of the cells have rather small membership and must be interpreted with care.

Table 2-19

Mother's Occupation Classified by Site, Race, and Child's Preschool Attendance

Lee County

	White HS	White NK	White PS	White T	Black HS	Black NK	Black PS	Black T	Total HS	Total NK	Total PS	Total T
White-Collar	3	44	29	76	9	1	0	10	12	45	29	86
Blue-Collar	3	35	5	43	113	16	7	136	116	51	12	179
Unemployed	9	111	66	186	105	20	1	126	114	131	67	312
Total	15	190	100	305	227	37	8	272	242	227	108	577

Portland

	White HS	White NK	White PS	White T	Black HS	Black NK	Black PS	Black T	Total HS	Total NK	Total PS	Total T
White-Collar	3	17	7	27	22	18	20	60	25	35	27	87
Blue-Collar	4	16	5	25	46	40	5	91	50	56	10	116
Unemployed	25	69	15	109	100	61	16	177	125	130	31	286
Total	32	102	27	161	168	119	41	328	200	221	68	489

St. Louis

	White HS	White NK	White PS	White T	Black HS	Black NK	Black PS	Black T	Total HS	Total NK	Total PS	Total T
White-Collar	0	8	0	8	5	8	1	14	5	16	1	22
Blue-Collar	1	9	1	11	18	41	5	64	19	50	6	75
Unemployed	20	48	0	68	41	69	2	112	61	117	2	180
Total	21	65	1	87	64	118	8	190	85	183	9	277

Trenton

	White HS	White NK	White PS	White T	Black HS	Black NK	Black PS	Black T	Total HS	Total NK	Total PS	Total T
White-Collar	0	7	7	14	6	10	15	31	6	17	22	45
Blue-Collar	0	6	1	7	24	30	17	71	24	36	18	78
Unemployed	4	52	1	57	70	80	6	156	74	132	7	213
Total	4	65	9	78	100	120	38	258	104	185	47	336

Total

	White HS	White NK	White PS	White T	Black HS	Black NK	Black PS	Black T	Total HS	Total NK	Total PS	Total T
White-Collar	6	76	37	119	42	37	36	115	48	113	73	234
Blue-Collar	8	66	12	86	201	127	34	362	209	193	46	448
Unemployed	58	280	88	426	316	230	25	571	374	510	113	997
Total	72	422	137	631	559	394	95	1048	631	816	232	1679

Table 2-20 presents a percentage summary of mother's occupation, for black and white children. The bottom margin contains the number on which the percentages were computed. We note first that a substantial proportion of mothers were not employed when these data were gathered, presumably remaining at home to care for the children. Overall, 67.5% of the white mothers stayed home as opposed to 54.5% of the black mothers. A large black-white difference in the proportion of mothers unemployed occurred in all sites. Of the white mothers who were employed, more had white-collar than blue-collar jobs.

Table 2-21 cross-classifies the occupation of the mother by the child's preschool attendance. The figures are presented separately for white and black children. The percentages add up horizontally, and the number of cases on which the percentage is based is shown in the right-hand margin. This table reflects the earlier finding that a very small percentage of the white children attended Head Start. There is a slight difference in the percentage of white children in the Head Start or no-known-preschool category between white-collar workers' children and blue-collar workers', but there is a substantially higher percentage of white-collar workers' children who attended other preschool programs. However, a substantially larger percentage of the children of unemployed mothers attended Head Start than of employed mothers. Thus, it would seem that employed white mothers did not take advantage of Head Start for their children, except in Portland, although a modest percentage of the children of unemployed white mothers did attend.

The pattern for black children is different. Overall, about 36% of the children of black mothers in white-collar jobs attended Head Start, 55.5% of blue-collar mothers' children, and 55.3% of those who were not employed.

Table 2-20

Percentage of Mothers in Occupational Group, Classified by Site and Race

		White	Black	Total
Lee County	White Collar	24.9	3.7	14.9
	Blue Collar	14.1	50.0	31.0
	Unemployed	61.0	46.3	54.1
	TOTAL	305	272	577
Portland	White Collar	16.8	18.3	17.8
	Blue Collar	15.5	27.7	23.7
	Unemployed	67.7	54.0	58.5
	TOTAL	161	328	489
St. Louis	White Collar	9.2	7.4	7.9
	Blue Collar	12.6	36.7	27.1
	Unemployed	71.2	58.9	65.0
	TOTAL	87	190	277
Trenton	White Collar	17.9	12.0	11.6
	Blue Collar	9.0	27.5	23.2
	Unemployed	73.1	60.5	65.2
	TOTAL	78	258	336
Total	White Collar	18.9	11.0	13.9
	Blue Collar	13.6	35.5	26.7
	Unemployed	67.5	54.5	59.4
	TOTAL	631	1048	1679

Table 2-21

Percentage of Mother's Occupation Group Classified by Site, Race, and Child's Preschool Attendance

	White				Black			
	HS	NK	PS	T	HS	NK	PS	T
Lee County								
White-Collar	3.9	57.9	38.2	76	90.0	10.0	0.0	10
Blue-Collar	7.0	81.4	11.6	43	83.1	11.8	5.1	136
Unemployed	4.8	59.7	35.5	186	83.5	15.9	.8	126
Total	4.9	62.3	32.8	305	83.5	13.6	2.9	272
Portland								
White-Collar	11.1	63.0	25.9	27	36.7	30.0	33.3	60
Blue-Collar	16.0	64.0	20.0	25	50.5	44.0	5.5	91
Unemployed	22.9	63.3	13.8	109	56.5	34.5	9.0	177
Total	19.9	63.3	16.8	161	51.2	36.3	12.5	328
St. Louis								
White-Collar	0.0	100.0	0.0	8	30.8	61.5	7.7	14
Blue-Collar	9.1	81.8	9.1	11	28.8	65.4	5.8	64
Unemployed	29.4	70.6	0.0	68	36.6	61.6	1.8	112
Total	24.1	74.7	1.2	87	33.7	62.1	4.2	190
Trenton								
White-Collar	0.0	50.0	50.0	14	19.4	32.2	48.4	31
Blue-Collar	0.0	85.7	14.3	7	33.8	42.2	24.0	71
Unemployed	7.0	91.2	1.8	57	44.9	51.3	3.8	156
Total	5.1	83.3	11.6	78	38.7	46.5	14.8	258
Total								
White-Collar	5.0	63.9	31.1	119	36.5	32.2	31.3	115
Blue-Collar	9.3	76.7	14.0	86	55.5	35.1	9.4	362
Unemployed	13.6	65.7	20.7	426	55.3	40.3	4.4	571
Total	11.4	66.9	21.7	631	53.3	37.6	9.1	1048

Employed mothers, whether in white-collar or blue-collar occupations, were more likely to send their children to Head Start than to other preschool programs. Although a substantial number of black children of mothers with white-collar occupations were enrolled in other preschool programs, very few of the unemployed mothers sent their children to other preschool programs. All in all, the differences in Head Start attendance seem to be related to racial differences.

Father's occupation: The occupations of fathers were classified into 10 groups using the Census Bureau categories. We have added an eleventh category for the unemployed. The classifications are:

01 Professionals
02 Farm Owners and Managers
03 Managers and Proprietors
04 Clerical and Kindred Workers
05 Sales Workers
06 Craftsmen, Foremen, Kindred Workers
07 Operatives and Kindred Workers
08 Service Workers (including private household workers)
09 Farm Laborers and Foremen
10 Laborers, Except Farm and Mine
11 Unemployed

Complete data on father's occupation for race x sex x site x preschool attendance are presented in Progress Report 70-20, Appendix A.

We have again for simplicity grouped categories 1 to 5 as white-collar and 6 to 10 as blue-collar. These data are shown in Table 2-22, analyzed by race and category of preschool attendance, separately for each site. Compared to mothers (N=1679), this information was available for only 1293 fathers.

Table 2-22

Father's Occupation Classified by Site, Race, and Child's Preschool Attendance

Lee County

	White				Black				Total			
	HS	NK	PS	T	HS	NK	PS	T	HS	NK	PS	T
White-Collar	2	86	83	171	7	0	2	9	8	86	85	179
Blue-Collar	11	91	17	119	139	25	4	168	150	116	21	288
Unemployed	0	11	0	11	8	0	0	8	9	11	0	19
Total	13	188	100	301	154	25	6	185	167	213	106	486

Portland

	White				Black				Total			
	HS	NK	PS	T	HS	NK	PS	T	HS	NK	PS	T
White-Collar	9	47	14	70	15	18	12	45	24	65	26	115
Blue-Collar	14	54	10	78	80	64	21	165	94	118	31	243
Unemployed	3	3	0	6	8	10	2	20	11	13	2	26
Total	26	104	24	154	103	92	35	230	129	196	59	384

St. Louis

	White				Black				Total			
	HS	NK	PS	T	HS	NK	PS	T	HS	NK	PS	T
White-Collar	1	6	0	7	6	4	9	10	7	10	0	17
Blue-Collar	15	53	0	68	25	53	5	83	40	106	5	151
Unemployed	3	1	1	5	7	11	1	19	10	12	2	24
Total	19	60	1	80	38	68	6	112	57	128	7	192

Trenton

	White				Black				Total			
	HS	NK	PS	T	HS	NK	PS	T	HS	NK	PS	T
White-Collar	1	22	5	28	3	5	3	11	4	27	8	39
Blue-Collar	3	37	3	43	47	70	17	134	50	107	20	177
Unemployed	0	0	0	0	3	10	2	15	3	10	2	15
Total	4	59	8	71	53	85	22	160	57	144	30	231

Total

	White				Black				Total			
	HS	NK	PS	T	HS	NK	PS	T	HS	NK	PS	T
White-Collar	13	161	102	276	31	27	17	75	44	188	119	351
Blue-Collar	43	235	30	308	291	212	47	550	334	447	77	858
Unemployed	6	15	1	22	26	31	5	62	32	46	6	84
Total	62	411	133	606	348	270	69	687	410	681	202	1293

There are many rather small cells which are difficult to interpret.

Table 2-23 presents percentages of white-collar, blue-collar, and unemployed fathers, separated by race and by site. The number of persons on whom the percentages were based is shown as a lower margin of each table.

We first note that a substantially larger proportion of the black children had unemployed fathers. The proportion was 9% overall for blacks and 3.6% for whites. The finding of a substantially larger percentage of unemployed fathers of black children was consistent from site to site. Of the employed fathers, there was a larger proportion of blue-collar than white-collar workers for both races and in all sites, but the total of blue-collar employees outnumbers white-collar employees about 10 to 9 among the whites and about 7 to 1 among the blacks (see Table 2-23). St. Louis was an exception where there was a larger tendency for the fathers of white children to be employed in blue-collar occupations than for the fathers of black children.

Table 2-24 presents the proportion of fathers in each type of occupation whose children attended Head Start, other preschool programs, or no-known-preschool program. This information is displayed separately by race. The right-hand margin of each table shows the numbers from which the percentages were computed.

The number of unemployed white fathers was only 22, so we shall not discuss percentages based on such a small sample. There was a differential pattern for white-collar and blue-collar fathers in sending their children to preschool programs, with white-collar workers' children more likely to have attended other preschool programs and blue-collar workers' children to have attended Head Start programs. The distribution of fathers in white-collar and blue-collar jobs differentiated similarly among blacks, except in St. Louis

Table 2-23

Percentage of Fathers in Occupational Group Classified by Site and Race

		White	Black	Total
Lee County	White Collar	56.8	4.9	37.0
	Blue Collar	39.3	90.8	59.1
	Unemployed	3.7	4.3	3.9
	TOTAL	301	185	486
Portland	White Collar	45.5	19.6	29.9
	Blue Collar	50.6	71.6	63.3
	Unemployed	3.9	8.7	6.8
	TOTAL	154	230	384
St. Louis	White Collar	8.7	8.9	8.9
	Blue Collar	85.0	74.1	78.6
	Unemployed	6.3	17.0	12.5
	TOTAL	80	112	192
Trenton	White Collar	39.4	6.9	16.9
	Blue Collar	60.6	83.8	76.6
	Unemployed	0	9.4	6.5
	TOTAL	71	160	231
Total	White Collar	45.5	10.9	27.7
	Blue Collar	50.8	80.1	66.1
	Unemployed	3.6	9.0	6.2
	TOTAL	606	687	1293

Table 2-24

Percentage of Father's Occupation Group Classified by Site, Race, and Child's Preschool Attendance

	White				Black			
	HS	NK	PS	T	HS	NK	PS	T
Lee County								
White-Collar	1.2	50.3	48.5	171	77.8	0.0	22.2	9
Blue-Collar	9.2	76.5	14.3	119	82.7	14.9	2.4	168
Unemployed	0.0	100.0	0.0	11	100.0	0.0	0.0	8
Total	4.3	62.5	33.2	301	83.2	13.5	3.2	185
Portland								
White-Collar	12.9	67.1	20.0	70	33.3	40.0	26.7	45
Blue-Collar	17.9	69.2	12.8	78	48.5	38.8	12.7	165
Unemployed	50.0	50.0	0.0	6	40.0	50.0	10.0	20
Total	16.9	67.5	15.6	154	44.8	40.0	15.2	230
St. Louis								
White-Collar	14.3	85.7	0.0	7	60.0	40.0	0.0	10
Blue-Collar	22.1	77.9	0.0	68	30.1	63.9	6.0	83
Unemployed	60.0	20.0	20.0	5	43.8	50.0	6.3	19
Total	23.8	75.0	1.2	80	36.8	57.9	5.3	112
Trenton								
White-Collar	3.6	78.6	17.9	28	27.3	45.5	27.3	11
Blue-Collar	7.0	86.0	7.0	43	35.1	52.2	12.7	134
Unemployed	0.0	0.0	0.0	0	20.0	66.7	13.3	15
Total	5.6	83.1	11.3	71	33.1	53.1	13.8	160
Total								
White-Collar	4.7	58.3	37.0	276	41.3	36.0	22.7	75
Blue-Collar	14.0	76.3	9.7	308	52.9	38.3	8.5	550
Unemployed	27.3	68.2	4.5	22	41.9	50.0	8.1	62
Total	10.2	67.8	21.9	606	50.7	39.3	10.0	687

where a disproportionate number of children of black white-collar fathers attended Head Start and children of black blue-collar fathers attended no known preschool program. And, except in Trenton, there was a tendency for the children of white-collar black fathers to attend Head Start rather than other preschool programs.

Age at time of testing: A description of the age of the children at the time they were tested is complicated by the fact that some children were tested over a several-month period. This happened because children who missed some of the week of testing were followed up and brought back whenever possible to the testing center for further testing. In the ordinary routine, children were given a common battery of instruments on their first day and then took three batteries during the rest of the week. For simplicity, we have selected at random one test from each battery and computed the mean age of the children at the time of testing, classified by preschool attendance. These data, separated by site, are shown in Table 2-25.

The table contains two entries in each cell: the number of children in that cell and their average age in months. One pattern shows up quite strongly: the children in St. Louis were on the average about two-and-a-half months older when they were tested than were the children in other sites. As discussed in the next chapter, it was necessary to begin testing later and also to extend testing by about three months in St. Louis in order to increase the sample size in that site. We note that these children are still of the appropriate age, but the age at the preliminary testing was about two-and-a-half months older.

There is also a very slight tendency for children enrolled in some preschool program (Head Start or other) to be slightly older than those in the no-known-preschool category.

Table 2-25

Average Age (in months) at Time of Testing
Classified by Site and Child's Preschool Attendance

Motor Inhibition Test - Day 1 Battery

	Head Start		No Known		Other Preschool		Total	
	N	Mean	N	Mean	N	Mean	N	Mean
Lee County	237	50.73	162	51.12	104	51.91	503	51.10
Portland	210	50.92	213	50.86	68	50.60	491	50.85
St. Louis	109	53.11	107	52.96	7	55.14	223	53.10
Trenton	118	51.28	177	50.04	51	51.27	346	50.64
TOTAL	674	51.27	659	51.05	230	51.48	1563	51.21

Preschool Inventory (Caldwell) - Battery A

	Head Start		No Known		Other Preschool		Total	
	N	Mean	N	Mean	N	Mean	N	Mean
Lee County	229	50.83	154	51.09	104	52.06	487	51.18
Portland	209	51.00	213	50.87	70	50.56	492	50.88
St. Louis	109	53.13	103	53.13	6	55.17	218	53.18
Trenton	116	51.22	174	49.93	51	51.25	341	50.56
TOTAL	663	51.33	644	51.03	231	51.51	1538	51.23

ETS Story Sequence - Battery B

	Head Start		No Known		Other Preschool		Total	
	N	Mean	N	Mean	N	Mean	N	Mean
Lee County	236	50.78	158	51.18	104	52.05	498	51.17
Portland	210	51.02	213	50.83	70	50.57	493	50.89
St. Louis	104	53.22	103	53.15	6	55.17	213	53.24
Trenton	115	51.37	168	49.99	50	51.18	333	50.65
TOTAL	665	51.34	642	51.08	230	51.49	1537	51.25

Boy-Girl Identity Task - Battery C

	Head Start		No Known		Other Preschool		Total	
	N	Mean	N	Mean	N	Mean	N	Mean
Lee County	226	50.91	154	51.15	103	52.01	483	51.22
Portland	186	51.21	186	51.09	63	50.70	435	51.09
St. Louis	102	53.31	98	53.16	7	55.29	207	53.31
Trenton	115	51.52	174	50.01	50	51.22	339	50.70
TOTAL	629	51.50	612	51.13	223	51.57	1464	51.35

CHAPTER 3--METHODOLOGY

Collection of Data*

Enumeration and Parent Interviews

The first phase of data collection, enumeration and parent interviews, was undertaken by the New York City firm of Audits and Surveys (A & S), under subcontract with ETS. Its 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. An eligible child was one who, on the basis of his birthdate, was expected to enter first grade in the Fall of 1971.

Since previous experience with similar surveys had demonstrated the importance of community support, cooperation through the use of local media and through contact with key community leaders was effectively sought. Interviewers, all female, were recruited from the community, with A & S staff responsible for both training and supervision.

During the enumeration phase, several problems were encountered. One of the most difficult involved development of individual location maps to monitor interviewer assignments. This was particularly difficult in rural areas of Lee County because frequently there were no named streets or official county roads. The problem was finally resolved by hiring several local long-term residents who traveled through the county making detailed maps of each school district. The problem of locating the expected number of Households was not

*See ETS, PR-69-12, "From Theory to Operations," for a more detailed accounting of Year 1 data-collection procedures.

unique to the rural areas of Lee County. In St. Louis, for example, it was found that many of the neighborhoods in the study have houses with entrances in alleyways that do not appear on official maps. Here, too, the solution involved reliance on the knowledge and cooperation of local residents. As a cross-check, to ensure that as few eligible households as possible were missed during pre-listing, a question about first-grade enrollment was used. However, unexpected variations in local enrollment practices did cause problems, several of which are discussed in ETS Progress Report 70-20.

Following initial piloting in the metropolitan New York area, a full-scale pilot test of about 10 completed interviews was conducted in each of the four study sites. The interviewing procedures paralleled the final design and execution to as great an extent as possible. Three interviewers in each city underwent an extensive briefing in order to conduct the pilot test. All three completed practice interviews and later had the opportunity to discuss their reactions and opinions at a group debriefing session. The debriefing report, supported by tape recordings of the discussions and independent analysis of the pilot-test questionnaires, proved to be extremely useful in the final revision of both questionnaire and training procedures.

Since changes in the interview involved only deleting or rewording a few ambiguously worded questions, or modifying format rather than the nature of an item, another pilot testing proved unnecessary. The actual interviewing of eligible mothers or mother substitutes went relatively smoothly, and each one was reviewed on a question-by-question basis for consistency, clarity, and completeness.

Individual Testing

Phase two of data collection involved administration of individual tests. From the beginning of the study it had been argued that using local testers would facilitate community cooperation, contribute to the validity of the data obtained, and provide training that would contribute to future employment possibilities for community residents.

The general procedures were the same in each site. Prior to the arrival of the ETS training team, the local coordinator preselected the tester trainees, choosing approximately 30% more than the number who eventually would be hired. Depending on a variety of factors (such as resources in the community, the local coordinator's preferences, publicity concerning the project, and intra-community relations), trainees varied both within and between sites. All trainees were female. The usual educational credentials were not required, but experience in working with young children was considered highly desirable, as was the ability to read and speak with ease. Our judgments as to the adequacy of the tester's affective reactions to children and her ability to learn the tasks were the two focal criteria for final selection. Most of the trainees were housewives who had limited work experience, and most were black.

The on-site training was undertaken at staggered intervals, starting March 17 in Auburn, March 31 in Portland, April 14 in Trenton, and April 28 in St. Louis Training at each site during the first two weeks took place in the local coordinator's office. After receiving a general orientation, trainees began practice on one of the simpler tasks on the first day. It was felt that facility in handling the variety of problems a tester was likely to encounter could best be developed in the context of a particular test. These general procedures were then repeated more meaningfully in the context of other tasks. As in training trainers, the tasks were first demonstrated, and then the trainees practiced by administering them to each other.

The first tasks demonstrated were those in the Day 1 battery. To reduce the number of tasks that she would be required to learn, each trainee was assigned to learn one of the three remaining batteries. Each task was demonstrated, and trainees then practiced administering it to each other and to children volunteered by other trainees, their friends and neighbors. Videotapes of the trainees administering tests and brief tests to assess the trainee's knowledge of the test in the battery were also used.

During the third week trainees moved to the actual testing centers. An ETS staff trainer was assigned to each center to ensure adequacy of physical arrangements and testing supplies, and to function temporarily as a center supervisor so that trainees could concentrate on improving their testing skills. The local coordinator arranged for practice subjects who would be comparable to sample subjects and provided for their transportation to and from the center. During the fourth (and sometimes fifth) week of testing practice, the trainees were observed by ETS staff--in all cases this included the project director and a senior member of the professional research team--in order to evaluate performance and to select those women who seemed best prepared to be center supervisor, tester, or play-area supervisor. In those cases where an individual was not selected, every attempt was made to structure the situation as a growth experience instead of a failure and to maintain the person's interest and involvement in the study.

Once evaluations were completed, each center operated one or two weeks more for a dry run. A Princeton Office trainer remained at each center to provide general assistance and additional instruction in testing while the center staff practiced their new roles. Once actual testing began, monitoring of center operations (except at Trenton) was assumed by ETS regional office personnel with the assistance of Princeton Office staff; Princeton Office staff monitored Trenton operations.

As in training interviewers, piloting of procedures was an essential part of the training process. Prior to initial selection, each measure had been administered to children similar in age and socioeconomic level. None, however, had been given by indigenous testers; typically, a research assistant or graduate student under the supervision of an ETS researcher had administered the tasks. Although considerable rewriting of test manuals and changing in test format to facilitate the handling of testing materials had taken place both before and during the training of tester-trainers, refinement of these procedures awaited piloting in the field. The first two sites (Lee County and Portland) were therefore used for continued simplification and clarification of testing and scoring procedures based on trainer experience and trainee suggestions.

Similarly, the pilot batteries for each of the four days 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 something to take home (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 difficulty of test administration 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 scoring systems. From such cooperative efforts were derived not only more adequate measurement techniques, but also valuable community-based feedback on research procedures. (Table 3-1 shows the final order of the tests in the batteries.)

Testing centers were located in churches or community recreation facilities in or near the districts where the children lived. Each center provided, at a minimum, six individual testing rooms or partitioned spaces and a larger play and rest area; most also included kitchen facilities. Each testing unit, 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, and the fifth day scheduled for makeups. A rigid schedule was not always possible nor 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 sample school districts within a community; and in the testing situations, testers were instructed to wait until the children were ready, with breaks taken whenever necessary.

Table 3-1

The Measures and Testing Sequence Used in the Initial Assessments

Day 1	Av. Time in minutes
First-Day-of-School Question	
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 (WPSSI)	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 I	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

The first longitudinal sample children were tested seven to eight weeks after the beginning of tester training. During the actual testing, the center staffs worked independently except for periodic visits by monitors who were responsible for providing general advice on both testing and administrative problems to the center staff and to the local coordinator, and for observations to determine whether standard testing procedures were being followed.

Despite initial predictions that all testing would be completed by early July, centers continued in operation throughout the summer in an attempt to test the desired number of children. Several factors contributed to delays. In some cases there were failures to obtain, at an adequate rate, the names of families interviewed, and in several sites there was some reluctance of parents to allow their children to participate. Increased project publicity and personal visits by the local coordinator and testing staff helped to combat the latter problem. Also, there was a greater turnover in testing staff than had been anticipated because of the temporary nature of the job, because of previous summer or other domestic commitments, and also due to various private emergencies which arose more frequently since many of our testers lacked personal support and back-up resources. The high turnover rate made it necessary to continue training activities throughout the summer, although actual training time was shortened, since the trainee could obtain more individual attention and the trainer could share his duties with regional office and local center staff.

Because children were still being tested at the end of August, particularly in St. Louis and Trenton, and it was necessary to have those children who would attend Head Start tested before they were exposed to the program, we did the following:

1. Head Start advance registration lists for all centers within the study districts were obtained and checked against the names of children already tested; those not yet tested were scheduled for testing as quickly as possible.

2. At the opening of Head Start, center directors provided each Head Start teacher with a list of all children who had been tested in the district, and provisions were made to have any untested children sent directly to the nearest testing center before they participated in the Head Start program. Assistance from the national Head Start Research Office was very important for this phase.

3. The testing of any "left over" non-Head Start children was completed in September after the last Head Start children had been tested.

We should stress again that these extraordinary efforts were mainly relevant to Trenton and St. Louis, although we also extended testing time in Portland and Lee County to obtain as complete samples there as possible.

Medical Histories and Examinations

The third phase of data collection involved medical histories and examinations. As is true for other aspects of the study, there were regional variations in the procedures for conducting the medical examination. In St. Louis, a Neighborhood Health Center was contracted to do the examinations. In Portland and Trenton, a single physician examined all the study children. Distances in Lee County made it impossible to concentrate the medical examinations in one location, so three physicians covered the children in their respective areas. Examinations were scheduled routinely following completion of the testing cycle.

Processing of Data

The various processing operations required for the Year I data included scoring and coding of the raw data, the construction and maintenance of the data base, and the design, programming, and execution of the various internal and cross-domain analyses. Many of the analyses described will not be discussed in this report. Some of these, the initial descriptive analyses of instruments, were reported in Progress Report 70-20; 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 (e.g., in the technical reports of the individual measures). A detailed account of the design of the data base was also presented earlier (PR-70-20); therefore, many of its aspects will not be included in this report.

Coding

All data were scored by several raters to establish reliability and, following resolution of scorer differences, double-coded at the Princeton Office. Each answer sheet was checked for tester error in administration (e.g., allowing the mother to be present, or interruptions on the Fixation Test within a sequence) or recording (e.g., not rounding to .2 second on timed tasks or not circling the final response to an initial multiple response) or 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 subsequent clues to understanding the processes involved.

Preparation of the Data Base

To permit flexible and economical retrieval of the data for present and future analyses, it was necessary to create a merged tape file that contained all the derived instrument scores and classification variables such as race, sex, preschool experience, and age in one contiguous information block for a given child. Continuing development and maintenance of this comprehensive, accessible, correctible data base was a major component of the analysis system.

It was necessary to create programs that would up-date, add or delete entries into the merged file. This programming was accomplished by using the building blocks of the F4STAT Statistical System (see Appendix A) to produce a flexible set of programs. In a system of this type there is the problem of tracking variables through the file updates, and of assuring that the information retrieval can be accomplished with minimum effort. To accomplish this tracking, subroutines were built into the merge update programs which concurrently update a catalog of scores on a separate disc file. This catalog contains all the pointers (locations) of the variables contained in the merged file, as well as their related headings and titles necessary to properly label the analysis output. The catalog also provides an up-to-date listing of all scores that are available; thus, a researcher could use it in selecting scores for analysis. As is customary, precautions were taken at every step to prevent accidental deletion or loss of any data on the tape. Back-up tapes were created at all critical points of up-dating, assuring rapid and complete recovery from any type of computer or programming error at any point in the process.

The score retrieval programs were created in FORTRAN Subroutine form, using the basic Input/Output routines available in the F4STAT computer statistics system. In this way a programmer is able to retrieve any set of variables for any sample on the merge file by first passing a list of the variable numbers to the first of two retrieval routines. The relative locations of the selected variables in the child's block of information would now be available to the programmer since they are returned by this access routine. The programmer could then access the actual variables from the merge tape by passing this list of pointers to the second retrieval routine. This routine would actually read the merged tape, and, using the list of pointers passed to it, would then extract the requested variables for use in analysis. For any given observation, which the programmer could select by querying any of the classification variables, the program could now decide whether this set of variables was a member of the sample needed. This system was developed in such a manner that the programmer need not be knowledgable of the actual format of the tape but only need concern himself with the list of variables he would like to select for his particular application. The access routines retrieve not only the data variables but also their mnemonic coded headings and titles to be used in labeling of output for easier interpretation by researchers.

In all file maintenance and analysis runs, a child's test data must be matched to his master-file data. The master-file data provides the necessary identification checks and information on the age at time of testing which must be computed for each instrument since the date of testing varies among instruments. It also includes information on sex, race, site, and preschool experience. A subprogram used by the maintenance routines and the analysis routine performs this function. The ability of both the maintenance and the analysis programs to use this common subprogram not only saves programming and testing time, but it also insures that the data used, at all stages of the analysis, are "clean" data.

It was stated earlier that each instrument was essentially an independent set of data. This fact was a major obstacle in the design of a generalized primary analysis program that could be used for all instruments (excluding the questionnaire type of instrument), since every instrument had a different decoding scheme. A further complication was that many scores had to be computed by a logical sequence involving many pieces of information in the child's record. It was decided to use a method developed at Educational Testing Service, involving the programming of a unique decoding subprogram for each instrument. Its function is to decode the child's record and create derived scores for the instrument. In this way a generalized analysis program can be designed and tables for any instrument can be computed by providing the correct input control cards, label cards, and decoding subprogram. A time-saving feature of this method was that the programming and testing of the primary analysis program could be accomplished while the decoding subprograms were being independently assembled and tested. This method of decoding the instruments has proven to be successful in this study, as it has been in earlier applications.

Analyses of Individual Instruments

The initial program written for descriptive analysis of a given instrument computed and printed for each site and for the four sites combined two factorially constructed tables containing descriptive statistics on the derived scores of each instrument. The first statistical table presented data by age at time of testing, by sex, and by race, with age subdivided into six three-month intervals starting with 42-44 months and ending at 57-59 months. Race was divided into two categories, black and white. (The races classified under "Other" in the sample were excluded from these analyses because of a paucity of data.) The second table presented data by preschool experience, by sex, and by race. This table did not include the Lee County data since Head Start information was not available at the time. Preschool experience was divided into three categories: Head Start, no known preschool, and other preschool experience. The tables were completely cross-classified with a Total row computed by collapsing all the cells into it. For each cell the information included the number of observations, mean, standard deviation, minimum score, maximum score, and a percent response for each possible score category. The percent-response option had a cell count separate from the count used for the mean. This occurred because the percentage of tester errors and the percentage of refusals were computed and printed in the percent-response part of the table, but these cases were excluded from the cell mean. When percentiles were used, the separate cell size was not printed since only the scores used in computing the mean were used to compute the percentiles. Tester errors and refusals were excluded from the percentiles as well as from the mean. When the percentile option was used and the number of observations in the cell was less than six, the printed output consisted of asterisks.

"Among" and "within" statistics were provided at the bottom of each table to enable researchers to perform "a posteriori" tests on the data. The square of the statistic printed under the standard deviation column in the "among" row was the mean square associated with all the non-zero cells in the table. The squared "within" statistic was the estimate of the within-cell variance computed by pooling the variances within all the non-zero cells. Race, sex, age at time of testing, and preschool experience marginal cells were also provided by this program.

The statistics for the Child Health Record and Parent Interview were prepared somewhat differently from those of the other instruments. Here questionnaire distributions were run, consisting of counts and percent responding for each response of every item. This information was provided by sex, by race, and by preschool experience, both across all sites and within each site. A Chi-square statistic was provided for all items for which the questionnaire had several categories (such as male, female). The Chi-square computation did not include the "No Response" category provided on every item. If a predicted cell size was smaller than five, the Chi-square statistic was flagged with the letters (NV) indicating that the statistic may not be valid. Items considered to be of a continuous nature were excluded from the questionnaire distributions, and separate frequency distributions were provided for each of these items using the same site, sex, race, and preschool categories as for the questionnaire items. The frequency distribution output also provided other useful information, such as the mean, standard deviation, minimum value, maximum value, sum of scores, sum of squared scores, percent below intervals, and an analysis of variance table for comparison of the categories involved.

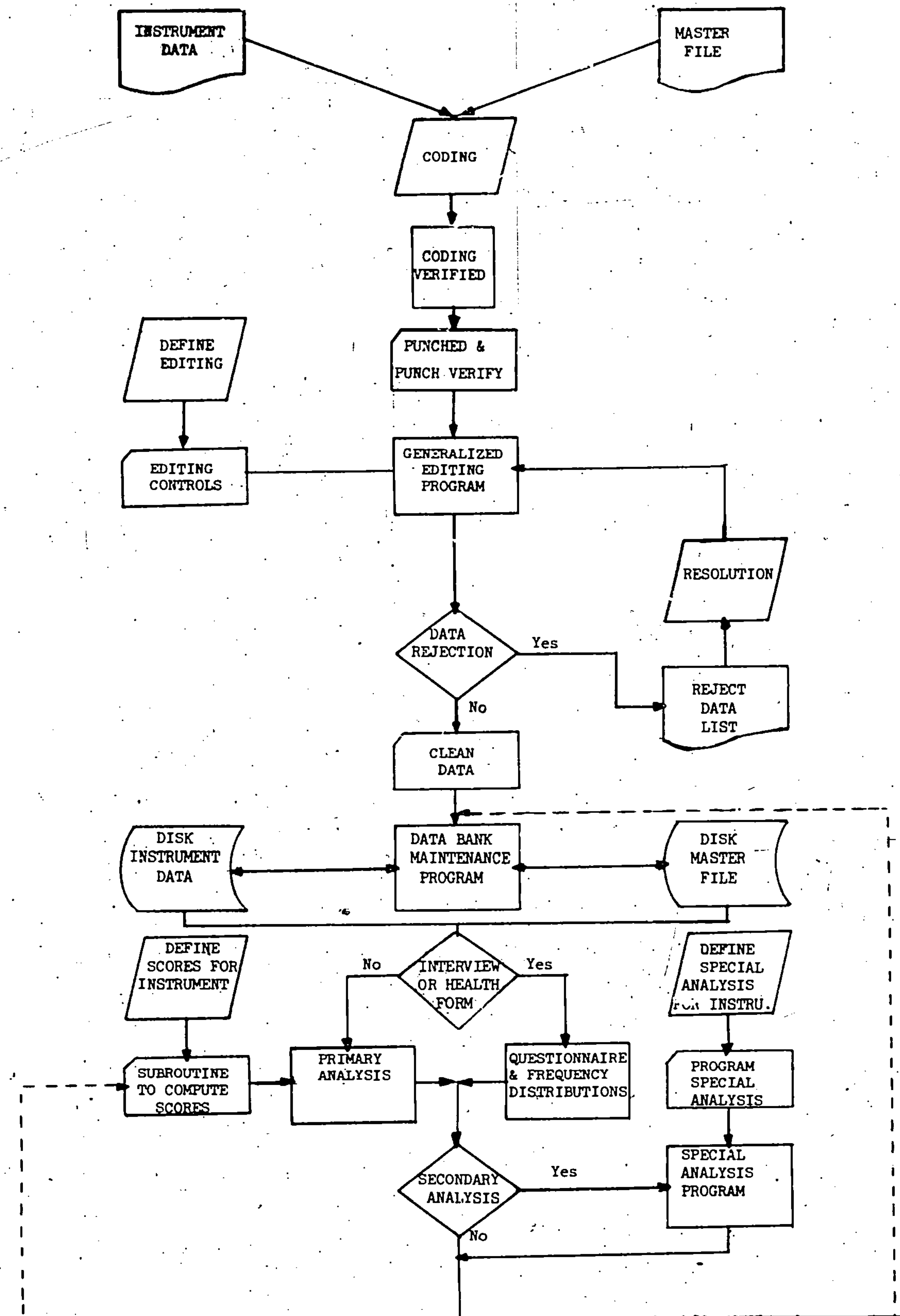
As shown in figure 3-1, special or secondary analyses have been run on all of the instruments involved in this report. Since the various instruments differ widely in content as well as in style and presentation, a wide variety of internal analyses were required. For all scores that were composites derived from right-wrong type of items, tables of item difficulty, biserial correlations of the items with the score, and KR-21 reliability coefficients were computed and printed. For other types of composite scores the alpha coefficient of reliability (KR-20) was computed. The alpha provides a lower bound for the true reliability of the composite score. Other secondary analyses were designed by researchers responsible for particular instruments. Used in these analyses were 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 used with 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.

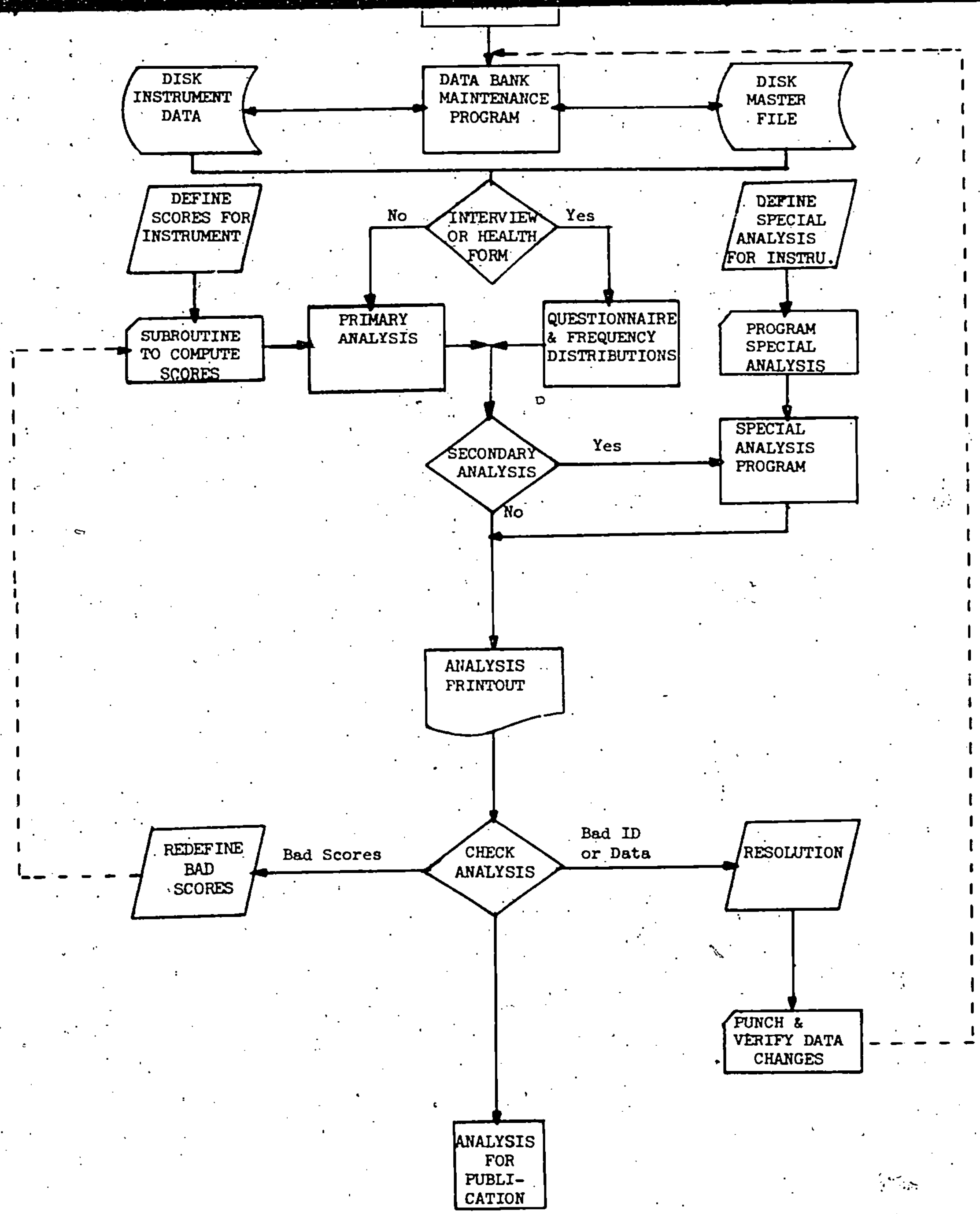
In this analysis program--as well as in the file maintenance program--label checks, data checks, variable checks, program checks, and input control-card checks were all carefully planned to prevent any possibility of incorrect use of any data, labels, or programs in a given computer run.

Structural Analyses

Missing data Pearson product-moment correlation tables were constructed for every variable that was placed on the merged file. Correlations were

Data Flow and Analysis System




DATA BANK MAINTENANCE PROGRAM
DISK INSTRUMENT DATA
DISK MASTER FILE
DEFINE SCORES FOR INSTRUMENT
INTERVIEW OR HEALTH FORM
No
Yes
DEFINE SPECIAL ANALYSIS FOR INSTRU.
SUBROUTINE TO COMPUTE SCORES
PRIMARY ANALYSIS
QUESTIONNAIRE & FREQUENCY DISTRIBUTIONS
PROGRAM SPECIAL ANALYSIS
SECONDARY ANALYSIS
Yes
SPECIAL ANALYSIS PROGRAM
No
ANALYSIS PRINTOUT
REDEFINE BAD SCORES
Bad Scores
CHECK ANALYSIS
Bad ID or Data
RESOLUTION
PUNCH & VERIFY DATA CHANGES
ANALYSIS FOR PUBLICATION

71

run for the total sample, and for the critical breakdowns such as sex, preschool experience, SES level, site, and for subclassifications based upon interactions among these main level factors.

The structural analyses run for the first year data analysis involved factor analysis techniques and the Guttman-Lingoes Smallest Space Analysis. Factor analysis is used most widely as a method for summarizing intercorrelations among large numbers of logically distinct scales in an attempt to infer underlying precursors or determinants of manifest test scores. The smallest space analysis is similar to factor analysis except for fewer parametric assumptions. It has been used in this study as a check on composites isolated through factor analysis, to insure that conclusions would not be based on results which were dependent on the method of analysis.

Before using either of these 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 this way, 46 variables were identified and placed into the structural analyses with an additional set of 5 variables placed into extension with the main set of variables.

The factor analyses were computed twice, first placing 1's in the diagonal of the correlation matrix (principal components analysis), and secondly placing an estimate of communalities into the diagonal. Communalities were estimated by Tucker's Adjusted Highest Off-Diagonal element procedure, which is explained in Appendix A. In both cases the initial factor loadings were rotated by varimax for 2 to 5 factors and then placed into promax oblique rotation using the same range of factors. Extension variables were also carried along during these steps to study their relations with the factors derived from the main set of variables. Again, all of these steps were conducted for breakdowns such as sex, age, SES level and preschool attendance controlled for Head Start eligibility. Six and seven factor varimax and promax solutions were also obtained for the composite sample and for narrower age groupings.

For every factor analysis, a parallel analysis was computed using the Guttman-Lingoes Smallest Space program. The Guttman-Lingoes program represents a non-metric technique for finding the smallest euclidean space for a configuration of points. To quote from the authors' description: "Briefly stated the problem posed for the program is: given a matrix of inequalities among pairs of points in a metric or nonmetric space, determine a set of euclidean coordinates such that the distances calculated from them are a monotonic function of the ranks or order among the inequalities" (Lingoes, 1965).

Unlike factor analysis techniques, this analytic procedure is sensitive to direction of scoring a variable. Therefore, before any of these smallest space analyses were computed, the algebraic signs of error scores were changed so that wherever it was clear that a variable reflected level of performance, high scores would indicate better performance.

After analyzing the results of the initial phases of the analysis it was decided to extract the variables that loaded heavily on the first factor and submit them to a separate analysis in an attempt to further separate them into subfactors. Both the factor analytic technique and the Guttman-Lingoes program were used in this secondary analysis. It was also observed that the Preschool

Inventory total score was the single most important component of the first factor. This led to a more complex analysis of this total score. An analysis of variance technique was used on the items of the Preschool Inventory in an item by sex by SES level ANOVA.

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. Preschool attendance and Head Start eligibility served as independent variables in the second set of ANOVAS. The results of these analyses of variance and the results of all the above analyses are reported in the succeeding sections of the report.

The importance of the merged tape for any type of analysis that will be done in the future can now readily be seen to be substantial. We have essentially reduced our data base from a collection of over 100,000 card images separated by instruments to one all-encompassing data file which contains approximately 2000 records or blocks of information, with each block of information containing all the information about a child. Equally important is the ease with which a programmer can access the file to perform an analysis. Illustrative information retrieval system flow charts are included in this section (Figures 3-2 and 3-3) in an attempt to provide an overview of the procedure.

Figure 3-2

Merged Tape Update System

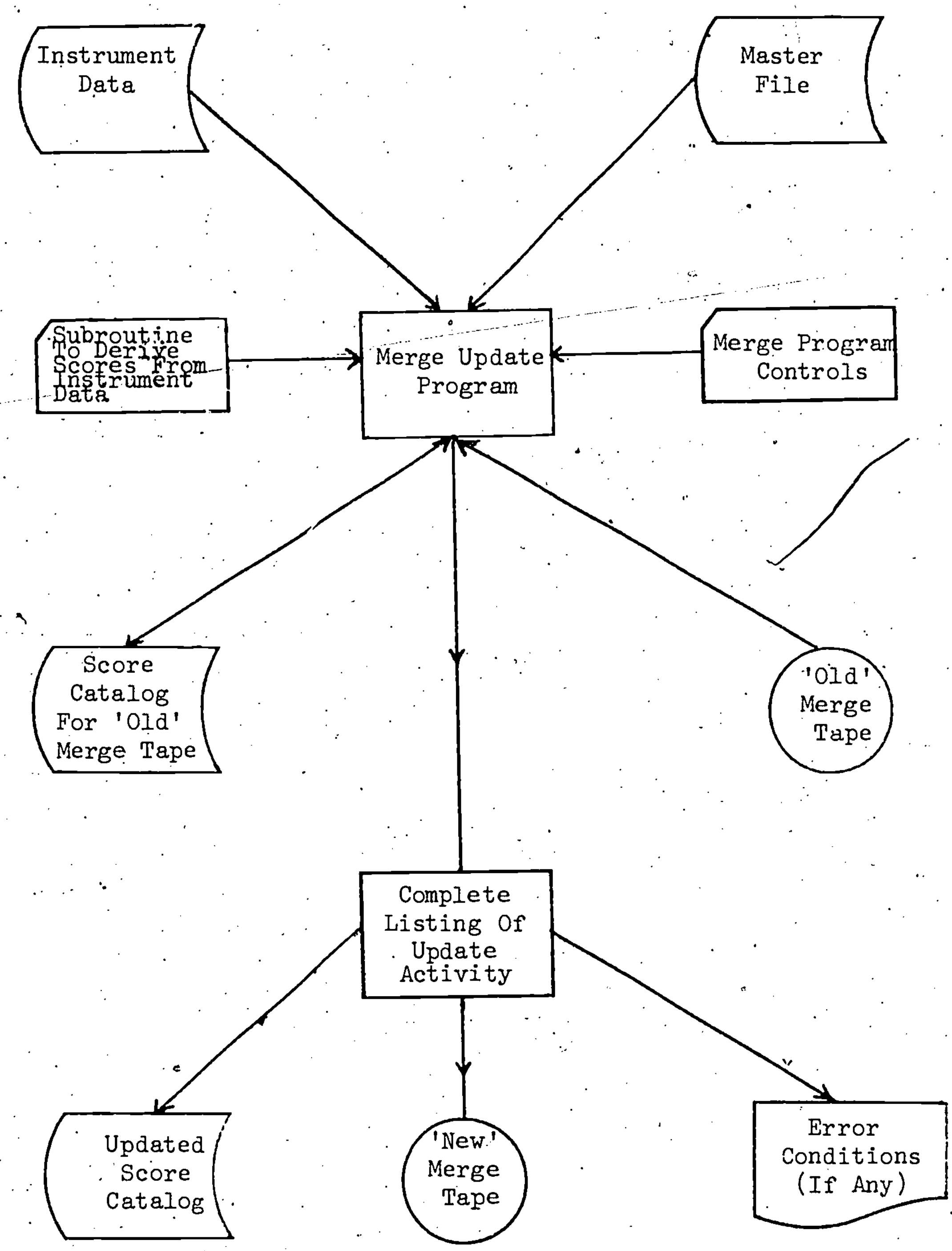
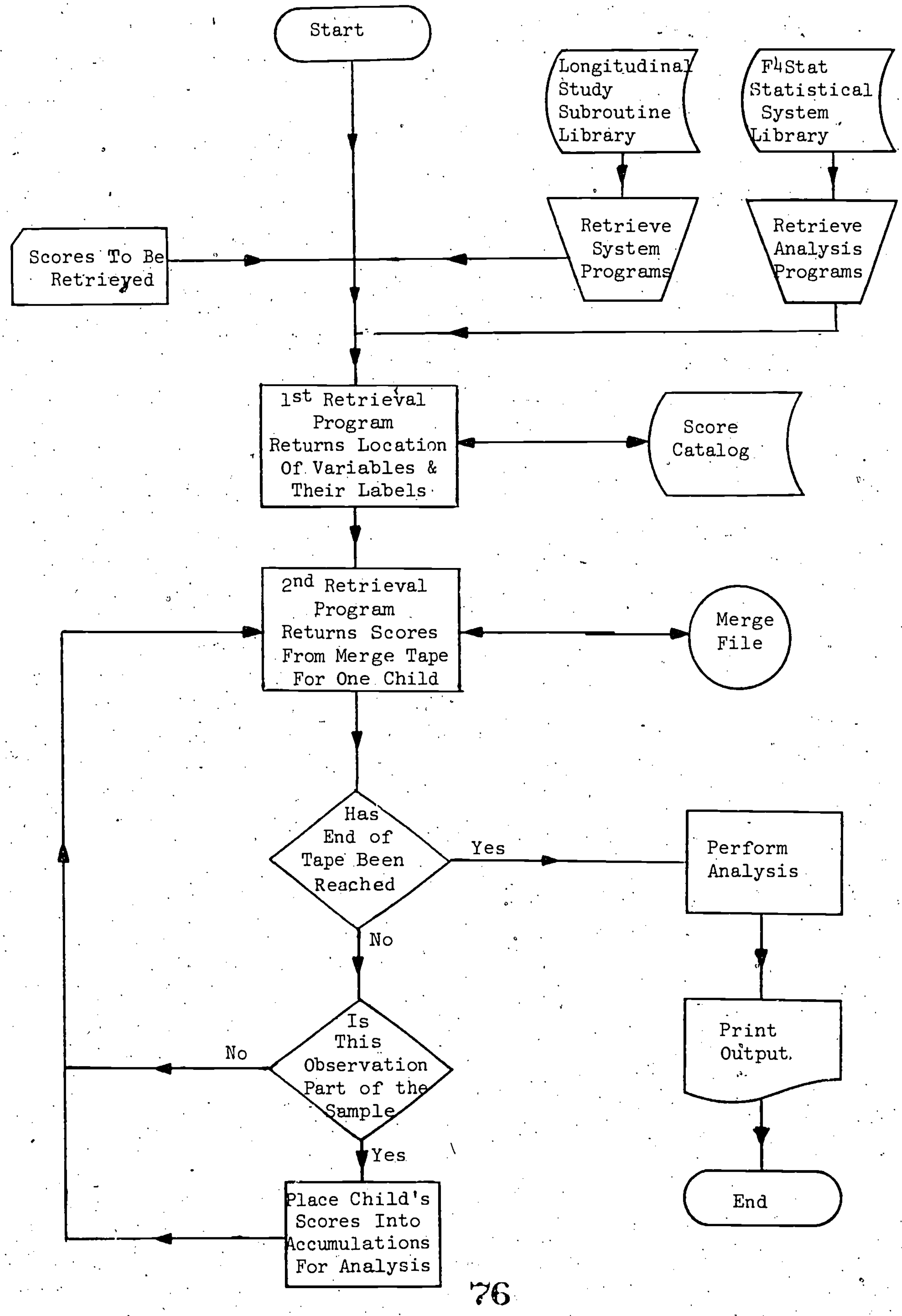


Figure 3-3

Merge Score Retrieval System



Chapter 4--RESULTS AND DISCUSSION

Initial Classification of Processes

In the first project report, Theoretical Considerations and Measurement Strategies (ETS, PR-68-4, Chapter C), a tentative outline was proposed by Messick for mapping the perceptual-cognitive-intellectual domain in a cross-classification scheme organized into hierarchical levels reflecting breadth of function and different orders of complexity. Basically an extension of Guilford's (1967) theory for the structure of the intellect, this outline combined features of dimensional, hierarchical, morphological and sequential models of intellect and incorporated variables derived both from the child development literature and from studies of adult performance. It also provided a guide for selecting instruments to represent the different types of contents, products, and operations delineated by Guilford.

Attention also was given to assessing those personality dimensions referred to as controlling mechanisms that cut across affective, personal-social and cognitive domains and thereby serve to interlace the cognitive system with other sub-systems of personality organization. In that same report (Chapter D) Emmerich delineated three other distinct but interrelated areas of personality investigation--social motives, attitudes, and interests--and suggested measurement strategies across the years of the study.

On the basis of this and other reviews of domains to be represented, variables considered salient for the study population were selected. Given the state of the art in measure development, tasks were selected which would allow: 1) continuity of measurement across age periods by using the same or vertically equivalent forms; and 2) multiple measurement of the same variable (within a context) across several age periods so that possible developmental shifts in expression could be monitored. Other factors affecting task selection were constraints related to available testing time, balance among areas and modes of response, sufficient knowledge of the appropriateness of the task for the intended population, and ease of administration.

Prior to statistical analysis, the Year 1 child test measures were grouped according to the above classifications (i.e., cognitive - perceptual - affective - physical) with certain sub-domains also suggested (e.g., within the cognitive domain, both Piagetian-derived measures and "academic" skills represented by verbal, quantitative and general information measures). Included in sub-domains were measures clustered according to contents, products or operations. Thus, verbal skill measures included comprehension of syntax, sequence and vocabulary, classification ability, and the ability to discriminate and mimic phonemes. In addition, cognitive styles (e.g., reflection-impulsivity, analytic functioning) and other controlling mechanisms such as risk-taking, curiosity, and attention deployment were delineated. A logical series of analyses were planned to study data within and across domains by mode and time of data collection. This report presents the results of the first "within-method" (i.e., individual testing) analyses, both within and across domains.

Overview of Structural Findings

As described in the previous section on data analysis procedures, following reduction to logically distinct scores for each task, principal components factor analyses using both unity and Tucker adjusted communalities on the diagonal were obtained. These analyses were performed for the composite sample and for major subject classifications; i.e., by age, sex, SES level, later preschool attendance controlled for Head Start eligibility, and Preschool Inventory [Caldwell] score with age partialled out. To facilitate interpretation, varimax and promax rotations of the first 2, 3, 4 and 5 principal components were performed successively. Six and seven factor varimax and promax rotations were also obtained for the composite sample. These same series of analyses were performed for a reduced set of variables posited to be in the cognitive domain. For these various analyses, 5 to 10 additional scores were included in extension analyses to study their relationships with factors derived from the main set of variables. In addition to the factor analyses, Guttman-Lingoes smallest space analyses (Lingoes, 1965) for 1, 2, and 3-dimensional solutions were performed on the same set of subject classifications and sets of variables.

The main findings of the factor and smallest space analyses of the data for the total group can be summarized as follows: 1) There was clear evidence of a general dimension accounting for most of the common variance among cognitive tasks. 2) A second orthogonal dimension relating to the child's speed of responding to a multiple choice task was obtained. 3) Additional factors that appeared were apparently tapping task-specific styles and behaviors (e.g., a factor principally defined by measures from the Open Field Task; a factor defined by two scores on the Fixation Task; a Spontaneous Numerical Correspondence factor; a Boy-Girl Identity Task factor). 4) Sub-clusters of tasks were not obtained; instead, considerable non-error specific variance was revealed for the many tasks used in the study. These findings were strikingly consistent across statistical methods and across subject classifications.* Given the saliency of the first factor and the small remaining common variance, slight differences in rotated solutions were very similar. The one exception was provided by children from ineligible families who later attended Head Start; for these children the latency measures split into the Spontaneous Correspondence and Open Field factors.

Table 4-1 presents the Tucker communality estimates for each score along with the estimated reliability where available. The estimates in Table 4-1 are based on the composite sample. Score abbreviations are included; task descriptions and a more detailed explanation of the scores used are presented in Appendix B. For all scores, coefficient alpha was the index of reliability. With few exceptions, estimated communalities were moderate to low, with considerable reliable but unique variance remaining.

Table 4-2 presents loadings for the first six unrotated principal components (using unities in the diagonal) and associated eigenvalues for the composite sample. Loadings with absolute values equal to or greater than .30 have been underlined. The generality of the first and second components and the specificity of the other components are clearly evident. The first principal component described in Table 4-2 accounted for 18.8% of the total variance; its eigenvalue was 8.6. The eigenvalue for the second component was 2.3, and it accounted for an additional 5% of the total variance. Subsequent components accounted for 3.9% or less of the variance. Fifteen components had eigenvalues of 1 or above. Utilizing communalities on the diagonal, the root for the first principal axis for the total sample was 8.1; it accounted for 50.6% of the common variance. Table 4-3 presents

*Among the six preschool attendance by eligibility categories, two groups, those Head Start eligible who attended a different preschool program and those who were not known to have attended preschool, had Ns too small to permit adequate comparisons of the factor structures.

Table 4-1

Estimated Communalities* and Reliabilities for Selected Scores

Score		Communality	Reliability
1	Hess and Shipman Toy Sorting Task: Total Score	.32	
2	Hess and Shipman Eight-Block Sorting Task: Total Score	.35	
3	Interaction Ratings: Mean Cooperation Rating (for 2 or 3 tasks)	.23	.81
4	Motor Inhibition Test: Average Time, Trial 2, for the Walking and Drawing Subtests	.26	.67
5	ETS Matched Pictures: Total Score	.21	.57
6	Preschool Inventory (Caldwell): Adjusted Total Score (minus items 52-55)	.68	.92
7	Form Reproduction: Total Score	.40	.65
8	Vigor 2 (Crank Turning): Average Number of Turns	.14	.86
9	Spontaneous Numerical Correspondence Task: Total Deviation Score	.35	.74
10	Spontaneous Numerical Correspondence Task: Total Configuration Matching	.54	.57
11	Massad Mimicry: Nonsense Words, Total Sounds (standardized by scorer)	.58	.91
12	Massad Mimicry: Meaningful Word Phrases, Final Sounds (standardized)	.53	.63
13	Risk Taking 2: Derived Score (0=toy only; 1=bag, trial 2; 2=bag, trial 1)	.03	
14	Picture Completion Subtest: Total Correct	.47	.89
15	Sigel Object Categorization: Total Grouping Responses	.33	.91
16	Sigel Object Categorization: Average Time to Response (Log 10)	.55	.77
17	Sigel Object Categorization: Total Correct Object Identification	.19	.62
18	Mischel Technique: Choice (0=smaller now; 1=larger later)	.02	
19	Johns Hopkins Perceptual Test: Total Correct	.46	.74
20	Open Field Test: Mean Play Complexity	.35	.61
21	Open Field Test: Number of Periods Child Talks to Himself	.09	.73
22	Open Field Test: Number of Periods Child Talks to Tester (1=if any)	.66	.81
23	Open Field Test: Number of Periods Child Approaches Tester (1=if any)	.07	
24	Open Field Test: Number of Periods Child Attempts to Leave Task	.18	
25	Open Field Test: Longest Simple Sequence	.28	.64
26	ETS Story Sequence Task: Total Score	.31	.50
27	Seguin Form Board: Fastest Time for Correct Placement	1.00	
28	Seguin Form Board: Number of Errors (for Trial with Fastest Time)	.35	
29	Matching Familiar Figures: Mean Log(X+1) of Response Times	.40	.90
30	Matching Familiar Figures: Mean Errors Per Valid Item	.58	.70
31	Fixation: Mean Recovery Time	.71	
32	Fixation: Mean Habituation	.40	
33	Brown Self Concept Task: Number of Items Omitted	.16	.91
34	Brown Self Concept Task: Self Concept Score (No. positive(1)/No. Coded 0 or 1)	.11	
35	Brown Self Concept Task: Smiling (1) or not smiling (0)	.04	
36	Preschool Embedded Figures Test: Total Correct	.20	.85
37	Preschool Embedded Figures Test: Average Time for First Response	.17	.77
38	Children's Auditory Discrimination Inventory: Total Correct	.52	.81
39	Peabody Picture Vocabulary Test, Form A: Total Correct to Criterion	.71	.96
40	Peabody Picture Vocabulary Test, Form B: Total Correct	.65	.93
41	Boy-Girl Identity Task: Task 1 (Girl), Item 1 Score	.67	
42	Boy-Girl Identity Task: Task 2 (Boy), Item 1 Score	.30	
43	Boy-Girl Identity Task: Sum of Task 1 Items 2, 3, 4, & 5	.01	.59
44	Boy-Girl Identity Task: Sum of Task 2 Items 2, 3, 4, & 5	.02	.64
45	Enumeration Task 1: Total Correct (Items 1 - 12)	.26	.85
46	Enumeration Task 1: Correct on Item 13 (counting)	.20	

*Communalities were obtained using Tucker's adjusted highest off-diagonal element

Table 4-2

First Six Principal Components for Total Group

Score*	1	2	3	4	5	6
1	0.56**	-0.06	-0.04	-0.04	-0.16	0.09
2	0.60	-0.01	0.00	-0.07	-0.05	0.10
3	-0.41	-0.04	0.23	0.17	-0.07	-0.06
4	0.53	0.08	-0.06	0.04	0.05	0.01
5	0.49	0.05	0.03	-0.04	0.07	0.00
6	0.83	-0.03	-0.01	-0.04	0.01	0.05
7	0.64	-0.03	-0.20	-0.10	0.02	0.05
8	0.39	0.12	-0.11	0.06	0.10	0.03
9	-0.21	0.29	0.37	0.21	0.15	-0.39
10	0.28	-0.24	-0.36	-0.29	-0.15	0.32
11	0.47	-0.08	0.30	-0.03	0.20	-0.11
12	0.46	-0.04	0.38	0.06	0.16	-0.07
13	0.02	-0.16	-0.01	0.01	-0.06	0.09
14	0.69	0.11	-0.08	-0.02	0.04	-0.06
15	0.60	-0.01	0.15	0.10	-0.14	-0.01
16	-0.09	0.66	-0.19	-0.20	0.12	0.00
17	0.40	-0.20	0.14	0.16	-0.05	-0.28
18	-0.03	-0.01	0.11	0.18	0.03	-0.02
19	0.55	-0.22	0.15	0.18	-0.16	-0.03
20	-0.08	0.26	-0.32	0.45	0.22	-0.01
21	0.03	0.35	0.07	0.26	0.12	0.15
22	0.21	0.36	0.34	-0.13	-0.06	0.20
23	0.02	0.26	0.20	-0.23	-0.05	0.18
24	0.00	0.17	0.41	-0.28	-0.24	0.12
25	0.06	-0.06	-0.47	0.48	0.32	-0.02
26	0.58	-0.27	0.12	0.13	-0.08	-0.01
27	-0.69	0.12	0.12	-0.05	0.05	0.09
28	-0.45	0.00	0.07	-0.02	0.01	0.22
29	-0.06	0.62	-0.08	-0.19	0.12	-0.09
30	-0.63	0.08	-0.08	-0.12	0.07	0.09
31	0.12	0.43	-0.14	0.26	-0.62	-0.11
32	0.10	0.36	-0.15	0.29	-0.65	-0.05
33	-0.39	-0.11	0.10	0.02	-0.12	0.05
34	0.32	0.10	-0.07	0.02	-0.10	-0.13
35	0.21	0.17	0.18	-0.02	0.10	0.20
36	0.43	0.19	-0.30	-0.09	-0.02	0.05
37	0.10	0.40	0.11	0.06	0.27	-0.05
38	0.65	0.20	0.08	-0.09	0.12	-0.08
39	0.77	0.16	0.06	-0.04	0.08	-0.11
40	0.75	-0.07	0.13	0.05	0.05	-0.14
41	0.30	0.10	0.19	0.28	0.08	0.52
42	0.19	0.05	0.17	0.40	0.03	0.56
43	-0.03	-0.05	-0.02	0.02	0.06	-0.12
44	-0.09	-0.06	0.01	0.14	0.10	0.30
45	0.48	0.01	-0.29	-0.22	0.07	0.06
46	0.46	0.03	-0.12	-0.18	0.12	0.04
***	8.63	2.31	1.79	1.56	1.46	1.39

*See Table 4-1 for score description.

**Loadings equal to or greater than .30 in absolute value are underlined.

***Eigenvalues: Although missing data correlations were used in these analyses, eigenvalues were obtained.

Table 4-3

Promax Correlations With Reference Factors*

Score	1	2	3	4	5	6
1	0.42**	-0.09	0.08	0.15	0.08	0.02
2	0.47	-0.02	0.01	0.12	0.07	0.05
3	-0.36	-0.08	0.03	-0.16	0.02	0.04
4	0.44	0.07	0.02	0.03	-0.04	0.04
5	0.41	0.05	-0.02	0.03	0.03	0.01
6	0.70	-0.01	-0.02	0.13	0.06	0.02
7	0.56	0.02	-0.02	0.18	-0.06	-0.02
8	0.34	0.09	0.02	0.01	-0.05	0.01
9	-0.07	0.19	0.03	-0.50	-0.03	-0.04
10	0.14	-0.16	-0.04	0.60	0.11	-0.01
11	0.46	-0.04	-0.16	-0.20	0.11	-0.07
12	0.41	-0.05	-0.10	-0.25	0.13	0.02
13	-0.01	-0.10	-0.01	0.06	0.00	-0.01
14	0.63	0.12	0.01	0.04	-0.04	-0.02
15	0.46	-0.09	0.10	-0.01	0.08	0.05
16	0.03	0.66	0.02	0.01	0.01	-0.04
17	0.34	-0.18	0.03	-0.10	-0.04	-0.05
18	-0.03	-0.03	0.00	-0.09	-0.03	0.04
19	0.42	-0.28	0.10	-0.04	0.01	0.04
20	0.01	0.23	0.06	-0.16	-0.36	0.09
21	0.02	0.19	0.06	-0.12	0.00	0.12
22	0.09	0.23	0.06	0.02	0.47	0.14
23	-0.01	0.13	0.02	0.03	0.21	0.02
24	-0.06	0.02	0.06	0.02	0.35	0.00
25	0.12	0.05	-0.04	-0.08	-0.48	0.08
26	0.44	-0.26	0.03	0.03	0.03	0.02
27	-0.65	0.12	-0.03	-0.03	0.29	-0.04
28	-0.43	-0.02	-0.01	0.06	0.16	0.02
29	0.06	0.56	0.02	-0.05	0.07	-0.07
30	-0.56	0.13	-0.09	0.07	0.02	0.03
31	0.06	-0.05	0.79	-0.05	0.06	-0.03
32	0.04	0.01	0.60	-0.02	0.04	0.00
33	-0.35	-0.10	0.01	-0.02	0.03	0.01
34	0.28	0.05	0.08	0.00	-0.01	-0.05
35	0.14	0.09	-0.02	-0.02	0.10	0.08
36	0.38	0.16	0.06	0.13	-0.08	0.00
37	0.14	0.31	-0.03	-0.15	0.01	0.03
38	0.60	0.20	-0.03	-0.05	0.07	-0.01
39	0.73	0.18	-0.02	-0.06	0.04	-0.02
40	0.68	-0.06	-0.04	-0.09	0.00	-0.02
41	-0.01	0.02	-0.05	0.01	0.00	0.74
42	-0.04	-0.04	0.02	-0.03	-0.03	0.52
43	0.00	-0.01	-0.02	-0.03	-0.04	-0.03
44	-0.10	-0.04	-0.03	0.01	0.00	0.08
45	0.42	0.07	-0.04	0.18	-0.07	-0.04
46	0.40	0.08	-0.06	0.10	-0.01	-0.02

* Using communalities in the diagonal

**Loadings equal to or greater than .30 in absolute value are underlined.

the 6-factor promax solution for the composite sample using communalities in the diagonal, with intercorrelations among factors reported in Table 4-4. As can be seen, the structure is highly similar to the structure evidenced in Table 4-2.

Table 4-4

Intercorrelations Among Promax Primary Factors for Six-Factor Solution

	1	2	3	4	5	6
1		-0.16	0.06	0.08	0.18	0.36
2	-0.16		0.18	0.14	-0.03	-0.02
3	0.06	0.18		0.08	-0.09	0.08
4	0.08	0.14	0.08		-0.27	-0.02
5	0.18	-0.03	-0.09	-0.27		0.12
6	0.36	-0.02	0.08	-0.02	0.12	

The task-specific nature of the 3rd through 6th rotated factors may be seen quite clearly in Table 4-3. (For further comparison, the 6-factor varimax solution using communalities in the diagonal is presented in Appendix C.)

Information - Processing Factor

Inspection of Table 4-2 reveals the diversity of tasks contributing to the first component. Out of forty-six scores from the twenty-six tasks, twenty-two of these tasks had a score with loadings of .30 or higher. As might be expected, the most general task in the test battery, the Preschool Inventory, had the highest loading (.83),* but the following all had loadings of .55 or higher: verbal measures -- receptive and productive vocabulary (Peabody A and B), classification skill (Sigel Grouping responses, Toy Sorting and Eight-Block Sorting Task scores); perceptual measures -- auditory

*It should be noted that the TAMA test of general information was not included in these analyses because a considerably smaller sample was available.

discrimination (Children's Auditory Discrimination Inventory), form discrimination and matching (Johns Hopkins Perceptual Test score and Matching Familiar Figures Test errors); and perceptual-motor measures -- visual-motor coordination (Seguin), form reproduction (Form Reproduction Test). Virtually identical results were obtained when comparing the ranking of deviation coefficients from the Caldwell (2-dimensional smallest space solution) with the ranking of weights on the first principal component. A factor that was very similar to the first principal component appeared after rotations of varying numbers of factors (between 2 and 7) whether unities or communality estimates were used in the diagonal of the correlation matrix. This is evident in the comparison of the first factor in Table 4-2 (first principal component using unities in the diagonal) and the first factor in Table 4-3 (first promax factor using communality estimates).

The first component seemed to be best defined as "g" or information-processing skills which contribute to level of performance on all of these tasks. For this sample, it was best represented by performance on the Preschool Inventory and Peabody Picture Vocabulary Test (PPVT) which correlated .62. The Preschool Inventory was developed to measure achievement in areas regarded as critical for successful kindergarten performance. To some extent performance on this task is an index of the child's ability to process general information from the environment. Millham et al. (1971) recently described scores on a vocabulary test as measuring associative information-processing ability. Both tests have been found to be highly sensitive to environmental impoverishment. Included in measures of "g," of course, are such "non-cognitive" aspects as ease and willingness to relate and assert oneself in the testing situation, attention, persistence, and task orientation. A common cognitive component is the ability to understand and follow directions. These aspects of "g" may, however, be age-specific.

Response Tempo Dimension

The second factor describing the overall correlational structure appears to represent a response tempo dimension; as defined by oblique rotation (Table 4-3), this factor was nearly orthogonal to the first factor. The correlation between the first two promax factors was -.16. It was best represented by the mean latency scores on the Sigel Object Categorizing Test and the Matching Familiar Figures Test (r=.47). The only other variable with a loading of .30 or greater on the second primary factor was the average time to first response on the Embedded Figures Test (loading = .31). Thus, response tempo, frequently used to measure the cognitive style of reflection-impulsivity, appeared as a consistent individual difference variable; however, for this sample during this age period response tempo was not related to performance level on the first factor. Similarly, latency and adequacy of response were not correlated within tasks (r = -.07 with grouping responses on the Sigel and .02 with errors on the Matching Familiar Figures Test). Response latency, therefore, did not have the same implication for performance as has been found with older and/or more advantaged subjects (Messer, 1970; Eska and Black, 1971), since it did not reflect individual differences in the degree to which the child considers the adequacy of his response. Perhaps, prior to school experience, there is a lack of anxiety or concern over error and/or fewer internalized standards of performance. These findings suggest that temperamental components have not yet become integrated into the cognitive domain.

Further Analyses of Cognitive Measures

Subsequent to these analyses, factor analytic and smallest space analyses were performed on those measures contributing to the first component in an attempt to delineate clusters, if any, among the cognitive measures. The resulting clusters of measures were not stable and the factors, as expected, were highly correlated. The verbal measures, for example, did not sort into receptive skills (e.g., comprehension of syntax [ETS Matched Pictures]; sequence [ETS Story Sequence - Part I]; and vocabulary [Peabody A]) and productive skills (e.g., labelling [Sigel Object Identification and Peabody B scores]; imitation of phonemes and real words [Massad Mimicry]) or into different subskills such as classification (Toy Sorting and Eight-Block Sorting scores and Sigel Grouping responses) or vocabulary comprehension (Matched Pictures [functor words] and Peabody A [content words]). Neither was there evidence of larger clusters, e.g., verbal, quantitative, perceptual. Also, a separate factor analysis of the Preschool Inventory yielded only a large general factor and item-specific factors rather than logically distinct clusters of quantitative, verbal, social and perceptual-motor items.

As would be predicted, a "g" factor became much more salient, when the non-cognitively defined measures were removed. The first factor accounted for 32.3% of the total variance; its eigenvalue was 8.08. Using communalities in the diagonal, the eigenvalue was 7.56, and 76.8% of the common variance was accounted for. Again, however, the magnitude of remaining non-error specific variance was clearly evidenced. Varimax and promax 5-factor rotations produced factors whose interpretation paralleled that of the larger factor analyses. The first factor appeared to represent a general ability dimension, the second a stylistic one, and remaining factors were task-defined. Thus, the general ability dimension evidenced on the first factor did not differentiate into various cognitive-intellectual domains; instead, attitudinal and/or highly specific determinants appeared to be operating.

Intercorrelations among the "cognitive" subset of measures further clarify these results. Intercorrelations among the receptive verbal measures were moderate to low.* Matched Pictures and Story Sequence correlated .38 and .41, respectively, with Peabody A, and .33 and .31, respectively, with the CADI; their correlation with each other was only .25. Thus, these measures appear to be tapping different verbal skills. The CADI, however, correlated .61 with Peabody A. As can be seen from the task descriptions, operations involved in both tasks are highly similar -- the child must point to the picture representing the oral stimulus. Moreover, the auditory discrimination task involves vocabulary skill to the extent that children had differential familiarity with the meaning of the real words.

Similarly, correlations among verbal productive measures were moderate to low. The sorting tasks correlated higher among each other (Toy Sorting and Eight-Block Sorting scores correlated .31 and .37, respectively, with Sigel Grouping responses), than with the Mimicry scores (.19 to .25) or with the number of objects identified correctly on the Sigel (.19 and .17, respectively). Except for the correlation of .49 between Toy Sort and Eight-Block Sort performance, correlations among the above tasks were highest with Peabody B (.33 [Mimicry Nonsense Words] to .43 [Eight-Block Sorting

*The intercorrelations among all of the scores listed in Table 4-1 are reported in Appendix C.

Test score and Sigel Grouping responses]). However, these measures correlated to a similar extent with the receptive verbal scores presented above; correlations with Peabody A were highest and virtually identical to those with Peabody B. Mimicry scores tended to have low intercorrelations with other tasks suggesting that imitation skill is distinct from understanding or producing language.

In Year 1, except for the counting and ordination items of the Preschool Inventory, measurement in the quantitative area consisted of two measures, Enumeration and Spontaneous Numerical Correspondence, both tapping what Piaget (1952) considered prerequisites for the later understanding and use of number -- perceptual ordering and articulation. These two measures did not form a quantitative cluster and their placements in the structure were quite different. Perhaps these measures are differentially related to general mathematical concepts and to numerical and computational skills. Correlations among the four task scores ranged from .11 to .18. The Enumeration scores, which are more closely linked to computational skills, loaded on the first factor and correlated in the 20's and 30's with other measures. Spontaneous Numerical Correspondence, which provides information on global intuitive responses of young children to problems in one-to-one correspondence, had correlations with all other tasks that were close to zero and defined a separate factor. (Factor 4 in the promax rotation reported in Table 4-3.)

Distinct from "academic" skills, but subsumed under the cognitive domain, are Piagetian-derived measures. Three tasks were included in the initial test battery to learn more about the preoperational stage in general and to chart its course in the socioeconomically disadvantaged child. Two of these tasks have just been discussed. The third, the Boy-Girl Identity Task, assesses the child's ability to maintain gender identity constancy despite changes in stimuli which increasingly resemble the opposite sex. As noted in the task description, although responses to item 1 (the "wish item") were moderately correlated (.45), correlations near zero were obtained with the other task scores. Item 1 was correlated to some extent (approximately .20) with those tasks highest on the general ability factor, whereas the other scores had virtually zero correlations with all other measures in the test battery. As with Spontaneous Correspondence, the two correlated scores formed a separate factor (Factor 6 in Table 4-3). Thus, this task did not seem to be tapping a cognitively based reality judgment of gender identity constancy in this population at this age.

The distinction between perception and cognition is obviously a fine one. Measures in Year 1 tapped form discrimination and recognition (Johns Hopkins Perceptual Test, Matching Familiar Figures Test), form analysis (Preschool Embedded Figures Test, WPPSI Picture Completion), eye-hand coordination (Seguin Form Board), and form reproduction (items from the Caldwell and Developmental Test of Visual Motor Integration). These tasks did show moderate-sized correlations with each other, and the relative size of the intercorrelations reflected task similarities. For example, the Johns Hopkins and Matching Familiar Figures tests, which make highly similar demands upon the child, correlated .52;* the two eye-hand coordination tasks, Seguin and Form Reproduction, correlated .47.* However, these tasks did not form a separate cluster. Instead, they all showed substantial loadings on

*Signs of correlations involving error scores (Matching Familiar Figures) and time to quickest solution (Seguin) have been reversed.

the first factor. As would be expected, those perceptual tasks which are included in the WPPSI, Picture Completion and Form Reproduction, had the highest correlations with the child's general achievement level as reflected in his Caldwell score (.59 and .54, respectively). Studying correlations of these measures with other tasks does provide clues, however, to task specific components. Sigel grouping responses, for example, correlated .41* with both Matching Familiar Figures and the Johns Hopkins, reflecting the child's ability to discriminate stimulus characteristics as a basis for sorting. The highest frequency of classification on the Sigel was by manifest stimulus attributes such as color or form.

Other Measures of Cognitive Style

Other cognitive-stylistic factors, such as analytic functioning, did not appear. Performance on both the Preschool Embedded Figures Test and the Picture Completion subtest of the WPPSI, which were included as potential markers of analytic functioning, loaded primarily on the first factor (.38 and .63, respectively); their intercorrelation was only .29. This finding could reflect insufficiency of measurement in defining this factor. Recent evidence suggests that the Picture Completion subtest is not a stable and consistent index of analytic functioning. Moreover, for many children in this sample performance was confounded by lack of understanding of the word "missing" in the directions. Both the Preschool Embedded Figures Test and Picture Completion Test exhibited substantial reliabilities, however, indicating consistent but specific functioning on these tasks.

Although speed of responding emerged as a factor in the overall analyses, the lack of relationship of the latency measures to other purported measures

*Signs of correlations involving error scores (Matching Familiar Figures) and time to quickest solution (Seguin) have been reversed.

of impulsivity (inability to inhibit a response or to delay gratification) suggests that impulsivity is not a unitary trait or generalized dimension in this population at this age. The correlation of mean latency on the matching Familiar Figures Test with mean time in the "slow" walking and draw-a-line trials of the Motor Inhibition Test was .04, and with choice of the delayed reward on the Mischel Task it was -.01; the correlation between these latter measures was -.07. As indicated in previous research (Maccoby et al., 1965; Massari et al., 1969; Ward, 1968), the ability to inhibit a response when appropriate (Motor Inhibition "slow" trial time), unlike a more stylistic variable such as latency, is positively correlated with IQ. In this study the Motor Inhibition score loaded only on the first factor and not on the second tempo factor. As noted in describing the Mischel task, impulsivity as defined by choosing the smaller but immediate reward is confounded by the child's understanding of the instructions and his faith in the tester. Delaying gratification is realistic only if one sees an opportunity to achieve gratification at a later time. Obviously, the task requirements for these measures are dissimilar (e.g., one response choice vs. two or more possible responses), and such method variance appears more potent than an underlying unifying personality dimension, at least in this population at this age. However, for primary grade children from low-income families, Hess et al. (1969) also found motor inhibition, reflection-impulsivity and delayed reward not to be correlated.

Other controlling mechanisms had near-zero communality estimates (e.g., risk-taking had an estimated communality of .03) or appeared as a task-defined factor (e.g., the four variables that define factor 5 in

Table 4-3 are Open Field Test measures). Such results could be interpreted as reflecting special abilities limited to one task and/or incomplete sampling of the processes represented by tasks. Given the different setting for the Open Field Test (unstructured play observation vs. adult-child testing situation), it is not known to what extent task-specific variance was due to the method difference rather than to the processes being tapped.

Attentional variables are among those that cut across relatively arbitrary distinctions between cognitive and personal-social functioning. Lewis and his associates (Lewis et al., 1970; Lewis, 1971) have found attention to be an index of early cognitive functioning. Not only may attention be a prerequisite of subsequent cognitive functioning, but individual differences in attention are likely to have direct effects on learning. Moreover, attention can be non-cognitively determined as well by the intentions and desires of the subject. As indicated earlier, the two Fixation Test scores used in the structural analysis appeared as a task-defined factor. (Factor 3 in the promax rotation reported in Table 4-3.) However, their lack of relationships to other measures and the low correlations across stimuli within the task prevent us from interpreting these findings further at this time.

Similarly, personal-social behaviors reflected in the Brown IDS Self-Concept Referents Test (smiling in the photograph, self-concept score), the Open Field Test (approaches E, attempts to leave) and the ratings of child cooperation in the mother-child structured interaction sessions had communalities less than .25 for the composite sample. However, the test battery did not sample enough of these behaviors to delineate factors in the affective and social domains. Given the present "state of the art" in valid measurement of these variables for this age, however, it is doubtful that other results would be obtained with more extensive measurement in test-like settings. In preliminary analyses the self-concept score, as an extension variable, mapped onto the first component only; in the promax 6-factor solution it loaded on the first factor only (.28), suggesting the importance of general intellectual competency as a critical component in positive self-evaluation. Similarly, the mean rating of the child's cooperation during the mother-child interaction sessions mapped onto the general ability measures and reflected the attitudinal, non-cognitive components of measures of "g". These results suggest that the child's task involvement and compliance with the mother in the interaction situations was highly similar to his behavior with the tester. We had intended to include tester ratings in the Year 1 battery, but the various ratings proposed were too numerous and too complex in format for testers to assimilate quickly and well. Since the various demands at the time did not allow adequate revision, these ratings were not used. Test ratings were included, however, in subsequent testing.

The one measure representing the physical domain, Vigor 2 (crank-turning), had a communality estimate of only .14, but given the lack of other similar measures no further interpretation of this score can be made at this time. As indicated in the task description, the extent to which this measure taps vigor, persistence, physical coordination and/or willingness to please the examiner is unknown. In Appendix B it is noted that another vigor measure had been administered (running), but since further inspection of the data suggested the presence of several confounding factors (e.g., fear of falling, closeness of the finish line to walls), the running speed score was not included in the structural analyses. These findings may reflect general instability of performance in this area at this age period.

Smallest Space Analyses

The smallest space solutions yielded highly similar results. The two-dimensional solution of the larger matrix produced one vector that seemed to define the first two factors (i.e., general ability and response tempo). Those tasks with highest loadings on the first factor were clustered together at one end, with the latency measures at the opposite end. In between these cognitive and tempo measures were the Fixation Test scores. The other measures were scattered around the space and no general clusters were evidenced.

Given the lack of clustering and low intercorrelations among the remaining measures, no clear interpretation of the second dimension can be made at this time. Tentative hypotheses refer to differences in the number of response options offered the child (one for skill measures vs. two or more for the Risk-Taking Task, Mischel, Boy-Girl Identity Task and Open Field Test, all of which were located at one end of the second dimension), and to the possible existence of a social dimension (smiling for the Brown Self-Concept Test photograph and talking to the tester in the Open Field Test). Thus, the task-defined factors discussed previously were located along the boundaries of the smallest space solutions. This smallest space solution did not aid, however, in further delineation of processes represented by those particular measures since there was no general clustering of measures.

Similar results were obtained when the measures assumed to be non-cognitive were removed. Measures on the first factor clustered together, defining the general ability component, with task-defined factors located along the boundaries of the smallest space solutions. Such tasks may be highly specialized (e.g., Massad Mimicry) and/or age-specific (e.g., Spontaneous Numerical Correspondence) in their implications. Their distance from the "g" dimension would seem to indicate that they have not yet been organized by central information-processing abilities. The smallest space solutions thus provided essentially similar information as to the major structural dimensions of the data -- suggesting general information-processing skill, response tempo and many task-specific aspects.

However, this analysis provided additional insight into the processes underlying the correlational structure. Moving away from those measures which were found to load highest on the first principal component and which, in the smallest space solution for the subset of measures, were centered around the Preschool Inventory score, it was discovered that those measures closest together were those which had been given in the same battery (i.e., same day and same tester). Thus, a secondary structuring variable for these data appears to be a contextual one. This battery effect seems equivalent to what Campbell and Fiske (1959) call method factors, and may reflect the operation of situationally determined variables in test performance. Thus, for example, self-concept and vigor scores were located close together not only because of their common lack of relatedness to the major skill-tempo vector, but presumably because of shared method variance in that the Brown and Vigor 2 tasks were both administered in Battery C. (See Table 3-1 in Chapter 3 for battery descriptions.) Given the fact, however, that it was the day of administration and tester which defined the battery effect, situational and tester characteristics rather than stable child style or personality variables may have created clustering. The smallest space solutions for the larger set of variables indicate, however, that such situational determinants were of secondary importance in organizing the data. Those tasks with logically independent scores (e.g., Open Field, Sigel, Matching Familiar Figures) had scores located in different parts of the space.

Factor Similarities Across Subgroups

Tables 4-5 and 4-6 present loadings on the first principal axes factors (using communalities in the diagonal) for the various subject classifications on the fifteen measures with highest loadings for the composite sample. The variables in Tables 4-5 and 4-6 are ordered in descending order of their loadings on the first principal axes factor for the total sample. As examination of the tables reveals, the loadings for the various subgroups are very similar to those of the total sample.

Although the factor analytic solutions by sex, age, preschool attendance, Preschool Inventory score and SES level were essentially the same with regard to pattern as for the total sample, a considerably smaller percentage of the common variance was accounted for by the first component in subjects who were younger and below their age-group mean on the Preschool Inventory (younger = 43.4 vs. older = 52.3 and low Caldwell = 31.1 vs. high Caldwell = 44.1, respectively). In contrasting eligibility by preschool attendance groups, it was found that a substantially smaller percentage of the common variance was accounted for by the first component for children from eligible families who later attended Head Start than for children from ineligible families who attended other preschool programs or no known preschool program (37.9% vs. 47% and 49.7%, respectively). The amount of common variance accounted for

Table 4-5

Task Loadings on First Principal Component* by Sex, Age, SES and Preschool Inventory Score Subgroups

		Sex		Age		SES		Caldwell	
Score	Total	M	F	Y	O	Lo	Hi	Lo	Hi
6	.82	.82	.82	.79	.83	.80	.85	.58	.70
39	.77	.77	.77	.75	.77	.75	.70	.62	.72
40	.74	.76	.74	.76	.72	.70	.65	.62	.65
27	-.72	-.73	-.70	-.66	-.71	-.68	-.63	-.66	-.68
14	.67	.65	.69	.60	.67	.66	.64	.48	.60
38	.64	.61	.66	.63	.65	.60	.61	.43	.58
30	-.62	-.61	-.63	-.59	-.61	-.60	-.62	-.47	-.60
7	.62	.61	.62	.55	.61	.62	.60	.42	.51
15	.57	.60	.54	.57	.54	.54	.58	.31	.52
2	.57	.61	.55	.54	.59	.49	.62	.32	.53
26	.55	.53	.57	.53	.56	.51	.55	.42	.49
19	.53	.51	.55	.51	.51	.49	.58	.40	.55
1	.53	.52	.54	.48	.54	.41	.59	.23	.49
45	.45	.45	.46	.43	.43	.42	.44	.42	.43
28	-.43	-.43	-.42	-.43	-.43	-.42	-.34	-.36	-.41

Table 4-6

Task Loadings on First Principal Component* by Head Start Eligibility and Preschool Attendance Subgroups

		Eligible	Ineligible		
Score	Total	HS	HS	PS	Not Known
6	.82	.77	.79	.84	.80
39	.77	.70	.70	.77	.77
40	.74	.65	.63	.62	.67
27	-.72	-.62	-.75	-.68	-.72
14	.67	.56	.63	.67	.64
38	.64	.50	.59	.67	.68
30	-.62	-.41	-.61	-.67	-.66
7	.62	.56	.46	.54	.58
15	.57	.43	.46	.60	.60
2	.57	.47	.54	.62	.53
26	.55	.44	.44	.58	.51
19	.53	.44	.57	.59	.52
1	.53	.44	.31	.63	.49
45	.45	.47	.46	.50	.43
28	-.43	-.36	-.43	-.41	-.52

* Using communalities in the diagonal

by the first component was essentially identical for boys and girls (49.7% vs. 50.0%) and for children whose families were classified as blue collar or white collar (46.8% vs. 46.6%). Tables 4-7 and 4-8 present the eigenvalues for the total group and each of these subject classifications. As can be seen, the first few factors account for less variance for the developmentally less mature subjects. These data reflect in part the greater instability and reduced variability of measurement for the less mature subjects, but they also suggest that less integration of the child's response repertoire has taken place. As noted earlier, considerable differentiated behavior was reflected in the large amount of unaccounted-for non-error variance. The above findings suggest that such differentiated behavior may reflect fragmented behaviors which have not yet been organized or integrated.

Although the pattern of performance was similar across groups, the level of performance was not. In a further attempt to understand interrelationships among measures, we looked at possible patterns of determinants effecting significant differences in performance. Three-way analyses of variance, Sex by Age (median split) by SES (blue-collar/white-collar occupation of head of household), were computed for each of the scores included in the overall structural analyses as were two-way analyses of variance using Head Start eligibility by Preschool Attendance classifications. The results of these ANOVAS are discussed in terms of the factors that emerged in the structural analyses presented earlier.

Mean Differences Among Sex, Age, and SES Groups

Table 4-9 summarizes the results of the Sex by Age by SES ANOVAS. Each score is identified according to the factor on which it had the highest

Table 4-7

Score Eigenvalues* by Sex, Age, SES and Preschool Inventory Subgroups

		Sex		Age		SES		Caldwell	
Index	Total	M	F	Y	O	Lo	Hi	Lo	Hi
1	8.11	8.07	8.25	7.39	7.95	7.28	7.91	4.52	6.60
2	1.75	1.79	1.70	1.98	1.59	1.83	2.00	2.08	1.63
3	1.21	1.24	1.32	1.32	1.18	1.16	1.66	1.34	1.43
4	1.00	1.03	1.11	1.15	1.08	1.03	1.28	1.03	1.12
5	.89	.94	.98	1.08	.93	1.01	1.05	1.00	1.07
6	.84	.91	.89	.89	.85	.90	.92	.96	.89
7	.80	.84	.77	.88	.75	.78	.87	.91	.75
8	.66	.70	.66	.76	.62	.72	.74	.87	.63
9	.61	.63	.62	.65	.57	.68	.69	.69	.63
10	.52	.52	.55	.61	.51	.60	.60	.64	.56
11	.47	.45	.50	.53	.44	.55	.57	.56	.41
12	.37	.36	.47	.50	.44	.44	.52	.52	.38
13	.27	.34	.34	.45	.31	.37	.40	.35	.32
14	.24	.28	.33	.35	.24	.32	.37	.33	.30
15	.18	.23	.30	.29	.23	.28	.36	.29	.26
16	.17	.20	.25	.25	.19	.21	.30	.25	.22
17	.11	.19	.18	.22	.17	.21	.28	.22	.17
18	.11	.15	.15	.19	.15	.16	.23	.19	.15
19	.11	.14	.14	.16	.11	.15	.19	.16	.12
20	.08	.12	.12	.12	.11	.12	.16	.14	.12
21	.05	.09	.10	.11	.07	.10	.13	.10	.11
22	.04	.07	.07	.08	.06	.08	.10	.09	.08
23	.03	.06	.06	.07	.05	.07	.08	.08	.04
24	.01	.01	.03	.06	.02	.05	.04	.06	.04
25	.01	.01	.02	.04	.01	.01	.02	.06	.02
26	-.01	.00	.00	.01	-.01	-.01	.01	.04	.01
27	-.02	-.01	-.02	.00	-.04	-.02	-.02	.02	-.01
28	-.02	-.03	-.03	-.01	-.04	-.03	-.08	.00	-.02
29	-.04	-.04	-.04	-.02	-.06	-.04	-.09	-.03	-.04
30	-.05	-.05	-.06	-.04	-.06	-.05	-.09	-.04	-.04
31	-.05	-.07	-.07	-.06	-.07	-.07	-.11	-.05	-.06
32	-.06	-.10	-.09	-.08	-.09	-.10	-.13	-.07	-.08
33	-.07	-.11	-.10	-.09	-.10	-.12	-.14	-.09	-.09
34	-.08	-.11	-.13	-.11	-.13	-.12	-.18	-.09	-.11
35	-.10	-.12	-.14	-.13	-.14	-.16	-.19	-.11	-.12
36	-.12	-.13	-.16	-.14	-.15	-.18	-.21	-.12	-.14
37	-.14	-.17	-.18	-.16	-.16	-.18	-.22	-.14	-.17
38	-.14	-.17	-.19	-.16	-.18	-.20	-.24	-.15	-.18
39	-.16	-.19	-.20	-.18	-.20	-.21	-.27	-.16	-.18
40	-.18	-.20	-.22	-.20	-.23	-.23	-.28	-.19	-.20
41	-.20	-.22	-.24	-.22	-.24	-.26	-.29	-.21	-.21
42	-.21	-.24	-.26	-.23	-.28	-.28	-.31	-.23	-.25
43	-.22	-.26	-.26	-.27	-.29	-.30	-.37	-.27	-.27
44	-.24	-.27	-.27	-.28	-.29	-.33	-.38	-.30	-.28
45	-.25	-.30	-.30	-.34	-.31	-.35	-.40	-.32	-.29
46	-.27	-.34	-.45	-.39	-.37	-.38	-.49	-.36	-.33
Trace	16.02	16.23	16.51	17.02	15.21	15.55	16.97	14.56	14.96

* Using communalities in the diagonal

Table 4-8

Score Eigenvalues* by Head Start Eligibility and Preschool Attendance Categories

Index	Total	Eligible HS	HS	Ineligible PS	Not Known
1	8.11	6.16	7.06	8.29	8.16
2	1.75	2.10	2.09	2.11	1.93
3	1.21	1.38	1.55	1.64	1.67
4	1.00	1.33	1.48	1.47	1.31
5	.89	1.19	1.27	1.37	1.14
6	.84	1.05	1.13	1.08	1.08
7	.80	.95	1.03	1.04	.94
8	.66	.83	.93	.87	.82
9	.61	.80	.90	.71	.72
10	.52	.79	.78	.63	.66
11	.47	.63	.70	.59	.54
12	.37	.56	.55	.51	.53
13	.27	.53	.51	.47	.49
14	.24	.48	.48	.44	.48
15	.18	.36	.43	.37	.38
16	.17	.33	.34	.36	.36
17	.11	.28	.30	.25	.28
18	.11	.24	.25	.21	.25
19	.11	.23	.22	.20	.17
20	.08	.19	.15	.17	.14
21	.05	.14	.14	.14	.12
22	.04	.07	.10	.08	.07
23	.03	.04	.10	.05	.05
24	.01	.04	.05	.04	.01
25	.01	.01	.01	.02	-.01
26	-.01	.00	-.01	-.01	-.02
27	-.02	-.01	-.03	-.05	-.07
28	-.02	-.04	-.05	-.06	-.09
29	-.04	-.05	-.09	-.09	-.10
30	-.05	-.07	-.12	-.13	-.13
31	-.05	-.09	-.14	-.15	-.16
32	-.06	-.10	-.17	-.16	-.19
33	-.07	-.14	-.19	-.21	-.21
34	-.08	-.18	-.21	-.21	-.23
35	-.10	-.19	-.24	-.24	-.26
36	-.12	-.20	-.26	-.27	-.28
37	-.14	-.23	-.29	-.29	-.30
38	-.14	-.24	-.32	-.31	-.32
39	-.16	-.27	-.34	-.32	-.35
40	-.18	-.30	-.35	-.34	-.37
41	-.20	-.30	-.39	-.37	-.38
42	-.21	-.32	-.40	-.40	-.42
43	-.22	-.36	-.42	-.41	-.46
44	-.24	-.42	-.45	-.44	-.48
45	-.25	-.43	-.52	-.48	-.50
46	-.27	-.49	-.54	-.53	-.54
Trace	16.02	16.26	17.01	17.63	16.41

* Using communalities in the diagonal

Table 4-9

Significant Analysis of Variance Effects by Sex, Age, and SES[1]

Score	Factor[2]	Sex[3]	Age	SES	SxA	SxSES	AxSES	SxAxSES
1	1		X	X				
2	1		X	X				
3	1	X^G	X	X				
4	1		X	X				
5	1		X	X				
6	1	X^G	X	X				
7	1	X^G	X	X				
8	1	X^B	X	X				
9	3		X					
10	3		X					
11	1	X^G	X	X				
12	1			X				
13	-	X^B						
14	1		X	X				
15	1		X	X				
16	2							
17	1		X	X				
18	-							
19	1		X	X				
20	4	X^B						
21	6	X^B						
22	4			X				
23	4							
24	4							
25	4							
26	1		X	X				
27	1		X	X				X
28	1			X				
29	2							
30	1		X	X				
31	5							
32	5							
33	1		X	X				
34	-		X	X				
35	-	X^G						
36	1		X	X				
37	2							
38	1			X				
39	1		X	X				
40	1		X	X				
41	6			X				
42	6			X				
43	-	X^B						
44	6							
45	1	X^G	X	X				
46	1		X	X				

1. X appears as an entry when the designated effect is significant ($p < .01$).
2. Numerical entries represent factor with highest loadings above .30 on 6-factor promax solution; a dash indicates that score did not load above .30 in this solution.
3. B signifies boys performed better; G signifies girls performed better.

loading (of those .30 or higher), and significant effects are noted ($p<.01$).

In chapter 2 we directed attention to disproportionalities and confounding of status classifications used in this report, pointing out the limitations this situation imposed on interpretation of findings reported in this chapter. Consequently, the findings reported below must be read with caution and any interpretations regarded as highly tentative.

None of the two-way interactions were significant at the .01 level and only one of the 46 measures had a significant three-way interaction. There were a substantial number of significant main effects. The description of the significant main effects is organized to correspond to the clustering of variables obtained in the factor analyses.

ANOVAS for measures on the first factor were consistent in showing significant SES differences and, with only three exceptions (Children's Auditory Discrimination Inventory, Mimicry Final Word Sounds and Seguin errors), significant age differences.* Thus, general information-processing skills, conceptual understandings, and favorable responses to the testing situation were greater for older children and for those from families of higher socioeconomic status. Not surprisingly, these findings indicate that cognitive-intellectual performance at this age period is a function of both developmental level and experience. Given the diversity of tasks represented on the first factor, age and SES were shown to influence a wide variety of behaviors. Differences were manifest on verbal, quantitative, and perceptual tasks. Nonverbal as well as verbal performances were affected, although for this sample at this age a verbal-nonverbal distinction may be difficult to make, since the verbal component of any task performance might

*Of these three, only the Seguin failed to show any age difference; the other two tasks showed marginally significant effects for age ($p<.05$).

be substantial. For example, the child could not point to the picture that matched the stimulus without understanding the meaning of the instruction.

Age and SES may be viewed, however, as differentially producing these effects. Given the relatively short age span, only three measures showed larger age than SES effects--Child Cooperation, Vigor, and number correct on the Enumeration task. These are all among the few measures which showed significant sex differences and, in accord with our later discussion of such differences, these results may indicate greater compliance and task persistence along with superior motor coordination in older subjects. On the other hand, those measures which showed SES effects which were largest relative to age differences were the Eight-Block Sorting Task, the Motor Inhibition Test scores, Sigel grouping responses, and Story Sequence. All these appear to require not only careful attention to verbal instruction, but also the demonstration through another modality of a verbally-based understanding of the task, or, conversely, verbal explanation of the child's own nonverbal performance. This difference suggests that higher SES is associated not only with a greater number of experiences--as age would be--but also with differences in the cognitive organization of these experiences.

Those tasks which appear to require more active environmental interchange showed larger SES and age differences. Thus, tasks requiring knowledge of specific information (e.g., Preschool Inventory) and communication skills (Peabody B) showed larger differences than those requiring form discrimination and matching or comprehension of syntax. The Caldwell was a particularly difficult test for these children. Mean performance for the total group was 27.9. In comparison with data reported for the Head Start pretest administration in the Fall of 1968, study children in a comparable age band (4-0 to 4-5) obtained a mean score of 27.7 vs. 30.0 for the test standardization sample. However, in looking at the performance of children from blue-collar workers' families only, mean performance was 24.3.

Particularly affected by SES were those measures assessing the child's ability to provide sorting rationales. Although approximately 50% of the children were able to categorize in the Toy Sorting Task, few were able to verbalize their reasons for doing so. As was found in previous research with this task (Hess & Shipman, 1965), differences were greater when the child was required to give a color rationale which taps abstract and categorical use of language as opposed to denotative and labelling usage. In the Eight-Block Sorting Task, the majority of children placed the test blocks correctly, but approximately 20% verbalized one dimension and only 11% both dimensions. Similarly, although most children were able to label the objects correctly on the Sigel Object Categorization Test, few were able to give appropriate verbal rationales for their sorts. When reasons were offered, they often had no discernible relationship to the grouping characteristics. Similar findings have been reported for low SES children by Hess and Shipman (1965); Melton et al. (1967); Meyer (1971); and Sigel and his associates (1967). These results would concur with Cazden's (1968) statement that basic grammatical structures seem to be learned despite differences in the child's linguistic environment; the manner in which children use language to express ideas, however, may be more sensitive to environmental manipulation.

Mean differences between groups were similar for the various perceptual tasks, but variations in performance level did suggest an ordering in complexity ranging from those tasks primarily involving form discrimination (Johns Hopkins and Matching Familiar Figures) to those requiring analysis (Preschool Embedded Figures and Picture Completion) and copying skills (Form Reproduction). Analysis skills would appear to require prior mastery of discrimination, which, in turn, presupposes figure-ground separation. Consequently one might expect these skills to be developmentally ordered, with the more complex functions developing at later ages than the simpler ones (Birch, 1963, 1967). The Preschool Embedded Figures Test and the Picture Completion Test proved to be of considerable difficulty. Moreover, although the children were able to differentiate between simple geometric figures, form reproduction appeared to be a much more slowly developing skill. Maccoby (1967) has stated that while holistic perception may suffice for a simple discrimination, it will not for making a copy; instead there must be perception of elements of a figure in addition to the whole.

A few measures loading on the first factor also showed significant sex differences: Cooperation rating, Preschool Inventory, Form Reproduction, Mimicry Nonsense Words, Enumeration, and Vigor. In accordance with previous findings, girls were rated as more compliant and task-oriented. The fact that girls performed significantly better on Enumeration and not on Spontaneous Numerical Correspondence suggests that this result is more attributable to coordinated sustained attention than to any basic difference in understanding of number. As pointed out in the Enumeration Test task description, in a repetitious task of this sort, style and persistence also play a major part in determining the score.

Sustained attention and compliance are also critical aspects of Mimicry test performance. It also has been noted (Bever, 1970) that perceptual strategies and auditory dominance develop earlier in girls with resulting facilitating effects on consonant perception. The fact that significant sex and age effects were shown only for the Nonsense Words score and not for the Real Words Final Sounds score may have been due to the different task demands; that is, the short phrases used in the real word items may have introduced a memory component for this age group. An equally plausible hypothesis, however, is that the difference is due to the relatively lower reliability that was obtained for the Real Words Final Sounds.

The significantly higher test performance of girls on the Caldwell and Form Reproduction items may reflect differential instruction in the home since both tasks would be highly sensitive to differences in training and practice. As observed in the mother-child interaction situation, girls appeared more attentive to the mother as a teacher. Preliminary findings from the interview data (ETS, PR-70-20) revealed a small but consistent trend for mothers of girls to be more involved in school relevant activities (e.g., reading to their child). These results also may 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 et al., 1968; Moss, 1967).

Unlike the above skill measures, boys obtained higher scores on the Vigor measure. Moreover, SES differences were relatively small as compared to age differences. These findings suggest that physical coordination and/or maturation was a relevant component of crank-turning performance for this sample at this age period. The significant sex difference obtained may reflect effects of sex typing, with greater reinforcement being given to boys at this age for more active and assertive behaviors. These findings, plus the low correlation of the Vigor score with other measures, suggest that only a small component of vigor performance is loading on Factor 1.

In contrast to the above findings, the three latency measures that defined Factor 2 showed no significant age or SES effects. Sex effects also were non-significant, although boys tended to have longer latencies on each of these measures. Thus the orthogonal relationship between the cognitive competency and tempo factors was paralleled by an apparent lack of similarity in the relationship of scores to age, sex, and SES.

The third factor in the 6-factor promax solution, defined by the two scores from the Spontaneous Numerical Correspondence task, showed only significant age effects. Since it was assumed that these abilities were undergoing rather rapid change during this age period, the potency of the age variable was expected. The data from this task were in accord with Piaget's view (1952) that the understanding of number at this preoperational stage is essentially perceptual in character, reflecting global rather than articulated intuition processes. Most children found the task difficult and matching to configuration was easier than matching to number. According to Kohlberg's (1968) analysis, Piagetian tasks are assumed to be relatively insensitive to specific instructional experiences and thus SES effects would have been expected to be minimized. It should be noted that the other quantitative measure, Enumeration, which is closer to computational skills and loaded on the first factor, did show SES effects and, assumingly, the greater effect of experiential differences.

The other Piagetian-derived measure, the Boy-Girl Identity Task, defined a separate factor (Factor 6) and showed a different pattern of effects than that for Spontaneous Numerical Correspondence. Paralleling its relationship to the first factor (r = .36), performance on the "wish" items showed SES effects. Boys gave significantly more constancy responses on the concrete Girl items, but since this was not replicated with the Boy items and no other significant inter-relationships among measures were obtained, such responses seem more appropriately described as "pseudo-constancy." This interpretation is supported by the fact that although children above their age-group mean on the Preschool Inventory were more likely to give a constancy response to the "wish" items, children who scored lower on that measure were more likely to give constancy responses to items two through five. Since developmental trends were not obtained with this task at this age period, the SES results may reflect differences in expectancies of the fulfillment of one's wishes rather than in achievement of a reality judgment on gender identity constancy.

Factor 4 was defined by four of the Open Field Test measures. Three of these showed no significant main or interaction effects, but boys obtained a significantly higher mean play complexity score, and SES differences in talking to the tester were significant. The former finding is consistent with previous studies investigating curiosity which have reported boys as showing greater preference for complexity when the behavior is manipulation rather than visual attending (Lucco, 1964). Since talking to the tester includes both task-related verbalizations (e.g., requests for help, directing the tester's attention to the task) and non task-related verbalizations (e.g., talking about his family) and could reflect positive or negative feelings about the task activity, no interpretation of these data can be offered at this time. Talking to self, which loaded only on Factor 6 (.32), showed a significant sex effect, with boys talking more. Further analysis of the data revealed that this was true for both task-related and non task-related speech.

The Fixation Test scores of mean recovery and habituation defined Factor 5 but showed no significant main effects or interactions. Future analyses will investigate effects on duration of attending and differential responsiveness to the nonsocial and social stimuli.

The two remaining scores, Mischel choice and Risk-Taking, were previously noted as having little, if any, relationship to other measures in the test battery. Similarly, they showed a different pattern of effects. No significant effects for the Mischel were obtained. As discussed earlier in this chapter, many factors other than the ability to delay gratification are represented in the child's performance and make interpretations at this time difficult. The majority of children chose the delayed reward, but as noted in the task description, those who chose the immediate reward were somewhat more likely not to give a reason for their choice or to offer an irrelevant or egocentric response. Risk-taking showed a significant sex effect, with boys more likely to choose the uncertain event. Those children who feel more capable of manipulating their environment may be more willing to take risks. The fact that a significant sex effect was obtained may reflect differential reinforcement of assertiveness and daring for boys and girls at this age. Given the paucity of present data in the affective domain, however, such hypotheses must be regarded as highly tentative.

The Brown self-concept score, which loaded .28 on the first factor only, showed age and SES effects as did the other scores loading on Factor 1. One should be cautious, however, when interpreting such findings to mean that children who are older or who come from families of higher socioeconomic status feel better about themselves. For children of this age, the findings may reflect instead differences in comprehension of the task and in ability to make a differential response about oneself. Moreover, the findings of Clark et al. (1967) indicate that self-perception responses to pictorial stimuli may be different from those made to verbal material covering the same content. Smiling for the photograph taken for the Brown test did not load on any of the six factors, but it did show a significant sex effect. Although the majority (67%) of children did not smile, girls smiled more than did boys. This may reflect greater ease in the situation and/or differential learning of social roles.

Mean Differences Among Head Start Eligibility and Preschool Attendance Groups

The ANOVAS obtained for Head Start Eligibility by Preschool Attendance categories were highly similar to those reported above (see Table 4-10). Differences in family income (eligibility) produced results similar to those obtained for differences in parent's occupation (SES). Where no significant effects had been reported (Table 4-9) (as, for example, for Fixation Test scores, Mischel choice, or latency measures), eligibility and preschool attendance were not shown to produce significant differences (Table 4-10). Similarly, when sex alone was significant (i.e., Risk-Taking, Open Field, Boy-Girl), Head Start eligibility was not a significant main effect. The one exception to this general conclusion was smiling for the Brown photograph. Although significant differences by SES had not been found, low income children

Table 4-10

Significant Analysis of Variance Effects
by Head Start Eligibility and Preschool Attendance[1]

Score	Factor[2]	E	PS	ExPS
1	1	X	X	
2	1	X	X	
3	1	X		
4	1	X	X	
5	1	X	X	
6	1	X	X	
7	1	X	X	
8	1		X	
9	3			
10	3			
11	1		X	
12	1	X	X	
13	–			X
14	1	X	X	
15	1		X	X
16	2			
17	1		X	
18	–			
19	1		X	
20	4			
21	6			
22	4	X		
23	4			
24	4			
25	4			
26	1	X	X	
27	1	X	X	
28	1	X		
29	2			
30	1	X	X	
31	5			
32	5			
33	1	X	X	
34	1	X		
35	–	X		
36	1		X	
37	2			
38	1	X	X	
39	1	X	X	
40	1	X	X	
41	6		X	
42	6			
43	–		X	
44	6			
45	1	X	X	
46	1	X	X	

1. X appears as an entry when the designated effect is significant ($p < .01$).
2. Numerical entries represent factor with highest loading above .30 on 6-factor promax solution; a dash indicates that score did not load above .30 in this solution.

were much less likely to have smiled when their photograph was taken. This may reflect the lower probability of the low-income family owning a camera and thus familiarizing the child with having his picture taken.

When comparing preschool attendance categories, it was found that those children who later attended Head Start were performing at a significantly lower level than those who attended other preschools. Mean performances among groups generally ordered themselves as follows, from low to high: Head Start eligible, attended Head Start; Head Start eligible, not known to have attended preschool; Head Start ineligible, attended Head Start; Head Start ineligible, not known to have attended preschool; and Head Start ineligible, attended other preschool. (Given the extremely small cell size for the Head Start eligible--attended other preschool category, no comparisons with these data can be made at this time.) Again, differences were obtained only for those tasks assessing cognitive competencies.

Summary of Findings on Mean Differences

The results of the ANOVAS thus paralleled the findings from the structural analyses. Those measures loading on the first factor and defining a general ability dimension showed significant age and SES effects, despite the relatively restricted ranges of both variables. SES effects, in general, were substantially larger than those for age--for the twenty-three measures showing both age and SES effects, age differences between those above and those below the sample mean ranged from .13 to .66 standard deviations, with a median of .33; while SES differences between blue-collar and white-collar families ranged from .21 to .98 standard deviations, with a median of .55. SES effects were larger than age effects for 20 of the 23 measures. Head Start-eligible children who later attended Head Start performed significantly poorer on these measures; both eligibility and later (self) selection for preschool experience were associated with performance differences. Measures defining other factors showed different patterns of effects. There were few significant sex differences; those that were obtained showed girls performing better when sustained attention was required, whereas boys gave the preferred response when assertiveness seemed a more critical component. It should be noted that the usual finding of superior performance by girls on verbal measures was not evidenced in these data. Girls generally obtained slightly higher scores on the verbal tasks, but for this sample at this age these differences were not significant. No such trends were present for the perceptual tasks. Nor did boys perform better on tasks tapping analytic performance, as has been found in previous research (Kagan, Moss, & Sigel, 1963; Sigel, 1965). Whether such sex differences in performance on verbal and perceptual tasks will emerge in subsequent years remains to be determined. With one exception, a Sex by Age by SES effect for Seguin time to quickest solution, there were no significant interaction effects. Age and SES did not show different effects for boys and girls, nor did SES have different effects for younger and older children. Also, new clusterings of tasks according to patterns of effects were not obtained.

In these analyses, SES (as defined by the occupation of the head of the household) and Head Start eligibility (as defined by income per household size) served as gross proxies for describing the child's environment. Future analyses will include further delineation of processes in the child's home environment provided by information from the parent interview and structured mother-child interaction sessions. Such information should provide more direct evidence on the environmental interactions accounting for differences in these test performances.

Chapter 5 - CONCLUSIONS

Structural analyses of the Year 1 child test data yielded (a) a general ability dimension (i.e., information-processing skills) cutting across contents and operations sampled in the cognitive test battery and (b) a stylistic response tempo dimension. Additional factors apparently tapped task-specific styles and behaviors. Although previous research has reported differentiated abilities in very young children ranging down to ages two and three (Hurst, 1968; Meyers, Dingman, Orpet, Sitkei, & Watts, 1964; Meyers, Orpet, Atwell, & Dingman, 1962; McCartin & Meyers, 1966), and several cognitive dimensions have been delineated in analyses of infants' and preschool children's performance on standard preschool scales (Stott & Ball, 1965; Ramsey & Vane, 1970), primary factors differentiated by content, operations, and/or products did not emerge in the present data. Given the generally high internal consistency of the various scores and their moderate to low communalities, considerable non-error specific variance remained. These findings were strikingly consistent across a variety of statistical methods and across major subject classifications (i.e., by sex, age, SES, preschool attendance controlled for eligibility, and Preschool Inventory score).

Many theorists, including Piaget and Guilford, emphasize the importance of interactions with the environment for intellectual development. Although the child may start with certain innate mechanisms, such as predispositions for Guilford's five operations or Piaget's invariant functions of assimilation and accomodation, the rate of progression and the variety of dimensions in cognitive functioning appear to depend upon the extent to which these mechanisms are exercised in interaction with a varied environment (Hunt, 1961).

Ferguson (1954, 1956) has suggested that cognitive factors represent behavioral domains that happen to have been learned together, along with those similar behaviors that become associated through generalization of learning and transfer. As Messick has pointed out (ETS, PR-68-4), some of the determinants of these shared learnings are developmental, in the sense that certain things are experienced together because they are appropriate to particular ages, but most of the determinants appear to be more directly sociocultural (Lesser, Fifer & Clark, 1965). It would be expected from a transfer theory of abilities that factor structures would be more clearly defined for subjects having had the benefit of more varied experience. Thus, the absence of differentiated factors in the present data may reflect a relatively narrow range of environmental variations experienced by this sample as compared to subjects assessed in previous studies. For example, Meyers et al. (1964), Stott and Ball (1965), and Ramsey and Vane (1970), who tested children from primarily middle-class and upper-class socioeconomic backgrounds, found greater evidence for differentiation at this age or even earlier.

There was a substantial amount of variance in the present data, however, that was not part of the general ability dimension. Using unities in the diagonal, less than one quarter of the variance for the total test battery was accounted for by "g." The lack of clustering prevents us from knowing at this time if such specific variance is related to special abilities limited to one task, incomplete sampling of the processes represented by tasks, or particular situational determinants. The fact that a general ability dimension or "g" was somewhat less evident in less mature subjects (i.e., in younger children and those performing below their age-group mean on the Preschool Inventory) suggests that for this sample at this age behaviors were being tapped at the beginning of a period of integration rather than during a period of differentiation. Subsequent measurement might be expected to reveal increasing common variance on "g" followed later by increasing differentiation in terms of contents, operations, and/or products. Discontinuity in cognitive structure would thus be indicated by changes in the number or size of dimensions over time and/or by changes in the meanings of dimensions as revealed in new patterns of correlates or factor loadings.

The tentative finding of a battery-defined clustering of cognitive measures suggested the role of situational determinants (i.e., day and tester) as a secondary structuring variable. Perhaps children from more restricted environments are less familiar with test-like settings and, therefore, show greater variation in performance across days. Some children may increasingly adapt to testing and generalize skills learned in the testing situation while others, especially those who find the tasks or the testing center situation more difficult and frustrating, may become increasingly alienated from the situation. This interpretation suggests the possibility of analyzing the data by categorizing groups of children by extent of consistency over the four testing days. Comparative data for groups of children so defined would, of course, have to be controlled for tester and order of administration of the test batteries. This battery effect, however, was least for those measures with highest loadings on the general ability factor. The child who appeared particularly able in one assessment was generally able in another.

The less clearly defined general ability dimension which was found for the younger, less "academically" prepared, and economically poorer children in the sample may reflect both greater susceptibility to situational determinants and less generalizable information-processing skills. That is, they may have been less able to apply what they know, or, in Piagetian terms, to "decenter." The difference in generalization of these skills would seem to reflect differential training and practice in the various task components as well as in the transfer of skills from one task to another. These data would suggest that the differentiation seen in the present data is task-specific and may not be under the control of generalized cognitive mechanisms. With continued practice and experience, cognitive mechanisms may come into play that will give order and consistency to these behaviors. It may be that, only after such integration and generalization of the specifics occur can differentiation into stable cognitive factors take place. Longitudinal data will enable us to chart such developmental patterns and assess the differential utility of various theories such as Garrett's (1946) hypothesis of a single general ability that differentiates over time or the more general notion that cognitive structure tends to become increasingly differentiated (and hierarchically integrated) during the course of development, as propounded by Werner (1957) and Lewin (1951).

A major question posed in the study was the relationship between cognitive style and skill. The concept of ability implies measurement of capacities in terms of maximal performance, whereas the concept of style implies measurement of preferred modes of operation. Both are necessary for a full understanding of cognitive functioning (Cronbach, 1960). Some controlling mechanisms represent dimensions of individual differences in the structural characteristics of the cognitive system itself. Other controlling mechanisms appear in the form of preferences or information-processing habits, which determine a person's typical modes of perceiving, remembering, thinking, and problem-solving. In the Year 1 test battery an attempt was made to assess the cognitive styles of reflection-impulsivity and analytic functioning. There were too few scorable sorting rationales given on the Sigel to assess differences in preferred categorization style at this age. Data from the Sigel in future years of the study should enable us to assess not only the child's classificatory ability over time but mode and stability of response style.

A general dimension defined by the three latency measures did emerge. However, the orthogonal nature of the tempo dimension to the general ability dimension suggested that for this sample at this age temperamental components have not yet been integrated into the cognitive domain. The orthogonal relationship between cognitive competency and tempo factors was paralleled by an apparent lack of similarity in the relationships of scores to age, sex and SES. Moreover, other possible indices of the impulsivity dimension--the ability to inhibit a response when appropriate and the ability to delay gratification--were not related to the latency scores or to one another.

In accord with recent findings (McGaw & Jöreskog, 1970; Wasik & Wasik, 1969), the factor pattern was relatively constant over a range of socioeconomic and ability levels. Inspection of standard errors of the means and patterns of correlates of the measures in the Year 1 battery indicated no major differences in construct validity of the tasks for the major classifications used in these analyses. Similar results were reported by Stevenson et al. (1971) who found the pattern of interrelationships among learning tasks for four- and five-year-old low-income children to be similar to that reported for middle-class preschoolers.

Mean performance levels did show significant group differences, however. Performance on those tasks defining the general ability dimension was shown to be a function of developmental level (age) and experience (socioeconomic status), despite the relatively restricted ranges of both variables.

Seventy-five percent of the study children who were eligible for Head Start did attend Head Start subsequent to our initial testing. According to the present results these children generally were performing significantly less well on a variety of cognitive-perceptual tasks prior to their enrollment. Such effects were particularly evident on tasks highly saturated with "g," including those tapping acquisition of information considered necessary for success in school (Preschool Inventory and Peabody) and use of language as cognitive tool (e.g., classificatory skills). Piaget has argued that classificatory structures which emerge during the preschool years are an essential foundation for later concrete operations (Flavell, 1963). Longitudinal data may enable us to assess the extent to which retardation in the development of logical operations in socioeconomically disadvantaged children may be due to inadequate foundations for such processes.

Those children who later attended other preschool programs were found to be more advanced prior to entering any preschool program. As pointed out in Chapter 2, however, the variables defining the several groups are confounded and thus, no simple main effect comparisons for classificatory variables such as Head Start vs. non-Head Start can be made without careful consideration of their interactions with other variables. For example, preschool attendance is confounded with site, race, and the four indicators of socioeconomic status (mother's and father's educational and occupational levels). Thus, to interpret simple mean differences for Head Start vs. non-Head Start groups would be quite unwarranted.

Assignment of tasks to the "cognitive domain" does not imply they are independent of motivation. For the young child especially, one cannot separate intellective and non-intellective factors. Specifically, motivational factors cannot be separated from the learning process. As was found, measures of persistence and cooperation loaded on the general ability dimension (cf. Zigler & Butterfield, 1968). Although there was probably insufficient sampling of such behaviors in the test battery to produce factors in the affective domain, the affective domain may not be highly differentiated at this age. However, there was rather strong evidence for differentiated personal-social characteristics of these children when seen later in the classroom setting (Emmerich, 1971).

For this sample at this age "non-cognitive" measures did not show the same SES differences found on the measures of cognitive competency. The majority of children were willing to choose the uncertain outcome, to accept the delayed reward (which also might be seen as an uncertain outcome), and to express positive statements about themselves. Consistent with the findings of other investigators (Brown, 1966; Clark et at., 1967), self-concept scores were predominantly high. This may reflect the fact that "significant others" at this age are primarily family members and particularly the mother, whom we might expect the child to perceive in a supportive role.* As the child grows older with increasing opportunity for interaction with others in a variety of situations, we would expect a more differentiated concept of self to emerge, resulting in greater variance of scores. For many low-income children, especially those of minority status, such interactions may lead to negative self-evaluations and markedly lower scores (Katz, 1969; Tannenbaum, 1967).

The above discussion should not lead us to overlook the most striking finding of all, namely the wide range of variation in performance exhibited.

*In subsequent analyses we will look at the child's self concept score in relation to the mother's supportiveness in the interaction tasks.

Although group differences were statistically significant, many mean differences in task performance were small relative to within-group variability. Low-income youngsters are not a homogeneous group. Children from low-income families span a much wider range of cognitive, personal-social and perceptual functioning than some would have us believe. The fact that the same factor pattern was found within the low SES group reflects this finding also.

In general, tasks in the Year 1 battery 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. Future analyses will investigate influences of tester characteristics on child performance. 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.

It should be noted that the present analyses imply a linear model of cognitive growth. Emmerich (1968) has made a distinction between "differential" and "classical" views of sequential structural development. The differentiation hypothesis, which assumes increasing differentiation of the cognitive domain with development, was discussed in the first section of this chapter. However, there is also the possibility that individuals pass through a developmental sequence of qualitatively different structural organizations, usually held to be in an invariant order, which is the more classical developmental view of stage progression. Several theorists have postulated such a developmental sequence of stages, usually involving three major phases that encompass similar phenomena from theory to theory but are labelled in somewhat different terms -- such as sensorimotor, perceptual, and conceptual (Werner, 1948); enactive, ikonic, and symbolic (Bruner, Olver & Greenfield, 1966); perceptual, imaginal, and conceptual (Thurstone, 1926); or sensorimotor, preoperational, and operational (Piaget, 1950). The emphasis in measurement would then be upon the assessment of qualitative features that are characteristic of particular stages of cognitive functioning and upon ordered sequences of tasks capable of gauging the transition from one stage to another.

The present analyses were based primarily on total scores which might mask differences in patterning or level of response; data reduction entails a risk of losing critical information. Moreover, few "markers" of stage level are represented in the Year 1 battery. An exception is the Sigel Object Categorization Test; analyzing the performance of children differentiated into three groups that may represent different developmental stages -- i.e., above the median on nonscorables, on color responses, or on form responses -- might give clues as to developmental stage. Recent findings reported by Katz (1971) suggest that the change from color to form sorting reflects a change in the tendency of children to go beyond perceptually dominant stimulus characteristics and to analyze, reflect upon, and use alternative dimensions. Younger children, because they tend to process impulsively and do not decode all relevant stimulus information, respond to color on color-form tests more than reflective children. Thus, changes in cognitive tempo associated with age (not age per se) may account for differential responding.

Those tasks that might have yielded scores representing different levels on a developmental scale (e.g., Boy-Girl Identity and Spontaneous Numerical Correspondence) did not do so at this time. The Boy-Girl Identity Task did not tap a cognitively based reality judgment of gender identity constancy in this population at this age, but instead yielded four reasonably orthogonal scores; children's performance on the Spontaneous Numerical Correspondence Task indicated that understanding of number at this preoperational stage was essentially perceptual in character, reflecting global rather than articulated intuition processes. Thus, the study children were generally preoperational. Measurement in subsequent years on these tasks might provide such scaled scores. Also, additional tasks amenable to such differentiated scoring have been included in later test batteries. The later use of conservation items with the Spontaneous Numerical Correspondence Task will enable assessment of changes associated with shifts from preoperational thinking to concrete logical operations.

Efforts will continue in deriving other, more sensitive, indices of level of performance and of task sequences. The generally quick test responses of this sample suggest that at this age most subjects could not inhibit long enough to enable cognitive processes to operate optimally in contexts where greater reflection would be functional. As suggested recently by Eska and Black (1971), future analyses may fruitfully separate children into "reflective" (long latency, low errors), "impulsive" (short latency, high errors), "quick" (short latency, low errors), and "slow" (long latency, high errors) groups in order to obtain a better understanding of the factors which influence and/or determine a particular response style. Moreover, given the likelihood that response sets are particularly important in the responses of young children (Damarin & Cattell, 1968), further internal analyses of tasks will be directed toward investigating such effects. For example, it was noted in describing the findings from the Children's Auditory Discrimination Inventory that children showed a differential preference for pointing to the real as contrasted to the nonsense picture. For these analyses, in which children would be grouped on various attributes, utilization of inverse factor analytic and other clustering techniques might be explored.

Further understanding of the present data will be provided by mapping out similarities and differences in sociocultural determinants. Planned analyses will assess the extent to which task-specific variance separates variously defined groups, both with and without partialling out "g." Tasks loading on the first factor range from general to specific, and common dimensions may be defined where shifts in determinants occur. Configurations of the data provided by the smallest space solutions could define clusters of behaviors with similar sociocultural determinants. Moreover, the variety of measures included in the study enables one to examine the components of related but not identical constructs (e.g., cognitive styles as defined by Witkin (1967) and by Kagan and his associates (1964).

The findings of McGaw and Jöreskog (1970) in studying the factorial invariance of ability measures in high school subjects differing in intelligence and socioeconomic status suggest the value of looking at SES differences within ability level. They found the gap in mean factor scores between low and high-SES groups to be much wider for low IQ than for high IQ; the facilitative effect of high SES thus appeared greater for low IQ subjects. Similar findings have been reported by Willerman et al. (1970) for infants and preschool children.

The above discussion illustrates some analyses presently planned to help tease out complex interrelationships among variables that must be investigated before one can understand the complexity of the child's functioning in the test situation. The next report will describe cross-method (i.e., data from tests, interviews and interaction situations)--cross-domain analyses essential to delineating underlying processes. Present analyses used occupation of head of household and income as gross proxies for assessing the child's environment. These indices assume constancies of meaning within and across groups, and they tell us little about the type of stimulation the child is being exposed to in the home environment. Within a given SES level, the range of home environments can be so great as to make any generalizations about SES level and development extremely tenuous (Pavenstedt, 1965; Tulkin, 1968; Zigler, 1968). More fine-grained analyses will become possible using indices from the parent interview and mother-child interaction sessions. For example, we will look at the effects of variation in experience on the child's ability to use language as a tool for symbolic or representational thinking. The present data as well as those recently reported by Golden et al. (1971) suggest that the effects of variety of experience are particularly salient for those behaviors reflecting the cognitive use of language. In the first project report (ETS, PR-68-4), Shipman and Bussis delineated other process variables in the child's environment that appear to be particularly influential in the child's cognitive development, especially the role of the mother in selecting, structuring and transmitting information about the environment to her child and in regulating his behavior in relation to both the environment and the information transmitted.

Data in subsequent years will enable us to evaluate hypotheses generated by the present structural findings; especially whether there is increasing integration of cognitive behaviors followed by differentiation into clusters, as those found in previous research. We will also be able to assess the effects of differential experiences provided by Head Start and other preschool programs. Such experiences would be expected to show differential effects depending on the nature of the processes involved and the level of the child's functioning.

APPENDIX A

DATA ANALYSIS PROCEDURES

F4STAT

F4STAT is a set of building blocks with which general or special purpose programs can be assembled with ease and without the user of the resulting program necessarily being aware of the existence of the F4STAT. F4STAT is a system of subroutines, compatible with both Fortran and Cobol main programs. This system was originally designed in 1964 for the IBM 7040 and was translated to the IBM 360/65 in 1968. It has been in continuous operation at ETS in both production and research functions since 1965 and is used by both Cobol and Fortran programmers. This long experience with operational aspects assures its freedom from programming errors.

The F4STAT system recognizes that mathematical manipulation is but a small part of data analysis. It is not unusual for the creation of a data file, including its editing, to be far more expensive and time consuming than the statistical manipulations. For this reason, F4STAT has a number of special editing features as well as a capability of handling complex files. It also has an extensive collection of statistical procedures. Both the data and statistical processes are based on unique basic modules. Data and files of data are operationally defined by procedures rather than aesthetic symbols in a computer memory. The operational definition allows not only complex transformation but also complex file organizations. The statistical processes are built on concepts of special matrix operators described by Beaton (1964). The combination of these facilities permits the same language to be used for both data manipulation and mathematical analysis, and indeed it is customary to do both in the same task over a data file. This process minimizes the number of reads and writes performed, minimizes temporary data files, and minimizes computer costs.

The design considerations in F4STAT are as follows:

1. Accuracy

The F4STAT subroutines are continually undergoing accuracy checks to assure the adequacy of the arithmetic in our programs. Whenever F4STAT can identify computational problems (e.g., multi-collinearity), it warns the user of possible problems. All additions to F4STAT are very carefully checked before admission to the system.

2. Generality

A statistical system must be able to perform any computable statistical analysis presented. Over the years, F4STAT has grown to include many subroutines for descriptive statistics, correlation analysis, factor analysis, and the other techniques mentioned in the data analysis section. The basic building blocks described by Beaton (1964) assure the very wide variety of application. The system is also flexible in that newly developed algorithms can be added.

3. Efficiency

In working with large data bases, efficiency is a very important factor. F4STAT is considerably faster than other statistical systems largely due to its efficient use of storage; that is, F4STAT does as many operations as possible in one pass over a data base. Its algorithms are efficient as well as accurate. Such properties are especially important in large analyses such as these.

4. Ease of Use

This project depended upon flexible, easy to use programs to do specific tasks with the data base. This was accomplished by building tailor made, easy-to-use programs for specific tasks from the F4STAT repetoire.

5. Transportability

Since Fortran is the basic language of F4STAT, the package is easily transportable to other computer systems.

6. Output Readability

F4STAT routines allow labeling of output so that a user may have the mnemonic labels as assistance in interpreting the output. There are a number of very general output routines as well as service routines for tailoring output for specific needs. Examples of some output of descriptive statistics are shown in the analysis section.

Derivation of Tucker's Procedure for Estimating Communalities

Professor Ledyard R. Tucker has suggested a modification of Thurstone's procedure of using the highest correlation to estimate the communality of a variable. A brief derivation of this estimate is presented here.

Given the factor analysis model

$$r_{jj'} \doteq \sum_k a_{jk} a_{j'k} ,$$

where $r_{jj'}$ is the correlation between variable j and variable j' and a_{jk} is the loading of the j^{th} variable on the k^{th} factor. If variables j and j' are co-linear in the factors, then an orthogonal rotation exists such that

$$r_{jj'} \doteq a_{j1} a_{j'1} + 0 .$$

Assume that variables j and j' $(j \neq j')$ are co-linear in the factors, where j' is the variable which correlates most highly with variable j. Then,

$$|r_{jj'}| = |a_{j1}||a_{j'1}| + 0$$

$$|r_{jj''}| = |a_{j1}||a_{j''1}| + 0 , \quad \text{where } j \neq j''$$

$$|r_{j'j''}| = |a_{j'1}||a_{j''1}| + 0 , \quad \text{where } j' \neq j''$$

and,

$$\sum_{j''} |r_{jj''}| = |a_{j1}| \sum_{j''} |a_{j}$$

$$\sum_{j''} |r_{j'j''}| = |a_{j'1}| \sum_{j''} |a_{j''1}| .$$

Solving for a_{j1}^2:

$$|r_{jj'}| \frac{\sum_{j''} |r_{jj''}|}{\sum_{j''} |r_{j'j''}|} = \frac{|a_{j1}| \sum_{j''} |a_{j''1}|}{|a_{j'1}| \sum_{j''} |a_{j''1}|} |a_{j1}| |a_{j'1}|$$

$$|r_{jj'}| \frac{\sum_{j''} |r_{jj''}|}{\sum_{j''} |r_{j'j''}|} = a_{j1}^2$$

a_{j1}^2 is then used as the estimate of the communality for variable j.

APPENDIX B

TASK DESCRIPTIONS

NOTE

Appendix B presents abbreviated task descriptions for child measures administered in Year 1 which were in the analyses for this report. The information on each task is intended to serve as an outline; data will appear in more comprehensive form as a compendium of Technical Reports as a part of the next report. The present task descriptions provide information on general task rationale, administration and scoring procedures, and score properties for those scores used in the present structural analyses. Also, sample characteristics are presented as defined by age and sex categories. The information in these tables typically includes the number of observations, the mean, the standard deviation, and the percentage of children responding for each possible response category (as identified in the table headings). Percent Response is replaced by Percentiles (10th, 25th, 50th, 75th, 90th), where continuous scores are used. The percent response option may provide a total cell count different from that used in the total mean, since cases of tester errors or children's refusals are included in the response total but not in the total mean. Where percentile intervals are computed, testers' errors and children's refusals are excluded from the total count so that the total in the percentile cells agrees with the total count for the mean. Also, in comparing the number of subjects by task, either across classifications or by composite totals used in deriving indices of reliability, the total N's for a task may vary slightly due either to the above adjustments or to periodic updating of master file information.

Boy-Girl Identity Task

Purpose

The Boy-Girl Identity Task was designed to assess the child's cognitive ability to preserve gender identity constancy despite changes in stimuli which increasingly resemble the opposite sex. The present instrument is a refinement of a technique introduced by Kohlberg (1966a) and used by De Vries (1969) in her study of bright middle-class boys.

Task Description

The present version of the instrument consists of two 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 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 Johnnie 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.

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 was ambiguous, the item was scored 0.5. (There were relatively few instances of such partial scores.) If the child exhibited constancy on a particular item, he was asked to give reasons for his response to that item. However, because of the generally low level of constancy achieved by the present sample at this age (see below), these explanations have not been subjected to a content analysis.

Properties of the Total Scale

Each S's total constancy score was the mean of scorable items for that S (usually 10 items). The average total score for the total sample was .22, with a standard deviation of .19 (N=1330). Thus, on the average, constancy occurred on about one out of five items; a considerably lower base rate than that found previously in bright middle-class children at this age.

De Vries (1969) found that bright middle-class boys of three years have some competence on this kind of task and that four-year-olds have attained considerable competence. Also, our own pilot testing found that gender identity constancy was maintained on the present instrument for about half of the items in a small sample of middle-class four-year-olds. Moreover, the present version of the instrument incorporated technical improvements which, if anything, should have made the task easier than the earlier Kohlberg and De Vries versions.

Internal consistency analyses (see below) and the fact that the total score did not increase monotonically with age in this sample suggested that this measure generally did not assess the cognitive achievement of gender identity constancy in this sample at this age, but rather was tapping certain preconstancy processes. As anititial check on this conclusion, item difficulties were computed at each of several total score levels separately by sex of subject. It was reasoned that once item difficulty ranks become similar across the boy and girl stimuli and across boy and girl subjects, we would be dealing with a total score level representing minimal competence in cognitively based gender identity constancy. It was found that (comparable) items were at different levels of difficulty across the two tasks and sexes within different total score levels below 50% correct. Also, below this cutting point, constancy on some items failed to increase monotonically with increasing total score level! Thus, the total scale score did not appear to have the same meaning across the tasks, sexes, and total score levels.

Item and Subscale Properties

Item and internal consistency analyses revealed the presence of four subscores which are reasonably orthogonal to one another in this sample at this age. The first component is the child's response to item I-1 ("If Janie really wants to be a boy, can she be?"). The second is the child's response to item II-1 ("If Johnny really wants to be a girl, can he be?"). The third 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 fourth index is the child's summed score on items II-2 through II-5, signifying constancy of the boy stimulus despite suggested changes in activity, clothes, and hair style.

Items I-1 and II-1 were uncorrelated with all other items (highest r=.14), although they were correlated with each other (r=.45). Subscores I-2-5 and II-2-5 had KR-20's of .59 and .64, respectively, but were uncorrelated with each other (r=.01). While the positive correlation between I-1 and II-1 indicates that they share common variance, it may be useful to keep them as distinct measures in future analyses of these subscales.

Conclusions

The total constancy score did not tap a cognitively based reality judgment of gender identity constancy in this population at this age, and, therefore, great caution should be exercised in its use. Moreover, internal analyses revealed four reasonably orthogonal scores derived from this instrument.

These outcomes are interpreted tentatively as follows: Sex-role identity becomes increasingly stabilized as the child's cognition of gender identity invariance becomes increasingly firm. When the child developes some competence (if not maximal performance) with regard to this basic reality judgment, this judgment influences his sex-role attitudes and behaviors (Kohlberg, 1966b). Prior to this point in development, however, sex-role processes such as those indexed by the present four orthogonal subscales probably are influenced by preoperational interpretations of the meanings of social relationships, reinforcements, and cultural stereotypes.

Future analyses will evaluate the hypothesis that prior to the achievement of a cognitive reality judgment of gender identity invariance, the four component measures are differentially influenced by diverse and interacting motivational, affective, and social factors measured in the study.

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 are relatively contented, are able to relate well to others, and have a generally positive self-concept. Yet, in contrast to the 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 (Wiley, 1961). Since a primary goal of most preschool programs is to increase the child's self-esteem, there has been much rhetoric on the need for adequate measurement in this area. For a variety of reasons, however, task development has been slow. The Brown IDS Self-Concept Referents Test, developed in 1966, is a technique for assessing self-concept using a photograph of the child to induce the young child to take the role of another toward himself. The task measures self-esteem and also the perception of oneself from the point of view of socially significant others. It was included in the present study since it was one of the few measures in the literature during the first year of the study relating to the child's evaluation of "self as object" and "self as subject" which had reliability data and evidence of validity for use with four-year-old disadvantaged children.

In Brown's (1966) study, 38 black (lower class) and 36 white (middle class) preschool children responded as to how they, their mothers, their teachers, and "other kids" perceived them. Self-perceptions of the black children were significantly less favorable than those of 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).

Task Description

A full-length color Polaroid photograph is taken of the child. After the tester ascertains that the child recognizes himself in the picture, the child is asked to respond to 15 bipolar adjectival items stated in the vocabulary of the four-year-old child (e.g., happy-sad; afraid of a lot of things-not afraid of a lot of things), each time referring to the child's picture. All items are presented in an "either-or" format, the more positive and socially desirable choice being scored one and the less socially desirable choice scored zero. (Positive choices were randomly assigned first and second position.) Since data with children three-six to four-six 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.

Scoring

In the present study, each item was scored as positive, negative, refused, indeterminate (e.g., multiple answers) or "don't know." Total number of items to which the child did not respond and total self-concept score were the two scores used in the present structural analyses. For these analyses, the self-concept score was adjusted in order to account for items which were refused or otherwise unscorable. Thus the scores reported here are percentages of "positive" responses based on the number of items clearly answered in a positive or negative way. In addition, the child was judged as smiling or not smiling in the photograph in order to investigate the relationship between the concrete stimulus and the response alternative chosen.

Score Characteristics

The KR-20 coefficient of reliability (alpha) for the unadjusted self-concept score was computed to be .71 for a subsample of three of the sites (N = 972). For this smaller group, R biserials for each item with the total score ranged from .48 to .73. Coefficient alpha for number of omitted items was .91.

Sample Characteristics

As has been found in previous studies, the distribution of the self-concept score (N = 1371) was markedly skewed (mean = 82.0%, S.D. = 14.6), indicating the strong tendency for the child to select positive attributes. The correction for indeterminate responses, however, would have spuriously inflated the score to the extent that this reflected a defensive response rather than the child's lack of differentiation with regard to a particular item. However, the correlation between the corrected and uncorrected self-concept score was .83. The mean number of items omitted was 1.5, with S.D. = 3.0.

Age and sex differences in mean self-concept score were insignificant. There were, however, differential item responses for boys and girls that merit further study. For example, the strong-weak item discriminated most between the sexes. As would be expected, total number of omitted items decreased with age. The majority of children (66.6%) did not smile for their picture, but females, white children and older children smiled more, suggesting that smiling may have been a reflection of differential social expectations or of the child's greater familiarity and ease in the situation. Smiling in the photograph correlated only .15 with the happy (rather than sad) response to item one. Thus, children did not seem to be responding primarily on the basis of immediate stimulus cues.

Future analyses will be directed toward investigating 1) differential item responses over time as they relate to differential sex role expectancies and other environmental influences upon the child as he extends his interactions with "significant others," and 2) the interaction among affective and cognitive responses. Utilizing Mead's notion of the evolvement of self-concept from one's perception of salient others' perception of self, we may observe the development of the young child's positive and negative conceptions of self as they interrelate with data obtained on specific teacher-child, peer-child and parent-child interaction behaviors.

Children's Auditory Discrimination Inventory (CADI)

Purpose

In assessing verbal skills, it is important to determine children's ability to discriminate oral verbal utterances. Such ability contributes to effective listening, reading, speaking, and writing behavior.

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 so that the results would not be confounded by irrelevant differences in vocabulary skills.

Task Description

The CADI, the result of Stern's work, is an individually administered measure designed to evaluate children's ability to identify, from among two pictures that have been given oral word equivalents, the picture that represents the orally presented stimulus word. The testing material includes 38 pairs of words and 38 cards with two pictures on each card. One picture in each pair is a real picture representing a familiar word and the second picture is a nonsense picture to be paired with a nonsense word. The real and nonsense pictures are randomly located on the right or left side of the cards to avoid positional responding by the child. E presents each pair of pictures orally, naming them as he points to each. Following the presentation of each pair, the child is asked to point to the one picture that represents the name he then says orally. There are two sample items prior to beginning the test. If the child does not respond to the sample items, the procedure is explained again. If, after the repetition, the child still does not respond, the test is discontinued. During the test an item may be repeated only once. Also, to prevent lip reading, E turns his head slightly away from the child when saying the test word.

Scoring

The name for each picture in each pair is given on the Answer Sheet with the test word underlined. E records whether the child's response was correct or incorrect, if a multiple answer was given, if the item was repeated, or if there was need to probe for the best answer. The score is the number correct (range = 0-38).

Internal consistency, using the Kuder-Richardson (Formula 20) estimate of reliability, was found to be at .87 (N=1443).

Sample Characteristics

Table 1 presents mean total score by age and sex subgroups for the composite sample.

Table 1

Means, Standard Deviations, and Percentile Distributions for Age and Sex Subgroups

Group	N	Mean	S.D.	Percentiles 10	25	50	75	90
42-44 mo.	84	27.69	4.92	20.80	23.17	27.18	31.44	34.12
45-47 mo.	310	28.42	5.10	21.36	23.93	27.69	31.93	35.17
48-50 mo.	327	28.70	5.57	21.34	24.04	28.33	32.53	35.72
51-53 mo.	382	28.94	5.39	21.26	23.79	28.40	33.15	35.66
54-56 mo.	274	29.20	5.36	21.87	24.13	28.46	33.42	35.90
57-59 mo.	61	29.84	4.28	23.89	26.50	29.11	32.93	34.78
Male	758	28.39	5.38	21.11	23.60	27.84	32.27	35.38
Female	680	29.22	5.19	21.86	24.56	28.68	33.05	35.67
Total	1438	28.79	5.30	21.45	24.01	28.22	32.67	35.52

The data indicated that children's auditory discrimination increases with age for children between 42 and 59 months. Further, it appeared that girls in this age range had slightly better auditory discrimination than boys.

Remarks

Real words and nonsense words do not appear to be measuring the same thing in this task. During testing it was noted that children pointed more often to the "real" picture. Subsequent analyses revealed that scores were higher for real words than for nonsense words (Mean for real words = 16.59, S.D. = 2.69; Mean for nonsense words = 12.21, S.D. = 4.49), and the correlation between the real word subscore and nonsense word subscore was .03. Internal consistency of these subscores, using the coefficient alpha estimate of reliability, was found to be at .76 and .85, respectively. In the present analyses the total score only was used, but future analyses will investigate the differential meaning of "nonsense" and "real word" subscores.

ETS Enumeration I

Purpose

The purpose of ETS Enumeration I was to measure a component of quantitative ability that does not require counting or reciting the name of numerals. The task assesses the child's ability to organize a field of figures and to keep track of two shifting sets--the set of figures "pointed at" and the set "not-yet-pointed at." It is patterned after a procedure described by Potter and Levy (1968) in the belief that the no-counting quality makes it a promising procedure with very young children.

Task Description

The child is asked to point once, and once only, to each figure in an array on a test-booklet page. No verbal response is requested. The figures, consisting of colored circles, are arranged on a page into one of three types of arrays: single line, rows, random. The number of figures within an array varies from six to nine. There is one practice item and 12 test items. (A thirteenth item, on which the child was asked to count aloud, was also included for the purposes of examining performance under conditions of counting instructions and of preparing a version of the measure for the second year of the study.)

Scoring

The tester's record of the child's performance on each item includes an indication of the nature of any errors made and the direction of hand movements. The item is scored "correct" if the child points to each figure in the array once, neither omitting nor repeating a figure. The possible range for the total score is 0 to 12.

Sample Characteristics

In their study of 58 nursery school children, ages 2 1/2 to 4, Potter and Levy found that accuracy of performance was clearly correlated with age. Data in the present study indicate a similar finding. As shown in Table 1 there is a steady rise in score as a function of age with a mean of 4.06 for the youngest group and 7.09 for the oldest. Table 1 also 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 6; the 25th and 75th percentiles are located evenly at scores of 2.98 and 8.64, respectively.

Table 1

Enumeration I: Distributions of Total Correct Score by Age*

Age	N	Mean	S.D.	Percentiles 10	25	50	75	90
42-44 mo.	82	4.06	3.12	0.09	1.31	3.50	6.18	8.19
45-47 mo.	306	5.07	3.66	0.28	1.90	4.82	7.93	10.41
48-50 mo.	323	5.74	3.53	0.61	2.71	6.01	8.51	10.54
51-53 mo.	367	6.39	3.39	1.57	3.70	6.58	9.02	10.96
54-56 mo.	259	6.64	3.27	2.13	4.10	6.95	9.20	10.93
67-59 mo.	58	7.09	3.15	2.63	4.37	7.40	9.37	11.37
Total	1395	5.89	3.52	0.81	2.98	6.04	8.64	10.71

*Range = 0-12.

Item Difficulty and Internal Consistency

Item analysis of the measure indicates that accuracy of performance is systematically influenced by the number and arrangement of figures on a stimulus page. Arrays containing six figures were consistently easier than arrays containing nine figures. Difficulty levels of six-figure-items ranged from 75% passing to 50% passing; difficulty levels of nine-figure-items ranged from 47% passing to 27%. For items of six figures, random arrangements were the most difficult, ordered arrangements of two rows were of moderate difficulty, while a single row of figures was the easiest. A parallel order of difficulty, associated with type of arrangement, was found for the items of nine figures. This contribution of number and arrangement of figures to item difficulty parallels the findings of Potter and Levy. An Alpha of .85 indicates satisfactory internal consistency.

Item 13: Counting

The counting item (seven figures in a single row) was administered at the conclusion of the test. Responses were coded in two ways. One coding system took into account whether the total of number names recited by the child corresponded, one to one, with the total of seven figures; the other system took into account the correctness of the sequence recited. Thus, a response of "1, 2, 3, 7, 8, 9, 10" was judged correct in correspondence (seven number names) but incorrect in sequence. The response: "1, 2, 3, 4" was judged noncorresponding but correct in sequence. Approximately 29% of the children were correct in both senses; 22% recited a correct but non-corresponding sequence; another 11% were accurate in correspondence but incorrect in sequence; 30% failed in both systems; 8% refused. Although there is a correlation of .36 between item 13 and the total correct for items 1-12, it would appear that the request to count changes the character of the task in the direction of making it more difficult.

Remarks

Perceptual abilities required in the organization of a field of figures would be expected to contribute to performance on this test. The observed pattern of item difficulty makes psychological sense in this context, insofar as an increase in number of figures on a page and/or an increase in complexity of their arrangement should have the effects on accuracy that were indeed evident. However, the extent to which this test may be regarded as a measure of enumerating abilities can be determined only through further examination of relationships to other measures within the study and to longitudinal evidence on the significance of these "precounting" abilities. In Piaget's analyses, perceptual ordering and articulation are viewed as necessary precursors to a conceptual understanding of number. Finally, it should also be stressed that on a repetitive homogeneous task of this sort, the child's style and persistence in responding can very well play a major part determining his score.

ETS Matched Pictures Language Comprehension Test I

Purpose

It is knowledge of grammatical or syntactic structure that allows the mature speaker to understand and generate an infinite variety of sentences. Thus, the degree to which a person has command of structural rules determines what linguists have come to call underlying language competence. Increasingly over the past years it is this aspect of language, rather than its lexical aspect (i.e., vocabulary), which has been the focus of study for those interested in development of language and in its relationship to other facets of cognitive growth.

Most studies of syntactic development have relied on spontaneous speech samples for their primary data. Productive speech, however, can be influenced by a host of factors which have little or nothing to do with language competence. Therefore, a comprehension test was devised for the Longitudinal Study as a means of obtaining data on syntactic development. The major overall purpose of such a measure (and of similar measures devised for succeeding years) is to study the developmental pattern of syntactic comprehension and the relationship of this pattern to family and school determinants. In addition, the immediate concern in Year 1 was to shed light on a question of current controversy and debate among educators and theorists. Is the child from a low-income environment retarded in syntactic comprehension, as Bereiter and Engelmann (1966) and Osborne (1968) suggest? Or, is even a minimal language environment sufficient for a child to develop the basic grammatical rules of adult language, as many linguists (e.g., Weksel, 1965; Lenneberg, 1967) would suggest? Analysis of the mother's verbalizations in the interaction tasks will provide an index of the restriction of the child's linguistic environment which can then be related to the various measures of linguistic competence used in the study.

Task Description

The ETS Matched Pictures Language Comprehension Test 1 utilizes Roger Brown's and Jean Berko Gleason's "matched pictures" technique. It consists of 20 cards, each card containing a pair of pictures. Both pictures in a pair contain identical 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 20 picture pairs are divided into four subtests, with counter-balanced design for the position of the "correct" picture (right or left on the card) and the sequence in which E names the "correct" picture title (first or second). The four subtests are as follows:

Future Tense	4 items
Past Tense	4 items
Negation	6 items
Prepositions	6 items

Scoring

Each item on the test is scored either right or wrong (1-0) and the test as a whole yields six scores: the four subtest scores, a Total Tense Score and a Total Score.

Score Properties

Table 1 presents the intercorrelations among subscores. Given the low estimates of internal consistency (coefficient alpha) for these subscores, the total score only (coefficient alpha = .57) was used in the overall structural analyses.

Table 1

Subscore Intercorrelations (N=1460)

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

* - coefficient alpha

Sample Characteristics

The distribution of Total Correct (range 0-20) for an N of 1435 Ss was approximately normal, with a Mean of 12.78, a Median of 12.83 and a standard deviation of 2.94 (see Table 2).

Table 2

Distributions of Total Score by Age

Age	N	Mean	S.D.	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
TOTAL	1435	12.78	2.94	8.98	10.82	12.83	14.86	16.64

Remarks

With respect to the various subtest scores, future and past tense discriminations proved to be quite difficult for children of this age; negations and prepositions proved quite easy, with the children obtaining median scores of 5.03 and 5.09 respectively on these 6-item subtests. While these results contradict the claims of Bereiter and Engelmann (1966) and Osborne (1968) who state that disadvantaged 4-year-old children do not understand prepositions or negation, they are similar to results obtained on the Matched Pictures test with other low and middle-income populations (ETS, PR-70-20) and to results obtained in a recent study of syntactic comprehension in Italian preschool children (Parisi, 1971). However, it will be necessary to study succeeding years of longitudinal data before drawing any firm conclusions about the effect of low-income environments on basic language competence.

ETS Story Sequence Task I

Purpose

Traditionally, the development of language skills in young children has been viewed in terms of the two major components of "receptive" vs. "productive" language. The "receptive" skills are identified as the recognition of language labels given in oral or written form, whereas the "productive" skills involve the use of these labels in such a way that the child "produces" (e.g., "says something") in response. Both types of language skills are present in intelligence tests, but tests which measure school readiness, such as the Metropolitan Readiness Tests (Hildreth at al., 1965), are composed almost entirely of "receptive" language items. However, the close relationship of both measures with actual school performances suggests the advisability of assessing both receptive and productive language at earlier ages.

There is a great deal of evidence for social class differences in language development (e.g., Loban, 1965; Raph, 1965; Weaver, 1965) and some evidence that productive responses are somewhat more difficult than receptive responses. The few studies which have compared the two modes of response with the same materials and procedures (e.g., Carson & Rabin, 1960) indicate that productive language is a much more difficult skill for the culturally disadvantaged child. Most of these studies have been limited to a comparison of receptive and productive skills regarding single-word comprehension versus asking the child to verbally label or describe the pictured item. An interest in looking at the young child's use of larger units (sentences and short paragraphs) requiring both receptive and productive responses led to the development of the ETS Story Sequence Task used in this study.

Task Description

The ETS Story Sequence Task was designed to assess the young child's understanding and use of language in story sequence under three different conditions varying in the degree to which the child is asked to use receptive and productive language skills. The materials are similar for all conditions, consisting of seven sets of cards with drawings of animals in various situations, including one instructional set and six test sets. There is no apparent sequence in any of the pictured situations--the sequence is provided either by the verbal cues used in the presented stories or by the story produced by the child. The stories were especially written for these tasks in order to avoid the problem of differential familiarity. It was decided to use animals as the "characters" in the stories, rather than children, because of the difficulty of "balancing" the distribution of sex, race and situations in a small number of items.

The items are divided into three types of tasks which require different kinds of responses from the child as follows:

Task Type	Description
1	Receptive language: the child selects and arranges card sequence while listening to a story told by the examiner. There is no inherent order in the pictured situations and the child is dependent on linguistic cues provided in the story.
2	Productive language using verbal recall: tester presents cards in order as she tells the story. Child is asked to retell the "same" story.
3	Productive language using child's own story: child chooses picture cards from an array and tells his own story about them.

Two practice items are given 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 an array of cards. The two test items which follow ask the child to select the appropriate sequence of pictures while listening to a story. In both test items, there is no replacement of cards in the array so that the size of the array diminishes with each choice, but the child could use the same card twice (e.g., the child may decide to reuse a card placed in the row rather than the last card 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.

Scoring

Story Sequence I is the first of the three parts of the total task, and only its scores are reported here. It is composed of two items which focus on the child's receptive language and his ability to use linguistic cues in the construction of a sequence. There are two sequences: Tommy Kitten (3 cards) and Timothy Mouse (4 cards). Each correctly selected card is given one point so that the possible range for the two items combined is 0-7.

The product-moment correlation coefficient between item 1 and item 2 for the composite sample of 1448 children was 0.33.

Sample Characteristics

The composite mean score for the task was 4.26 with a S.D. of 2.25. Reliability (coefficient alpha) was .50. The combined site scores (with only a small discrepancy in the oldest group) showed a consistent progression with age (see Table 1). This increase was consistent across all age ranges in Auburn, Portland, and St. Louis but showed some very slight discrepancies in the youngest and oldest groups in Trenton (however, these groups have very small N's, e.g., five children in oldest group). The potential range of scores from 0-7 was found in each age group of children. Sex differences in mean scores were negligible. The composite score for the females (4.43) was slightly higher than for the males (4.12), and similar differences were reported for each of the sites.

Table 1

Mean Total Score by Age

Age	N	Mean	S.D.	Range
42-44 mo.	89	3.86	2.41	0-7
45-47 mo.	315	3.87	2.19	0-7
48-50 mo.	331	4.11	2.26	0-7
51-53 mo.	383	4.48	2.23	0-7
54-56 mo.	270	4.68	2.19	0-7
57-59 mo.	60	4.58	2.29	0-7
Total	1448	4.26	2.25	0-7

Conclusions

The findings are in general agreement with the results of the earlier use of this measure with preschool and kindergarten children in New York City (Melton et al., 1968). In that study, a significant SES difference was reported. Further, the results showed an interaction of SES with sex and age ($p<.04$) which is relevant to the current study. That is, within the middle SES, the girls were superior to boys at both age levels, whereas in the low SES, the boys did better in the older group and there were no sex differences in the younger group. The children in the present study form a similar SES population of "four-year-olds" and it is interesting to note the parallel finding of no sex differences at this age. Later data will be examined to see if the parallel holds at older ages. Moreover, addition of the recall and story production items in later years will provide some information on the relationship between the receptive and productive language of the child using similar stimulus materials.

Fixation Time

Purpose

Working with infants and young children, Kagan and Lewis (1965) and Lewis and his associates (Lewis et al., 1970; Lewis, 1971; Lewis, 1972) have demonstrated that attention, at least in the early years of life, is an index of early cognitive functioning. Moreover, individual differences in attention may also have some direct effects on learning; for example, the child who cannot concentrate or who grows bored quickly cannot obtain as much information from his environment as the child who can. Thus, in the child of preschool age, attention may serve to relate to later as well as current cognitive functioning. Attention may be noncognitively determined as well, for example, by the intentions and desires of the subject (Messick, ETS, PR-68-4). In this respect attention may fall within the personality domain as well as serving as an index of cognitive functioning.

Task Description

The task used in this study obtained a measure of the amount of time a child fixates or looks at a given picture as it is repeated over a number of trials. Of interest also was the degree to which a child was able to discriminate between this redundant stimulus and a variation of it.

Two series of slides were used in the fixation task. Series one consisted of six trials of a redundant nonsocial visual stimulus (twenty chromatic straight lines) and a seventh trial of chromatic curved lines. Series two, the social array, consisted of a chromatic schematic representation of a family: a man, woman, and young child, shown for six trials, and a seventh presentation which consisted of the same schematic without color. Each slide was shown for 30 seconds and was followed by a black slide for 30 seconds before the next presentation. An observer positioned behind a peg board partition watched the child's behavior. Fixation time was recorded as the amount of time a child looked at each picture before turning away the first time. (In past research the interobserver reliability for determining whether the child was or was not looking at the screen had varied between .60 and .99. [ETS, PR-68-4]). 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 at that point. A short break was allowed between series but not between slide (trial) presentations. This procedure has been used previously in laboratory settings under highly controlled conditions; however, these conditions did not always exist for the present study and the results may have been affected by external noise, visibility of the observer, and other distracting stimuli.

Scoring

By presenting slides in two series, each consisting of a repeated and then some varied event, it was possible to obtain three measures of attention: response decrement, stimulus differentiation, and amount of attention. Response decrement, or habituation, is measured by the change over trials in response strength to a repeated event; stimulus differentiation is measured by response recovery when a variation of the repeated event is presented (difference between trial 7 and trial 6); amount of attention is the total fixation time over all trials. For the structural analyses, mean habituation and mean recovery for the two series were used. Recovery scores represent the difference between trial 7 and trial 6. Habituation scores were computed as the difference between trial 1 and trial 6 weighted by trial 1 time.

Score Properties

The data that follow are based on three-site totals (Trenton, Portland, Lee County) as equipment failures did not allow collection of fixation data in St. Louis.

Table 1 presents the intercorrelations between mean recovery and mean habituation times for the social and nonsocial series.

Table 1

Intercorrelations Between Recovery and Habituation Scores (N=1168-1195)

	1	2	3
1. Non-Social, Recovery	--	--	--
2. Non-Social, Habituation	.48	--	--
3. Social, Recovery	.21	.07	--
4. Social, Habituation	.11	.14	.60

The relationships between recovery and habituation for each series indicate that greater recovery is associated with greater change in response strength; that this relationship is higher for the social stimuli may be due to the fact that the social stimuli were attended to longer. However, correlations of habituation and recovery scores for the two types of stimuli were quite low. Correlations between recovery and habituation within a single task may be spurious because they are both dependent on the score in trial 6.

Sample Characteristics

Table 2 presents data for three sites for each stimulus picture in the social and nonsocial series. As can be seen, response decrement is shown in the form of a negative exponential function for both social and nonsocial stimuli. Response recovery is shown on trial seven, with a very small indication that this is greater for the social stimuli; however, the social stimuli were attended to longer.

The mean response decrement for the two series combined was .43, S.D. = .37; mean response recovery for the two series combined was 9.79, S.D. = 8.31; and the correlation of these mean scores was .53 (N=1222).

Table 2

Mean Fixation Time, Recovery and Response Decrement for Social and Nonsocial Stimuli

	Nonsocial			Social		
	N	Mean	S.D.	N	Mean	S.D.
Trial 1	1221	12.29	8.50	1220	19.63	9.52
Trial 2	1227	9.89	8.02	1212	14.28	9.37
Trial 3	1224	8.32	7.50	1210	11.40	9.21
Trial 4	1223	7.80	7.40	1214	10.06	9.01
Trial 5	1216	7.61	7.20	1214	8.96	8.32
Trial 6	1204	7.40	7.59	1206	8.66	8.31
Trial 7	1217	16.75	9.79	1207	18.86	9.44
Recovery (7-6)	1195	9.32	10.50	1200	10.25	10.72
Response Decrement	1194	.36	.50	1201	.50	.46

Form Reproduction

Purpose

Success in the ability to reproduce geometric forms graphically relies on a complex integration of visual-motor, visual, proprioceptive, verbal and conceptual functions. This ability is highly linked to developmental level both within and across age levels. Beery (1967) reported a correlation of .89 between chronological age (2-15 years) and number of items correct on the 24-item Developmental Test of Visual-Motor Integration (VMI). Also related to this ability is the level of achievement in reading (e.g., Beery, 1967) and writing. Reported correlations with measures of intelligence are .48 for Verbal, .60 for Performance, and .58 with Full Scale Scores on the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) (Wechsler, 1967). Tests of form reproduction are also indicators of integrative skills at a nonverbal level, thus increasing ease of administration with a non-verbal child.

Interjudge scorer reliability for the VMI was reported as .98 by Beery (1967) and by Buktenica (1966). Test-retest reliabilities for this instrument ranged from .80 to .90 for intervals from two to eight weeks (Beery, 1967). Wechsler (1967) reported a test-retest reliability of .62 for form reproduction for a retest group, and estimated it at .78 for the standardization sample. Internal consistency (Kuder-Richardson) for Beery's sample (age 3-14 years) was .93 on a sample of suburban subjects. Buktenica (1966) reported a Kuder-Richardson coefficient of .78 for a first grade sample.

Task Description

The Form Reproduction Test for Year 1 consisted of the four form reproduction items from the Preschool Inventory (vertical line, circle, square and triangle) and the right oblique line and oblique cross items from the VMI developed by Beery (1967). The procedure adapted for use in this study required the child to copy the four Preschool Inventory forms in the standard manner and order for this test; when the Preschool Inventory was completed the child was asked to copy the two forms from the VMI. It should be noted that the first four items are included in the VMI forms, although their ordering is somewhat different.

Scoring

Scoring of the six forms was done using a combination of the Beery system of 0 or 1 (for the vertical line, triangle, right oblique line and oblique cross) and the WPPSI scoring criteria of 0,1,2 for the circle and square, resulting in a maximum score of 8. WPPSI scoring was used whenever possible since the Beery and WPPSI use essentially the same criteria, but the WPPSI provides more differentiated scoring and norms based on larger preschool samples.

Score Characteristics

Inter-item correlations generally were low (see Table 1). The highest relationship was between the two most difficult items, the square and triangle ($r = .46$). (Successful production and integration of vertical and horizontal lines appears common to both of these forms.) All items have moderate correlations with total score, the highest being the circle (.72) followed by the cross (.66), square (.63), right oblique line (.56), vertical line (.55) and triangle (.51). These item-scale correlations are part-whole correlations and have not been corrected for overlap. Reliability (coefficient alpha) for the total score was .61.

Table 1

Inter-item and Total Score Correlations (N = 1318-1411)

	Line	Circle	Square	Triangle	Rt. Oblique	Cross
Circle	.28					
Square	.19	.32				
Triangle	.15	.24	.46			
Right Oblique	.20	.23	.23	.18		
Cross	.21	.31	.37	.29	.25	
Total	.55	.72	.63	.51	.56	.66

Sample Characteristics

The mean total score distribution for the six age groups presented in Table 2 indicates a clear linear relationship between age and success in reproducing forms, although differences between adjacent age groups appear small. Examination of percent passing each item indicates that although all forms show some age increase, these increases are not consistent across age intervals. Sex differences were negligible for total score distribution.

Table 2

Mean Total Score* by Age and Sex

Group	N	Mean	SD	Percentiles 10	25	50	75	90
42-44 mo.	88	1.49	1.24	.00	.70	1.03	1.92	3.23
45-47 mo.	318	1.65	1.34	.00	.72	1.10	2.95	3.37
48-50 mo.	346	1.82	1.41	.01	.79	1.67	2.97	3.87
51-53 mo.	384	2.40	1.65	.13	.96	1.92	3.32	4.80
54-56 mo.	271	2.53	1.72	.16	1.00	1.98	3.36	4.97
57-59 mo.	71	2.84	1.88	.78	1.63	1.97	3.94	5.17
Male	783	1.93	1.56	.01	.78	1.67	3.09	4.11
Female	685	2.27	1.59	.12	.94	1.86	3.24	4.25
Total	1468	2.09	1.58	.05	.84	1 76	3.17	4.18

*range = 0-8

The 50th percentile ages for passing the right oblique line are 52 months (males) and 48 months (females). In the 48-50 month age group 25.8% of the sample passed, while in the 51-53 month age group 36.3% passed this item. Beery's age norms indicate that by age 4-6 (males) and 4-3 (females) 50% of a sample should be expected to pass the square; for this sample, 15.8% of the 51-53 months age group received full or partial credit, and 22.9% of the 54-56 month age group received full or partial credit. These data may indicate poorer visual-motor integration in this sample, or sampling fluctuations due to a substantially smaller sample in Beery's work. These questions await further research.

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 were seriously limited in their use with some groups of children: children with functional or organically determined speech defects; culturally deprived children with limited verbal and experiential repertoires; children with motor handicaps; and 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 3-6 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 investigators (Frostig, Maslow, Lefever, & Whittlesey, 1964; Kephart, 1960; Koppitz, 1964) have postulated the existence of a neurological developmental hierarchy underlying cognitive skills. Although the evidence is inconclusive, it can be hypothesized that children who lack certain discrimination skills, whether through a developmental "lag" or through physical impairment, will not be able to benefit from many normal learning experiences.

Task Description

The test is one of form discrimination involving black geometric figures printed on cards. It consists of 30 test items, preceded by 3 practice items, in which the child is presented with a stimulus form and asked to point to the matching one from among several alternatives. There are two types of forms; some are purely random and some figures are related to each other along a continuum of known variation. Complexity is defined by the number of angles in the figure and by the number of alternatives given the child (either 2, 3, or 5).

Scoring

Items were scored as correct, incorrect or indeterminate (e.g., multiple answers). The score is simply 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 (5) of alternative responses) makes it difficult to tease out process differences among items. Given the above, only the total correct score was used in the overall structural analyses.

The coefficient alpha index of reliability for total score for an N of 1419 was .74.

Sample Characteristics

Table 1 presents the means, standard deviations and percentile distributions for total score by age.

Table 1

Distributions of Total Score* by Age

				Percentiles				
Age	N	Mean	SD	10	25	50	75	90
42-44 mo.	80	16.47	5.23	9.50	12.83	16.72	20.25	23.00
45-47 mo.	295	15.68	4.83	8.88	12.58	15.76	19.45	22.05
48-50 mo.	328	16.19	4.81	9.98	13.10	16.06	19.45	22.34
51-53 mo.	379	17.93	4.60	11.65	15.11	17.84	21.43	24.12
54-56 mo.	270	17.67	4.84	11.00	14.26	17.55	21.32	23.80
57-59 mo.	59	17.97	4.55	11.80	14.25	18.42	21.22	23.53
Total	1411	16.93	4.86	10.25	13.68	16.94	20.53	23.18

*range = 0-30

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. Sex differences were consistently negligible across sites, with girls (Mean = 17.1, S.D. = 4.86) scoring slightly higher than the boys (Mean = 16.7, S.D. = 4.86). There was a trend for mean score to increase with age, but with the exception of St. Louis, no significant mean differences by age were found.

Remarks

Although Rosenberg referred to this task as a nonverbal test of general mental ability, the correlation with the Peabody Picture Vocabulary Test, Form A, was only .33. However, it did correlate substantially with other tests involving visual discrimination (.52 with errors on the Matching Familiar Figures Test and .41 with quickest time to correct solution on the Seguin Form Board Test). Future analyses will investigate possible differential effects of experiential factors on performance in contrast to effects on verbal measures, as well as exploring relationships to indices of possible neurological involvement available from the children's health data.

Massad Mimicry Test 1

Purpose

Since children generally learn language through imitation and tend to rely on the auditory-vocal system throughout their learning experiences, it would appear only natural to employ this system as one means of observing children's linguistic competence. Research by Slobin and Welsh (1967), Keeney (1969), and Fraser, Bellugi and Brown (1963) has indicated that children's linguistic competence may be assessed through controlled, elicited imitation and that imitation ability may be a separate skill from understanding or producing language.

Task Description

The test is an individually administered measure intended for 3 1/2-year-old children. Part I of the test evaluates children's ability to reproduce phonemes in thirty (30) nonsense words upon hearing them no more than three times from a tape-recorded model. Part II, using a tape-recorded model, assesses children's ability to reproduce meaningful words and phonemes as they occur in word phrases (primitive sentences) and two simple sentences.

E uses two tape recorders--one for playing the model (stimulus) tape and the other for recording during the testing session. Both the model utterances and the child's responses are recorded on the latter tape.

A child must be able to listen as well as repeat what he hears in order to perform the tasks. Prior to testing, some rapport must be established so that the child is talking, realizes that he is making a recording, and understands that the equipment is not to be played with while making the recording. The test is proposed as a game of "Follow the Leader" in which the model utterance is to be repeated by the child exactly as given by the model. A warm-up session using three sample utterances recorded on the model tape occurs before the beginning of each part of the test. Positive reinforcements of the child's responses are given only at four designated times during the actual testing; specifically, at four designated times when the model tape is being played and the child is responding.

Scoring

A Scoring Guide is used with a SCRIBE answer sheet for each child. It is necessary to use the Scoring Guide to determine what is to be listened for at each numbered space of the SCRIBE answer sheet. Only the first two spaces (A or B) of the answer sheet are used in scoring. "A" is marked when the scorer judges the child has (a) correctly repeated the specific sound or (b) been able to say the word or some form of it as given by the model tape (when applicable). "B" is marked if the scorer judges that the child has not successfully repeated the sound or a comprehensible form of the word given on the model tape (as indicated in the Scoring Guide).

Part I gives three scores: Initial Sounds (Possible Score, 30), Medial Sounds (Possible Score, 28), and Final Sounds (Possible Score, 30). There are two scores for Part II: Final Sounds (Possible Score, 10) and Model Word or Some Semblance of It (Possible Score, 35). The three scores of Part I may be totaled for a score on Nonsense Words; however, the scores in Part II reflect distinct capabilities and should not be totaled.

In addition, the Medial Sounds include twelve long vowels and thirteen short vowels which may be looked at independently. However, if a score is given for each of these two types of vowels, it must be remembered that the scores are interdependent with the score for Medial Sounds.

Score Characteristics

To identify trends in language development regarding the ability of children to reproduce initial, medial, and final phonemes of utterances, 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.

The intercorrelations among the three scores of primary importance are presented in Table 1. The data indicate that, while they measure different things, the three parts are significantly interrelated.

Table 1

Intercorrelations Among Mimicry Scores (N=approx. 1000)

Part	2	3
1. Nonsense Words Total Sounds	.56	.53
2. Meaningful Words in Phrases, Final Sounds		.47
3. Meaningful Words in Phrases, Model Word or some semblance of it		

Interscorer Reliabilities

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

Tables 2 and 3 indicate interscorer reliabilities for Part I and each of the first three subsections of Part I and the two subsections of Part II. In each table, 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 with the assumption that differences in means between judges are systematic and should not be considered part of the error variance.

Table 2

Interscorer Reliabilities Between Judges A and B

Part	Subsection	N	Reliability	
I. Nonsense Words		289	.71	.86
	A. Initial Sounds	289	.81	.82
	B. Medial Sounds	291	.71	.83
	C. Final Sounds	289	.45	.77
II. Meaningful Words in Phrases				
	A. Final Sounds	273	.66	.66
	B. Model Word or Some Semblance of It	227	.78	.80

Table 4

Estimated Reliabilities for Internal Consistency

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

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

The data indicate that a satisfactory degree of internal consistency exists within Part I and the various subsections. Part ID and Part IE, consisting of 12 and 13 items, respectively, and each being independent of the other but both 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 for ID and IE, contain no less than twenty-eight. Further investigation is planned for those subsections with the lowest reliabilities.

Sample Characteristics

Because of the demonstrated inconsistency of scoring across scorers as well as by scorer (and taking into consideration the nature of the population), scores were standardized by scorer. As a result of standardization of scores, the data obtained from each of the three scorers could be combined and viewed as though there were no differences of judgment among or within scorers. However, there exist certain limitations on interpretation of the resultant data in that the means, standard deviations, and ranges of scores cannot be interpreted in the same manner as raw score data. The data do indicate, nevertheless, trends in language development for the population studied. Tables 5 through 7 represent data based on the total population studied at the four sites.

Table 5

Total-Group Ranges

Part	Subsection	No. Items	N	Range (For Adjusted Scores)
I. Nonsense Words		88	1098	-4.05 - 2.49
	A. Initial Sounds	30	1101	-4.09 - 2.58
	B. Medial Sounds	28	1105	-4.65 - 2.04
	C. Final Sounds	30	1100	-2.68 - 2.95
	D. Long Vowels	12	1101	-4.87 - 2.20
	E. Short Vowels	13	1139	-4.45 - 2.05
II. Meaningful Words in Phrases				
	A. Final Sounds	10	1060	-2.0 - 3.0
	B. Model Word or Some Semblance of It	35	954	-5.8 - 1.7

Table 6

Nonsense Words, Total Sounds: Means, Standard Deviations and Range

Group	N	Mean	S.D.	Range
42-44 mo.	62	-0.35	1.05	-4.05 - 1.58
45-47 mo.	211	-0.23	1.09	-3.97 - 2.13
48-50 mo.	265	-0.07	1.00	-3.34 - 2.13
51-53 mo.	292	0.14	0.92	-2.28 - 2.44
54-56 mo.	222	0.14	0.93	-2.84 - 2.49
57-59 mo.	46	0.34	0.94	-1.54 - 2.21
Male	569	-0.10	1.01	-4.05 - 2.49
Female	529	0.11	0.98	-3.97 - 2.44

Table 7

Meaningful Words in Phrases, Final Sounds: Means, Standard Deviations, and Range

Group	N	Mean	S.D.	Range
42-44 mo.	62	-0.13	0.92	-1.49 - 1.74
45-47 mo.	198	-0.06	1.00	-1.95 - 2.39
48-50 mo.	254	-0.04	0.99	-2.00 - 2.83
51-53 mo.	285	0.02	1.01	-1.95 - 2.97
54-56 mo.	215	0.08	1.01	-1.95 - 2.97
57-59 mo.	46	0.12	1.07	-1.95 - 1.98
Male	545	-0.07	0.98	-1.95 - 2.83
Female	515	0.07	1.01	-2.00 - 2.97

The data indicated that children's ability to reproduce phonemes as well as meaningful words in phrases increases with age between 42 and 59 months. Further, it appeared that boys and girls in this age range differ in these abilities, girls tending to achieve at a higher level.

Matching Familiar Figures Test

Purpose

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). 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. Reflectiveness is, however, related to performance on tests of reasoning (Kagan, Pearson, & Welch, 1966) and of word reading (Kagan, 1965) in early elementary children. Its implications for performance in children below school age are not known, but the dimension has been found to be present in kindergarten children (Ward, 1968), and in middle-class nursery school children (Lewis, Rausch, Goldberg, & Dodd, 1968). Inclusion of a measure of the dimension in the present battery, along with several other measures of impulse expression and control, will allow 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 and must point to the one figure among the four which is <u>identical</u> to the standard. Latency to first choice and number of errors (to a maximum of two per item) are recorded.

<u>Scoring</u>

Two scores were obtained: mean response time and mean number of errors. The latencies were Windsorized to a maximum of 20 seconds and then transformed by log (X + 1) before averaging, since their distributions were positively skewed, and it appeared desirable to decrease the effect of a single unusually long latency on the score. 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.

<u>Sample Characteristics</u>

Item data were examined for a subsample consisting of the first 853 cases available for analysis. Over the various items, the number of subjects whose first response was correct ranged from 37 to 84 percent, with a median of 50 percent. The correct alternative was the modal first response for sixteen of the eighteen test items, and was nearly so for the remaining two items. The most favored distractor was chosen with a frequency ranging from 8 to 47 percent of the subsample, and a median of 25 percent. The test, therefore, appeared to possess an appropriate difficulty level for the present sample, and none of the items had unacceptable distributions of responses.

Mean response time and mean error scores were examined for possible age and sex differences. Scores were obtained for the first eight test items, the last ten test items, and all test items, to allow examination of whether any systematic differences in performance were to be found between early and later items on the test. No major differences were evident; only the total scores need be considered.

There was no sex difference for either score. Mean errors decreased with age within the sample (r = -.20), while mean response time had a negligible correlation (-.07) with age. However, when the data were examined for each three-month age group in the sample, mean response time showed a regular decrease with increasing age which was consistent across testing sites and as large--around four-tenths of one standard deviation--as that for mean errors. (See Tables 1 and 2.) This latter finding is inconsistent with expectations from other work where, over a broader age range, older children have shown longer response time and lower error scores than have young subjects.

Both the time and the error scores possessed substantial internal consistency. For response time, coefficient alpha was .90, while for the error score it was .70. However, contrary to previous findings with the reflection-impulsivity dimension, these two scores were unrelated--over the entire sample their correlation was .02, and this coefficient did not differ for males versus females, or for younger as compared to older subjects. This result suggests the need for caution in interpretation; the children in the present sample show the same consistency in response tempo which has been obtained with older children, but this variance in tempo does not appear to have the same implications for quality of performance for them as it has in older subjects.

Table 1

Mean Response Time by Age (Transformed by T = Log[T + 1])

Age	N	Mean	S.D.	Percentiles 10	25	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
Total	1399	0.60	0.12	0.45	0.51	0.59	0.67	0.74
(In seconds)		2.93		1.81	2.26	2.90	3.68	4.54

Table 2

Mean Number of Errors per Item*

Age	N	Mean	S.D.	Percentiles 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
Total	1399	0.61	0.30	0.25	0.37	0.59	0.81	1.07

*Range = 0-2

Mischel Technique

Purpose

The concept of delayed gratification or postponement of reward is derived from psychoanalytic theorizing on the organism's development of attention, choice, and other ego processes. In the experimental situation S is offered a choice between a smaller, immediate reward and a larger but delayed reward. Research by Mischel (1961), Mischel and Metzner (1962), and Mischel and Gilligan (1964) has shown stability of this measure over time. Delay of gratification has also been related to greater social responsiblity (Mischel, 1961), to higher intelligence test performance and age (Mischel & Metzner, 1962; Hess, Shipman, Brophy & Bear, 1969) and to more accurate estimates of future time perspective(Mischel & Metzner, 1962; Mischel & Gilligan, 1964). As one of several measures of impulsivity included in the present study, it also affords the opportunity to investigate further the dimensionality of impulsivity.

Task Description

In the procedure adapted from Mischel used in the present study the child was simultaneously presented with a large and small piece of candy (four vs. two sections of a Tootsie Roll). He was asked to identify the larger section and then to choose whether he wanted the smaller one now, or the larger one at the end of the morning (afternoon) testing session. An inquiry followed as to why the child chose as he did. Those children who chose the larger piece were asked to recall E's instructions when it was presented. Inclusion of an initial size identification question and a concrete specification to the child of the length of the delay seemed to be important procedural modifications for use with young children.

Scoring

Scores were obtained for correctness of the child's identification of the larger piece, for his choice, the reason for this choice, and for memory of the instructions. The only score used for the present structural analysis was choice of the immediate or delayed reward.

Sample Characteristics

Percent distribution of immediate and delayed reward choices for the total sample are presented in Table 1.

Table 1

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

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	36.8	59.8	3.4
54-56 mo.	271	31.7	67.2	1.1
57-59 mo.	63	33.3	65.1	1.6
Male	785	33.8	63.2	3.0
Female	686	36.6	61.7	1.7
Total	1471	35.1	62.5	2.5

Although the oldest groups showed greater preference for the delayed reward than the youngest groups, there was no linear relationship with age, and sex differences were negligible.

Table 2 presents distribution data for the use of different choice rationales by age and sex subgroups.

Table 2

Percentage of Children Using Different Choice Rationales by Age and Sex

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
Male	777	35.8	1.4	3.6	0.4	22.4	7.2	4.5	24.7
Female	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 Codes:

2 = Egocentric ("I like it," "I wanted 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 big/bigger/biggest; to eat now/later).

7 = Nonexclusive reason ("this is candy"; "it tastes good").

8 = Seeming Irrelevance.

9 = "Don't know"; "Because"; No answer.

Although there was a tendency for test-defined reasons to be given more by older children, other categories showed no age trends. Except for a tendency for girls to refer more often to bringing the candy home, and for boys to give more irrelevant responses, sex differences in category usage were negligible. As might be expected for this age sample, egocentric reasons were the most frequently given.

Table 3

Percent Use of Different Rationales, Classified by Choice

		Reason Categories							
Choice	N	2	3	4	5	6	7	8	9
Small Now	529	36.1	1.7	1.5	0.8	20.6	8.5	4.3	26.5
Large Later	919	35.7	1.3	5.3	0.3	23.7	7.6	3.5	22.5

The frequency of different rationales did, however, vary with the nature of the child's choice. As indicated in table 3, those children who chose the immediate reward were somewhat more likely not to give a reason or to offer an irrelevant, nonexclusive or egocentric response; children who chose the delayed reward were more likely to give a test-defined response or to say that they wanted to bring the candy home.

Remarks

One should not assume equivalence of immediate choice of the small piece with the child's inability to delay oral gratification; many testers noted that once the child had made his choice, he then saved the candy to take home. Immediate choice may reflect instead a lack of trust in the adult or in the fulfillment of expectations.

Mother-Child Interaction Tasks

(Hess and Shipman Toy-Sorting and Eight-Block Sorting Tasks)

The three tasks used in this study for investigating mother-child interaction were the Hess and Shipman Toy Sorting and Eight-Block Sorting Tasks and the Hess and Shipman Etch-A-Sketch Interaction Task. Discussion here is limited to the two sorting tasks included in the present analysis, and to only those data concerning the child's performance in the test situation and the tester's rating of his degree of cooperation during the teaching sessions.

Purpose

In studying the effect of environment on the development of the young child, it is essential to delineate the principal mechanisms of exchange that mediate between the child and his environment. In this respect, the mother may be seen as the most significant figure in the organization of the child's early experiences. One method of studying mother-child communication is by observing interaction situations structured so that the information to be conveyed to the child is held (relatively) constant for all subjects, but each mother is free to choose her preferred mode or technique of communicating this information. In the Toy Sorting and Eight-Block Sorting Tasks, mothers are asked to teach their children to sort objects in specific ways and to make clear the principles which underly the resultant groupings. Sorting tasks of this nature are particularly useful for studying the mother's ability to transmit specific information to her child, her manner of presenting the task, and her ability to recognize and adjust to difficulties which the child may experience in the situation. Also, the mother's teaching strategies are likely to have consequences for the child's ability to grasp concepts or learn lessons in other specific teaching situations; they thus affect the cognitive structures the reasons for these groupings. She was encouraged to use any method she desired to manipulate the toys as she wished. The entire teaching session was tape recorded. At the end of 15 minutes (or sooner if the mother indicated the child was ready), the child was tested. He was asked to sort the toys into the two groups his mother had shown him, and then to give his reason for sorting the toys as he did. A maximum of three trials were administered to elicit the two different sorts.

Scoring

On both the object and color sorts, scores were given for correctness of placement (0-1) and for the verbal rationale (0-2, with 1 being given for partially correct responses such as the label for one group). Points for verbalization were not given unless the child had sorted correctly. In addition, since this task was tape-recorded, scorers checked all tapes for circumstances during the teaching or testing period that might invalidate the scores obtained. The child performance scores obtained were: placement and verbalization scores for both the object and color sorts, total placement and total verbalization score, and total task score. For the overall analyses, total score only was included. For comparability to previous research data, subjects taught by other than the mother or maternal surrogate were excluded from these analyses.

Score Characteristics

Total placement and verbalization scores (object and color sorting) correlated .78 and .94 with total score; their correlation with each other was .54. For the object sort, the placement and verbalization scores correlated .68; for the color sort these scores correlated .56.

Sample Characteristics

The distribution of summed placement and verbalization scores and total score for the composite sample is presented in Table A-1.

Table A-1

Means, Standard Deviations, and Score Distributions for Placement, Verbal and Total Scores

					Score (%)							
Score	Range	N	Mean	S.D.	0	1	2	3	4	5	6	8*
Total Placement	0-2	1497	1.12	0.65	14.0	50.6	25.3					10.1
Total Verbal	0-4	1497	0.92	1.22	47.8	7.7	21.0	2.2	5.3			16.0
Total	0-6	1497	2.04	1.67	14.0	24.6	11.6	16.7	5.7	2.2	5.3	19.8

* 8 = indeterminate

The mean total score for this sample (2.04, S.D. = 1.67) was virtually identical to that obtained for the low SES subjects in the Hess and Shipman study (Mean = 2.0, S.D. = 1.65). Similar to the findings of the earlier study, a substantial percentage of children in the present sample (approximately 50%) exhibited ability at this age to categorize on this task while few were able to verbalize their reason for doing so. Although a higher percentage of the present children sorted correctly by color (55.4% vs. 50.6%), fewer were able to use color verbally as a sorting principle (15.0% vs. 20.6%).

Table 2 presents the distribution of total score by age and sex for the composite sample. Girls performed somewhat better than boys on this task, particularly with regard to providing correct rationales for sorting. As would be expected, total score increased with age.

Table A-2

Distribution of Total Score by Age and Sex

Group	N	Mean	S.D.	Score (%) 0	1	2	3	4	5	6	Indet.
42-44 mo.	100	1.59	1.63	26.0	22.0	14.0	12.0	3.0	2.0	4.0	17.0
45-57 mo.	331	1.69	1.51	16.9	30.2	8.5	16.3	3.0	2.1	2.7	20.2
48-50 mo.	344	1.91	1.60	15.1	24.1	12.2	15.1	6.1	1.7	3.8	21.8
51-53 mo.	390	2.10	1.61	11.0	27.2	13.6	17.9	6.7	2.1	5.1	16.4
54-56 mo.	272	2.52	1.80	9.2	18.4	11.8	18.0	8.5	2.2	8.8	23.2
57-59 mo.	60	2.86	2.06	13.3	13.3	8.3	21.7	5.0	6.7	15.0	16.7
Male	791	1.92	1.64	15.4	24.5	11.4	14.3	5.7	2.1	4.2	22.4
Female	706	2.17	1.70	12.5	24.8	11.9	19.4	5.8	2.3	6.5	16.9
Total	1497	2.04	1.67	14.0	24.6	11.6	16.7	5.7	2.2	5.3	19.8

The child's performance data will later be related to variables involving: mother's teaching style (the degree to which she provides specific pre-response instructions and specific post-response feedback); information-processing (use of feedback, orienting, specificity of directions); encouragement of verbalization (use of questions vs. commands); and reinforcement strategies (differential use of approval and disapproval), after partialling out an index of the child's learning ability.

B. Eight-Block Sorting Task:

Task Description

In this task the mother is to teach her child to use two criteria simultaneously in sorting eight blocks; that is, to group together blocks of the same height (tall or short) and with the same mark (X or 0), and to explain the reasons for these groupings. During the initial instruction period with the child absent, each mother was brought to the same learning criterion of three consecutive, errorless trials, each involving both placement of blocks and verbalization of the sorting principle. After completion of training, the mother was observed teaching her child. As in the toy sorting task, the mother was encouraged to teach by whatever method she thought best. After the teaching was completed (or after 25 minutes, whichever occurred first), the child was asked by the tester to place two new blocks (short O, tall X) into the appropriate group on the board and was asked to verbalize his reason for placing them where he did.

Scoring

The child's performance on the post-task test was scored on the following basis:

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

If the child placed a block correctly by one dimension he was asked if it could go anywhere else. Following his second choice, the child was then asked to indicate where the block went best. Similarly, if the child verbalized only one dimension after placing it in the correct group, he was asked if there was any other reason. In combination these scores yielded a range of 0-8 points. Points for verbalization were given only if the child had placed the block correctly according to the relevant dimension (height or mark).

The child performance scores obtained included placement and verbalization scores for each of the two test blocks (short O, tall X), total placement and verbalization scores, and total score. Total score only was used in the structural analyses.

Score Characteristics

Pearson product-moment correlations for placement and verbalization subscores with total score were .81 and .86, respectively; their correlation with each other was .32. Estimated reliabilities (coefficient alpha) for total placement and total verbal scores were .55 and .86, respectively. In scoring the protocols it was discovered that many testers had not requested and/or not indicated the child's best choice after a multiple response. In all such cases the child was given the benefit of doubt and credited with the higher score. The placement score, therefore, is inflated. This accounts partially for the lower reliability of the placement score as well as the low correlation between the placement and verbal scores.

Sample Characteristics

The distribution of summed placement and verbalization scores and total score for the composite sample is presented in Table B-1.

Table B-1

Means, Standard Deviations, and Score Distributions for Placement, Verbal and Total Scores

Score	Range	N	Mean	S.D.	0	1	2	3	4	Indet.
Total Placement	0-4	1495	3.18	1.09	3.5	4.3	17.1	19.2	54.6	1.3
Total Verbal	0-4	1495	0.86	1.29	60.9	11.0	14.6	4.4	8.0	1.0
					Percentiles	10	25	50	75	90
Total Score	0-8	1462	4.06	2.00		1.86	3.04	3.96	5.81	7.21

Although the majority of children could place the blocks correctly (72.2% for the short 0 block and 64.3% for the tall X block), few were able to verbalize correctly the sorting criteria (approximately 20% verbalized one dimension correctly and 11% both dimensions). These data are comparable to those obtained for children whose parents were in the "unskilled" low SES group in the earlier Hess and Shipman study.

Table B-2 presents the percentile distribution of total score by age and sex for the composite sample. As on the Toy Sorting Task, girls obtained higher verbal scores than boys, but the differences in total score were negligible. Scores tended to increase with age, and the difference in performance for those above and below the group's mean age was significant. The correlation obtained for total score on the Toy Sorting and Eight-Block Sorting tasks was .49.

Table B-2

Percentile Distribution of Total Score* by Age and Sex

Group	N	Mean	S.D.	Percentiles 10	25	50	75	90
42-44 mo.	95	3.74	2.00	1.37	2.11	3.85	5.10	7.05
45-57 mo.	325	3.67	1.96	1.31	2.10	3.86	5.14	6.11
48-50 mo.	335	3.89	1.92	1.85	2.20	3.91	5.27	6.16
51-53 mo.	387	4.25	1.94	1.94	3.15	4.01	5.88	7.30
54-56 mo.	266	4.42	2.14	1.89	3.17	4.07	6.02	7.89
57-59 mo.	54	4.72	1.94	2.03	3.85	4.14	6.06	7.89
Male	777	4.01	1.98	1.84	3.04	3.96	5.39	7.10
Female	685	4.11	2.03	1.87	3.04	3.97	5.83	7.33
Total	1462	4.06	2.00	1.86	3.04	3.96	5.81	7.21

*Range = 0-8

As with the Toy Sort child performance scores, the child performance scores on the Eight-Block Sorting Task will later be analyzed in relation to such maternal variables as teaching style, use of feedback, orienting, reinforcement, encouragement of verbalization, and child variables including linguistic competence, classification performance, cooperation, and motivation (e.g. persistence).

C. Interaction Ratings (Rating Scale for Child Cooperation)

Purpose

Following each of the Interaction Tasks the tester rated both the mother and child on the Fels Behavior Rating Scales for Maternal Affectionateness and

Child Cooperation (Baldwin et al., 1949). Only the latter is considered here. The Child Cooperation dimension is rated on a nine-point scale which characterizes the child's cooperation during the task as a whole and provides a useful index of the degree to which the mother finds it necessary to motivate or control the child in addition to teaching test-specific information during the interaction session. This rating is based solely on the child's attention and cooperation--independent of the actions of the mother or the child's successes or failures in task-specific responses.

This scale had been used previously by Hess and Shipman (1968) to rate child cooperation on the Eight-Block and Etch-A-Sketch Tasks. After principal component factor analyses of the child measures, the Child Cooperation ratings for the Eight-Block and Etch-A-Sketch tasks loaded consistently on a principal "Resistance" factor, together with scores for attention and percent of negative task involvement. Thus, the Child Cooperation data in the Hess and Shipman Study is reported only as a component of the Resistance factor. Correlations of this factor with child performance scores on the Interaction tasks were -0.28 (Toy Sorting), -0.30 (Eight-Block), and -0.21 (Etch-A-Sketch).

Scoring

The scale ranges from a high point for cooperation (1) defined as: "The child was fully tuned in to the mother--pliable, interested, attentive. No difficulty or conflict arose," to a rating of (9) for resistance: "child ignored the mother's teaching efforts and/or actively resisted the task throughout the interaction," with the midpoint (5) being defined as: "child was periodically inattentive, but inattention was not prolonged, and there was no resistance to the mother or the task." The child's mean cooperation rating across the three interaction tasks was the score added to the present analyses.

Score Characteristics

The 4-site correlation for ratings made on the Eight-Block Task and the Toy Sorting Task was 0.64. In the present study, the correlation between the mean child cooperation rating and the child's performance on the Eight-Block Task and the Toy Sorting Task was -0.33 and -0.26, respectively.

Sample Characteristics:

Percentile distributions for the mean ratings obtained on the Toy Sorting and Eight-Block Sorting Tasks and for the average rating across interaction tasks is presented by age in tables 1, 2 and 3, respectively.

Table C-1

Percentile Distribution of Mean Child Cooperation Ratings* on the Toy Sort Task by Age and Sex

Group	N	Mean	S.D.	Percentiles 10	25	50	75	90
42-44 mo.	49	3.47	2.34	1.00	2.07	2.97	4.86	8.02
45-57 mo.	202	2.78	2.09	1.00	1.04	2.81	3.15	6.82
48-50 mo.	232	2.78	2.02	1.00	1.08	2.34	3.17	6.16
51-53 mo.	262	2.49	1.74	1.00	1.06	2.21	3.09	5.03
54-56 mo.	187	2.37	1.86	1.00	1.01	2.06	3.05	5.05
57-59 mo.	36	2.58	1.59	1.00	2.03	2.31	3.04	4.88
Male	515	2.78	2.01	1.00	1.07	2.35	3.16	6.05
Female	453	2.51	1.86	1.00	1.04	2.19	3.09	5.07
Total	968	2.65	1.95	1.00	1.06	2.27	3.13	5.16

*range = 1-9

Table C-2

Percentile Distribution of Mean Child Cooperation Ratings* on the Eight-Block Sorting Task by Age and Sex

Group	N	Mean	S.D.	Percentiles 10	25	50	75	90
42-44 mo.	46	3.83	2.18	1.06	2.30	3.11	5.16	6.99
45-47 mo.	200	3.53	2.26	1.00	2.02	3.01	5.01	7.02
48-50 mo.	230	3.49	2.20	1.00	2.04	3.01	4.99	6.98
51-53 mo.	260	3.04	1.99	1.00	1.13	2.91	4.24	6.27
54-56 mo.	184	2.82	1.93	1.00	1.08	2.81	4.06	6.18
57-59 mo.	36	2.97	1.75	1.00	2.05	2.89	4.13	6.16
Male	507	3.42	2.18	1.00	2.02	2.99	4.94	6.97
Female	449	3.04	2.00	1.00	1.13	2.90	4.26	6.36
Total	956	3.24	2.10	1.00	1.16	2.95	4.83	6.89

*range = 1-9

Table C-3

Percentile Distribution of Mean Child Cooperation Ratings* Across Interaction Tasks

Group	N	Mean	S.D.	Percentiles 10	25	50	75	90
42-44 mo.	55	3.73	1.88	1.11	2.30	4.05	5.01	6.26
45-47 mo.	200	3.32	1.91	1.01	2.12	3.04	4.35	6.20
48-50 mo.	227	3.13	1.79	1.00	2.07	2.93	4.28	6.17
51-53 mo.	256	2.88	1.66	1.00	2.03	2.39	4.18	5.09
54-56 mo.	184	2.70	1.61	1.00	2.01	2.30	4.07	5.07
57-59 mo.	34	2.97	1.47	1.14	2.18	2.90	3.18	5.12
Male	503	3.19	1.80	1.02	2.11	2.95	4.32	6.14
Female	453	2.90	1.71	1.00	2.02	2.39	4.20	5.12
Total	956	3.05	1.76	1.00	2.06	2.88	4.25	6.05

*range = 1-9

As the mean ratings by age indicate, the youngest children in the sample exhibited the most inattention and resistance, but overall the children appeared tuned in to the task. As would be expected, cooperation decreased from task 1 to task 2, but this is somewhat confounded by the fact that the Eight-Block Sorting Task is a more demanding one. Girls, on the average, were rated as more cooperative, but the difference between groups was negligible.

Future analyses will investigate relationships between the child's cooperation and specific maternal behaviors during the teaching session as well as the predictive power for estimating the child's behavior in other learning situations.

Motor Inhibition Test

Purpose

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 impulsivity and of its implications for intellectual performance in young disadvantaged children.

Task Description

The test required the child to perform three motor acts--walking a line, drawing a line, and winding a toy jeep up to the rear of a toy tow truck. He practiced each act and then repeated it as slowly as he could. Maccoby, Dowley, Hagan, and Degerman (1965) found, with middle-class nursery school children, that the time taken under the "slow" instruction was highly correlated across tasks and that it was positively related to IQ. Their results were replicated by Massari, Hayweiser, and Meyer (1969) with lower-class preschool children, and by Ward (1968b) with eight-year-old middle-class boys. The ability to slow down a response thus appears to be either a component of general intellectual ability, or a style which contributes to performance on intellectual tasks. This ability has also been found to be related to individual differences in reflection-impulsivity (Kagan, Rosman, Day, Albert, & Phillips, 1964; Ward, 1968a).

Scoring

The data consist of six scores--for each of three subtests, log $(X + 1)$ of the time taken on the practice trial and on the "slow" instruction trial.

line and draw-a-line--the intercorrelation was .50, indicating the presence of such a dimension. The third subtest, in which the child had to wind a toy jeep up to the back of a tow truck, showed lower correlations with the first two, approximately .25. 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 poorly designed and 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, reflecting shared method variance; but there was little consistency among practice times: those from the walking and drawing subtests correlated .17 over the sample, while time from the truck subtest had near zero correlations with time from each of the others.

Correlational results showed no sex differences; for example, walk-a-line slowly and draw-a-line slowly correlated .51 for males and .49 for females. Partialling age out of the correlations also had no effect.

Conclusions

These results suggest that the most appropriate motor inhibition score 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 discarded, and the test as given in future years of this study will not include this subtest. The lack of intertask consistency in practice time indicates that there is no need to "correct" the motor inhibition score for practice time.

Sample Characteristics

Several features of these data are noteworthy. First, there were no differences in results associated with the child's age or sex. Over the six trials, age correlated with log time from -.08 to .14, while sex correlated .02 to .05.

Second, children in the present sample performed the motor acts relatively quickly. The mean number of seconds to complete the walking subtest under slow instructions was 6.4; for drawing, it was 5.9; and for the tow truck, it was 50.0. 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 on the practice trial for each task. Mean time scores under slow instructions represented an increase over practice times of 23% for the tow truck subtest, 36% for the walking subtest, and [illegible] for the drawing task. 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. 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.

Score Characteristics

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-

Open Field Test

Purpose

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 possible that there are important dimensions of behavior which are measured poorly or not at all in such situations, and which might be assessed by observing the child in a relatively unstructured play environment. 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.

Task Description

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; these were two dolls (one dark-skinned, one light), a truck, alphabet blocks, "Rising Towers" (more complex plastic building blocks), clay, crayons, felt-tipped markers, plain paper, and a 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, she recorded and described every play activity involving each object, along with a variety of nonplay activities.

Scoring and Sample Characteristics

Scores were developed to assess 1) the quality and quantity of play activities with the test objects, 2) the nature of verbalizations directed toward the tester or the child himself and 3) nonplay activities such as attempting to leave the testing room or approaching the tester.

Three aspects of the child's play activities with the ten objects were measured. The first of these was the number of half-minute periods out of the twenty during which he engaged in any play activities. This score did not discriminate well among children; the mean number of periods of play for the entire sample was 18.70 (S.D. = 3.84), and the median was 19.85. The "typical" child, then, remained involved in play throughout virtually the entire test period.

The second aspect 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 object; 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. An adequate range of scores was obtained: over the entire sample, mean complexity of play was 3.10 (S.D. = .32), and the median was 2.98.

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. Length of the longest such sequence, possibly a measure of the child's capacity for involvement in a self-imposed task, averaged 14.40 of the 20 time periods (S.D. = 5.32; median = 14.89). A comparable score was obtained for "complex" activity sequences--sequences in which at least part of the time was spent in play involving an integration of two or more of the objects. However, only slightly more than one-third of the sample engaged in any play at this level of complexity; therefore the most reasonable score involving complex activity sequences is simply whether or not any such sequences occurred for the child.

The remaining scores all concern nonplay activities during the testing session. The tester recorded all verbalizations by the child, categorizing them as directed either toward the tester or toward the child himself. Scores were obtained for each of these major categories, and also for several subcategories within each. Both of the major categories yielded usable, although skewed, distributions when scored for the number of thirty-second periods during which the child spoke. For child verbalizations directed to the tester, the overall mean was 2.21 (S.D. = 3.75), and the median was 0.44. The least talkative 25% of the sample directed no verbalizations to the tester, while the most talkative 25% spoke to her in 3.06 or more observation intervals. Similarly, for verbalizations made by the child for his own benefit, the overall mean was 2.09 (S.D. = 3.98), and the median was 0.31. The least vocal 25% of the sample did not talk to themselves at all, while the most vocal did so in 2.45 or more of the twenty periods.

Subcategories of verbalizations occurred infrequently. For child verbalizations directed to the tester, the following distinctions were made: (a) attempting to direct tester's attention to the task; (b) seeking help or direction; (c) attempting to discontinue the task; (d) other verbalizing, including nontask-oriented conversation as well as unclassifiable verbalizations. For self-directed verbalizations, the distinctions were these: (a) task-related, (b) nontask-related, and (c) unclassifiable. With the exception of the unclassifiable subcategories, none of these occurred with a median frequency larger than .10 of twenty observation periods.

The three remaining scores obtained also represented rare events. These were (a) number of periods during which the child approached or remained with the tester; (b) number of periods in which he made an overt attempt to end the task or to leave the testing room; and (c) number of periods in which he engaged in no overt activity, remaining inactive and inattentive. None of these scores had a median frequency of more than .07 out of twenty periods.

Of the scores examined above, five, because of their greater variability in these data, merit most attention in further analyses. These are mean complexity of play, length of longest simple sequence, presence or absence of complex sequences, verbalization directed toward the tester, and verbalization directed by the child to himself. None of these showed notable differences with age or with sex, although there was a tendency for males to have somewhat higher scores than females on mean complexity of play: For males, M = 3.17, S.D. = 0.35; for females, M = 3.03, S.D. = 0.27.

Reliability

Several types of reliability are relevant to this situation--the reliability of the child's behavior, the recording of this behavior by the tester, and that of the interpretation of the written record to provide scores. An estimate of the reliability of the behavior was obtained by scoring first

Insert p. 207

Peabody Picture Vocabulary Test

Purpose

The "Peabody" is one of the best known and most widely used verbal tests for the age group of interest. There are two forms, A and B, each containing 150 vocabulary items. The same response booklet is used with both forms; 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 (Dunn, 1965).

Vocabulary is a major component of many "general intelligence" tests. For Form L of the Stanford-Binet, the reported correlation between the score on the 16-item vocabulary subtest (Year IV) and total score was .74 (McNemar, 1942). The uncorrected correlation between the vocabulary score and the Full Scale Score for four-year-olds on the WPPSI was .71 (Wechsler, 1967).

Task Description

The Peabody was administered in two ways to the Longitudinal Study sample:

1. In the standard way, to obtain an estimate of receptive vocabulary, the stimulus word is presented orally and the child is required to point to one of four pictured choices. Testing is terminated after S makes six errors on eight consecutive items. For the Longitudinal Study administration, the maximum number of words administered included the first 75 stimulus words; E began always with the practice items followed by item 1. Form A was used.
2. In a modified form, so as to obtain an estimate of productive vocabulary, E pointed to one of the four pictures on a page and asked the

child to tell what it was or, in the case of verbs, to say what the child (or the children) in the picture was <u>doing</u>. The first 75 items of Form B were used.

Scoring

On Form A, each item was scored as follows: correct, incorrect, child refused to respond, or indeterminate response. A total correct score was obtained for each child up to the point where he made six errors on eight consecutive items. In addition, each child received scores based on the percentage of verbs correctly identified out of those attempted (total possible number in the first 75 items of Form A is 14) and the percentage of nouns correctly identified out of those attempted (there are 61 nouns in the first 75 items).

More than one word could serve as a correct response to an item in the modified version of Form B. Therefore, lists of acceptable synonyms were generated, and the child was given credit if he produced any one of the acceptable responses for an item. Three scores were given: total correct, percent verbs correct out of verbs attempted, and percent nouns correct out of nouns attempted. There were 14 verbs and 61 nouns. For the present structural analyses, total scores only were used.

Sample Characteristics

The mean total correct score for the conventional administration of Form A was 26.34 (N = 1198 for the four sites combined); the standard deviation was 12.85. Reliability (alpha coefficient) for 1451 cases was .96. Given the cut-off criterion in the test, this estimate is probably inflated. The alternate form reliability given in the test manual is .77 for children in the age range 4-0 to 4-5.

The mean percentages of verbs correct/verbs attempted was 57 (standard deviation 26) and of nouns correct/nouns attempted was 73 (standard deviation 11). The N's were 1439 and 1449, respectively.

Mean total scores on Form A increased consistently across the six age classifications (in three month intervals), but by varying amounts, ranging from .32 to 3.13. The results are summarized below:

Table 1

Percentile Distributions of Total Score on Form A by Age

				Percentiles				
Age	N	Mean	S.D.	10	25	50	75	90
42-44 mo.	68	22.84	10.45	10.58	15.30	20.85	30.10	36.39
45-57 mo.	273	23.50	11.95	9.45	14.33	21.83	31.56	40.46
48-50 mo.	276	25.17	12.66	9.93	14.64	23.81	34.17	44.19
51-53 mo.	313	28.30	13.32	11.18	17.86	26.76	38.85	47.01
54-56 mo.	217	28.62	13.43	11.73	17.48	26.50	39.53	48.34
57-59 mo.	51	30.72	11.37	13.63	22.88	31.16	40.92	44.46
Total	1198	26.34	12.85	10.55	15.91	24.43	36.05	45.21

Girls and boys performed at a highly similar level; Mean = 26.8 (S.D. = 12.58) for girls as compared with a mean of 25.9 (S.D. = 13.10) for boys.

When Form B was administered as a productive vocabulary test, the mean total correct score was 19.12 with a standard deviation of 8.91 for 991 cases. The reliability (alpha coefficient) was .93.

As in the case of receptive vocabulary (as measured by Form A), the children seemed to have more difficulty with the verbs than with the nouns. The mean percentage of verbs correct/verbs attempted was 49 (standard deviation 33) and the mean percentage score for the nouns was 68 (standard deviation 11).

In general, scores on the modified Form B of the Peabody increased with age. Summary information by age is given below:

Table 2

Percentile Distributions of Total Score on Form B by Age

				Percentiles				
Age	N	Mean	S.D.	10	25	50	75	90
42-44 mo.	51	15.51	7.15	7.09	9.67	14.07	20.98	25.38
45-47 mo.	221	17.22	7.84	6.57	11.95	16.52	22.24	26.37
48-50 mo.	223	17.99	8.18	7.32	11.46	18.05	23.50	28.03
51-53 mo.	249	20.14	9.64	8.35	14.12	18.87	24.53	30.95
54-56 mo.	195	21.76	9.35	10.37	15.66	21.91	26.43	35.34
57-59 mo.	52	20.69	8.97	8.27	14.51	21.54	24.55	32.30
Total	991	19.12	8.91	7.90	12.99	18.67	23.97	29.36

As previously found, children performed better on a task requiring receptive rather than productive skills.

versus second half of the test for a sample of 100 ramdomly drawn cases. For the several scores chosen to represent the task, split-half reliability coefficients were .61 for mean complexity of play, .64 for presence-absence of complex activity sequences, .81 for verbalization directed at the tester, and .73 for self-directed verbalization. This procedure was not appropriate for assessing the reliability of the length of sequence score.

Reliability of recording was obtained by having two judges simultaneously record the Open Field behavior of each of eight children. Over all categories of behavior recorded, the judges agreed exactly--i.e., on the occurrence of a given behavior in a given time period--on 80.2% of the behaviors recorded. The same index calculated over activities involving the ten play objects indicated 88.0% agreement, while that calculated over non-play activities was 41.5%. Thus, behaviors involving the play objects were recorded with reasonable reliability; the rarer non-play activities were less adequately observed.

Finally, reliability of coding--converting the tallies and descriptions on the answer sheets to the coded form from which computer-calculated scores were derived--was assessed. Here, each coder showed between 88% and 93% of possible agreements with a "standard" coding over twenty practice cases.

Picture Completion (WPPSI)

Purpose

The Picture Completion subtest of the Wechsler Preschool and Primary Scale of Intelligence is a downward extension of the identically named subtest of the Wechsler Intelligence Scale for Children. Half of the items are taken directly from the WISC and half are new (Wechsler, 1962). In studies by Goodenough and Karp (1961) and by Cohen (1959) the WISC Picture Completion subtest helped to define a factor interpreted as analytical functioning. It was included in the present study in the hope that it would, along with the Preschool Embedded Figures Test, mark such a factor of analytical functioning.

Task Description

The subtest consists of 23 pictures, each of which has some important portion missing. The child is asked to name or indicate the missing part in each picture. Testing proceeds through all 23 pictures or until there have been 5 consecutive failures.

Scoring

Summed scores were obtained for correct verbal responses, correct nonverbal responses, and total correct (including items marked correct but where insufficient information was recorded to determine the verbal versus nonverbal nature of the response). Other scores obtained but not analyzed due to their infrequency of occurrence or lack of range were number of repeats, number of elaborations, total task time, and number of indeterminate (not scorable) responses.

Sample Characteristics

Means, standard deviations, and reliabilities (coefficient alpha) are given below for the total sample (N = 1403).

Score	Mean	Standard Deviation	Reliability
Nonverbal Correct	1.34	1.92	.72
Verbal Correct	3.11	3.47	.85
Total Correct	5.45	4.71	.89

The reliability of the total score compares favorably with the reliability coefficients (corrected odd-even) reported in the WPPSI manual. These ranged from .81 to .86 for ages 4 through 6 1/2.

Score ranges and percentile distributions for the scores were:

			Percentiles (N = 1400)				
Score	Min.	Max.	10	25	50	75	90
Nonverbal Correct	0	16	0.0	0.01	0.55	2.10	4.05
Verbal Correct	0	19	0.0	0.24	1.98	5.09	8.36
Total Correct	0	21	0.13	1.59	4.34	8.47	12.49

Age breakdowns at 3-month intervals are presented in Table 1 below. A pronounced age trend is apparent. As can be seen by comparing the median scores for the study sample with the median scores from the WPPSI norm-conversion tables, the medians are consistently lower for our sample. Not shown below, but apparent from comparisons of score distributions for the two subject pools, there are considerably higher proportions of very low scores in the present study sample as compared with the WPPSI norm group but, interestingly, not an appreciably smaller proportion of high scores. There was no difference in mean performance for the two sexes.

Table 1

Percentile Distributions for Mean Total Correct by Age

				Percentiles					
Age	N	Mean	S.D.	10	25	50	WPPSI Median Raw Score*	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

*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.

Preschool Embedded Figures Test

Purpose

The Preschool Embedded Figures Test (Coates, 1969) is one of several embedded-figures measures of field dependence-independence or analytical functioning. Several studies (Cohen, 1957; 1959; Goodenough & Karp, 1961; Karp, 1963) have found separate verbal, attention-concentration, and analytical functioning factors in the Wechsler scales. More recent studies (Witkin, personal communication; Witkin, Faterson, Goodenough, & Birnbaum, 1966) reported that among individuals classified as mildly retarded, a startling number of individuals were found whose "verbal comprehension" factor scores were quite low but who had near normal prorated IQs on the "analytical" factor. Apparently the "retarded" label was applied as a function of the verbal comprehension level of the child and independently of his level of analytical functioning. In view of such findings it seemed judicious to attempt to measure analytical functioning independently of verbal comprehension.

Task Description

The test contains 27 items, 3 practice items and 24 test items, in each of which is embedded a simple equilateral triangle. The child is presented a card on which is printed the 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 taken through the three practice items a maximum of three times. If he fails to meet the criterion of two correct items on either practice trial two or practice trial three, 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. An attempt is made to administer all 24 test items.

Scoring

Scores obtained on the test were number correct, total testing time (time from beginning of practice to end of test), average time for the first response whether or not correct, and the average total time for correct responses.

It should be pointed out that the sample for this test is reduced by the number of children who could not successfully complete the practice items or otherwise failed to complete at least 75 percent of the test items.

Score Characteristics

The coefficient alpha index of reliability obtained for the total correct score was .85 for the total sample (N = 1288) and was between .85 and .86 among the four sites. The coefficient alpha for the latency measure was .77.

Means, standard deviations, and percentile distributions for the total sample are given below in Table 1 (Ns ranged from 1142 to 1288).

Table 1

Percentile Distributions of Selected Scores

Score	Mean	S.D.	10	25	50	75	90
Total number of items correct	12.1	5.55	3.9	8.6	12.8	16.1	19.5
Total testing time (minutes)	17.1	5.88	9.6	13.5	16.2	21.0	24.3
Av. time for first response (seconds)	6.6	2.95	3.1	4.5	6.3	8.3	10.5
Av. total time for correct response (seconds)	6.9	3.39	3.0	4.5	6.5	8.7	11.1

The number correct score and the total testing time were negatively correlated (r = -.35, N = 1178). Relatively independent from these two scores were the scores for response time. The two latency measures (average time to first response and average time to correct response were correlated at a high level (r = .70, N = 1291) which is not surprising since most correct responses were first responses.

Sample Characteristics

Age breakdowns for the total correct score are presented in the table below. As can be seen in this table, older children tended to get higher scores. Consistent with this result was a tendency for older children to require less time to take the test. There were no clear age trends for the latency measures.

Table 2

Percentile Distributions of Total Correct by Age

				Percentiles				
Age	N	Mean	S.D.	10	25	50	75	90
42-44 mo.	76	10.20	5.83	2.84	6.60	9.76	15.00	18.22
45-47 mo.	258	11.62	6.00	3.56	6.93	11.80	15.98	20.22
48-50 mo.	300	11.50	5.49	3.48	7.50	12.16	15.59	18.60
51-53 mo.	342	12.50	5.41	4.11	9.28	13.09	16.32	19.37
54-56 mo.	255	13.16	4.99	6.24	9.73	13.99	16.86	19.50
57-59 mo.	57	13.82	5.19	6.42	10.35	15.07	17.35	20.49
Total	1288	12.14	5.55	3.91	8.64	12.77	16.12	19.46

Preschool Inventory

Purpose

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

The original version of the instrument was developed by Bettye Caldwell, with assistance from Donald Soule for data analysis, and strong encouragement from Julius Richmond and Edmund Gordon, then of Project Head Start. It was first used in the initial summer of Head Start and has been widely administered in that program ever since. The edition used with the Longitudinal Study sample is the 1970 revised form.

Task Description

The current edition of the Inventory, frequently referred to as "the Caldwell," contains 64 items: general knowledge, 21 items; listening, word meaning, 2 items; listening, comprehension, 10 items; writing, form copying, 4 items; quantitative, 24 items; and speaking, labeling, 3 items (see ETS, PR-68-4, p. C-55 f.). The items are classified in the Inventory Manual under four main headings: Personal-Social Responsiveness (18), Associative

Vocabulary (12), Concept Activation-Numerical (15), and Concept Activation-Sensory (19). The majority of the items require an oral response from the child (e.g., "What does a dentist do?"), but some items require perceptual and/or motor responses (e.g., copying, following directions in manipulating objects, pointing to a body part).

Scoring

Each item was scored as follows: right, wrong, child said "don't know," child refused to answer, or indeterminate. The following scores were tabulated: total number of items correct, total number of items in which E had to repeat the instruction or question, total number of items to which S gave multiple answers, total number of "don't know" responses, total number of items to which S gave elaborated responses, total number of refusals to answer. In a future technical report, the characteristics and interrelationships of all of these scores will be discussed, but the present overall analyses concentrate on total correct score or, where appropriate, total correct score minus the score on the four form reproduction items. Factor analyses did not support use of separate subscores.

Sample Characteristics

For a total of 1474 cases in the four sites combined, the mean total correct score was 27.92 and the standard deviation was 11.91. The mean "don't know" score was 2.54 and the mean refusal score was 3.67.

Reliability (alpha coefficient) for 1467 cases was .92. Reliabilities for the standardization sample for the 1970 edition of the Inventory ranged from .88 for three-year-olds to .92 for children aged 5 years 6 months to 6 years 5 months.

Girls in the sample earned somewhat higher scores (mean 29.12, standard deviation 11.83, N = 694) than boys (mean 26.85, standard deviation 11.89, N = 780). The mean total scores increased quite regularly from one age level to the next. However, the scores for the Longitudinal sample seem to be slightly lower than those obtained by children of similar ages in the test standardization sample. Age data, including percentile ranks, are summarized in the table below:

Table 1

Percentile Distributions of Total Score by Age

Age	N	Mean	S.D.	Percentiles 10	25	50	75	90
42-44 mo.	89	22.27	11.44	8.16	13.41	20.67	28.94	38.05
45-47 mo.	317	24.98	10.85	11.21	17.27	22.96	31.37	40.07
48-50 mo.	348	26.41	11.45	11.68	17.71	24.46	33.33	42.40
51-53 mo.	392	28.95	11.55	13.56	20.41	27.97	35.83	44.07
54-56 mo.	270	32.08	12.04	17.00	22.28	31.17	40.82	47.20
57-59 mo.	58	35.28	12.62	17.80	25.60	34.20	44.50	50.60
Total	1474	27.92	11.91	12.38	19.03	26.60	35.55	44.38

Risk Taking 2

Purpose

Locus of control (Rotter, 1966), or the subject's belief that his actions either are or are not capable of producing consequences in the environment, has been shown to be an important motivational construct for predicting performance in school and in other intellectual and cognitive tasks (Lewis & Goldberg, 1969). Also, it has been hypothesized that individuals who feel capable of manipulating their environment and receiving consequences from that manipulation are more likely to be those willing to take risks in a risk-taking experiment. In the absence of a measure of locus of control for three- and four-year-old children, the task described here was devised to assess the possibility of relating risk-taking behavior to the variable of locus of control -- a measure which is administered in subsequent years of the study.

Task Description

The Risk Taking 2 Task investigates what children of this age will do when presented with a choice if they do not know what the results of that choice will be. The child is asked to choose between a certainty -- a toy placed in front of him -- and a paper bag which he had previously been shown might contain five toys or none at all. Small plastic cars were used for boys and small paper parasols for girls. If a child chose the certain (i.e., visible) toy he was shown that the paper bag actually contained five of those toys and he was administered a second trial with a different bag. He was again informed that the bag would either be empty or would contain five toys. The paper bag always contained the five items, and if the child did not choose the bag he was given the toys upon completion of the task.

Scoring

Responses for each trial were scored either 0 for those children choosing the toy, or certainty outcome, or 1 for those choosing the paper bag. Refusals were also noted. A derived score was computed for the structural analyses indicating degree of risk-taking (i.e., 0 for not choosing the bag on either trial, 1 for choosing the bag on trial 2, and 2 for choosing the bag on trial 1).

Sample Characteristics

Examination of the four-site total for this task showed that 60% of the subjects selected the uncertain outcome on the first trial. That is, 60% elected to take a risk and choose the paper bag rather than accept the single toy on the first trial. Of the 40% who were administered the second trial, 46.8% chose the bag in preference to the certain item. Thus, after two trials approximately 79% of the subjects at this age were willing to take a chance and choose an uncertain outcome.

The 4-site distributions by age and sex for the two trials are shown in Tables 1 and 2.

Table 1

Percent Choosing Toy or Bag on Trial 1

Group	N	0	1	7 *
42-44 mo.	84	26.2	73.8	0.0
45-47 mo.	313	46.3	53.7	0.0
48-50 mo.	343	41.4	58.6	0.0
51-53 mo.	381	39.5	60.2	0.3
54-56 mo.	266	33.8	66.2	0.0
57-59 mo.	58	46.6	53.4	0.0
Male	764	33.9	66.0	0.1
Female	681	46.7	53.3	0.0
Total	1445	39.9	60.0	0.1

*0=toy; 1=bag; 7=refusal

Table 2

Percent Choosing Toy or Bag on Trial 2

Group	N	0	1	7 *
42-44 mo.	22	50.0	50.0	0.0
45-47 mo.	140	58.6	40.7	0.7
48-50 mo.	140	50.0	50.0	0.0
51-53 mo.	150	50.0	50.0	0.0
54-56 mo.	88	55.7	44.3	0.0
57-59 mo.	26	50.0	50.0	0.0
Male	254	45.3	54.3	0.4
Female	312	59.3	40.7	0.0
Total	566	53.0	46.8	0.2

* 0=toy; 1=bag; 7=refusal

Boys were more likely to choose the uncertain outcome on the first trial; however, there was no linear relationship with age. Of those administered the second trial, 54.3% of the boys selected the paper bag and only 40.7% of the females did. The second trial also showed no linear relationship with age.

Seguin Form Board

Purpose

This test is a measure of form perception and eye-hand coordination, and is part of the Merrill-Palmer Scale of Mental Tests (Stutsman, 1931). Most normative data have been gathered on white, middle-class populations. Thus, the Longitudinal Study provides an opportunity to obtain comparative data on perceptual-motor skills in a low SES population. While the development of perceptual-motor abilities is of importance in developmental theory, "above a basic perceptual threshold, no relationship is hypothesized between perceptual skill and educational development" (ETS, PR-68-4, p. C-23). One question is whether or not low SES status does inhibit the growth of these basic abilities below a certain requisite level.

The Seguin might also be viewed as a learning measure in the area of perceptual-motor skills, since it assesses increments in performance over trials. Because speed is emphasized in the test and time and error scores are obtained, the measure might also tap stylistic as well as ability factors.

Task Description

The test materials consist of ten differently shaped blocks (circle, star, triangle, etc.) and a large form board with recesses corresponding to the various shapes. The board is placed in front of the child and the blocks are stacked in three piles and placed on the far side of the board (i.e., 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 in the board." The child is given three trials each with similar instructions indicating "how fast" or "how much faster" he can place the blocks in the board. A trial is terminated when (a) all blocks are placed correctly; (b) a three minute time period elapsed; or (c) the child indicates he is finished, even though his placements are incorrect or unfinished and he has been encouraged to continue.

Scoring

The test is scored in two ways. First, time (in seconds) required for each trial is recorded. Secondly, the number of errors for each of the three trials is recorded. An error in this case is considered to be any definite attempt to put a block into the wrong recess on the board. Because the time score distributions were skewed to the right as anticipated, all time scores were transformed by a Log 10 transformation.

Score Properties

Two derived scores from the Seguin were used in the structural analyses.

1. Quickest Time to Correct Placement out of three trials (Log 10 transformation). It should be noted that this score automatically eliminates subjects who never finished the trials and it represents each child's best performance. While there is no measure of internal consistency (alpha) for this score, trial intercorrelations for the time scores were as follows:

Table 1

Intercorrelations Among Trials of Time to Correct Placement

	Trial 2	Trial 3
Trial 1	.67	.62
Trial 2		.74

2. Number of Errors Made on "Quickest Time to Correct Placement" Test-retest estimates of reliability are seen in the intercorrelations of the Error Score over trials given in Table 2. The Error Scores intercorrelate at a slightly lower level than the time scores.

Table 2

Intercorrelations of Error Scores

	Trial 2	Trial 3
Trial 1	.63	.57
Trial 2		.66

Sample Characteristics

The log transformed Quickest Time Score for the total group (N = 1129) approximated a normal distribution, with M = 1.63, and S.D. = .19. Sex differences were negligible; largest difference in scores between groups was accounted for by age. Means, standard deviations, and percentile distributions by three-month age intervals are presented in Table 3.

Table 3

Percentile Distributions for Quickest Time to Correct Solution (Log Transformed)

Age	N	M	S.D.	Percentiles 10	25	50	75	90
42-44 mo.	55	1.74	.18	1.50	1.64	1.73	1.85	1.95
45-47 mo.	226	1.69	.19	1.46	1.56	1.69	1.81	1.97
48-50 mo.	255	1.67	.18	1.45	1.55	1.66	1.78	1.88
51-53 mo.	309	1.59	.17	1.39	1.48	1.58	1.70	1.81
54-56 mo.	229	1.56	.18	1.36	1.43	1.55	1.68	1.79
57-59 mo.	55	1.55	.18	1.31	1.40	1.56	1.68	1.79
TOTAL	1129	1.63	.19	1.40	1.50	1.62	1.75	1.88

The median error score over all three trials reported in the Merrill-Palmer manual for children at the 44.3 months age level is 9.0 (Stutsman, 1931, p. 190). The median error score over all three trials for the composite Longitudinal sample is 6.02. Granting that the Longitudinal Study is composed largely of children older than 44.3 months, and more recent data on the Merrill-Palmer might show a slightly lower Error Score, it nonetheless does not seem warranted to postulate major differences in the development of perceptual-motor abilities (as measured by Seguin) as a function of SES and/or race.

Time/Error Scores and Impulsivity. Since children are encouraged to increase their speed or go "as fast as they can" on each trial, an increase in errors over trials might be expected from the impulsive child. By and large, this did not occur, with both errors and times decreasing over trials and being positively correlated. Thus, impulsivity did not appear to affect performance for the majority of children, and the test may legitimately be viewed as an ability/learning measure. Median Errors, Median Times and their correlations are shown below for the composite 4-site sample.

Table 4

Median Error and Time Scores and Their Intercorrelation by Trial

	Md. Error	Md. Time	r
Trial 1	2.21	1.76	.59
Trial 2	2.05	1.68	.60
Trial 3	1.76	1.64	.59

Sigel Object Categorization Test

Purpose

The Sigel Object Categorization Test (SOCT) is a method of assessing young children's classification behaviors. Basically, the task assesses the criteria children employ in organizing stimulus events into categories. For example, given an array of familiar objects, we ask two questions: (1) can the child create a category and give a rationale for it, and (2) what is the criterion he uses, e.g., identical elements such as color, form or function, etc. (Sigel and Olmsted, 1968).

Research has revealed that classification behaviors and styles vary with age, sex, type of materials, and personality characteristics of the child (Sigel 1953, 1954, 1965; Sigel, Anderson & Shapiro, 1966; Sigel & McBane, 1967). Age differences indicate that older children produce a wider array of response styles for each item, with younger Ss exhibiting fewer overall responses, more stimulus-bound responses and less verbalization of rationales. Hess, Shipman, Brophy and Bear (1968, 1969) found classification responses and style positively related to maternal language style, and to vary as a function of age and social status. Comparing classification behaviors using familiar two- and three-dimensional stimuli has shown discrepancies between abilities in young children to deal with both modes of presentation and has led to work investigating representational thought (Sigel, 1968; Meyer, 1971).

The SOCT is of theoretical importance within a developmental framework as it can be used to investigate broadening of cognitive skill, changes in classificatory ability over time, and stability or change of response style as the child matures.

Task Description

The task materials consist of 12 familiar objects that may be related in various ways. The child is first asked to identify the 12 objects, and then E selects a different object on each of 12 trials, and the S is asked to select things that "are the same or like it," "belong with it" or "go with it." After each trial, S is asked why he chose the objects he did. Latency for the child's initial response on each trial is also recorded. Given the greater difficulty of the "passive sort" condition wherein E selects a group of objects and asks the child for a possible reason for the grouping (Sigel & Olmsted, 1968), only the "active sort" condition as described above was administered.

Scoring

Rationales are scored on two aspects: (1) verbal level, which indicates that the child did or did not produce a grouping response and which are scored as Grouping, Non-grouping, Nonscorable or Global depending on the appropriateness of the verbalization to the objects chosen; and (2) type of classification, which represents the child's basic rationale for the grouping. Classification scoring falls into three main styles: (a) descriptive, groupings based on objective stimulus attributes (e.g., color, form, shape); (b) relational-contextual, groupings based on functional or thematic relationships between the stimuli; and (c) categorical-inferential, groupings based on a class membership concept that involves some inference about the stimuli. Latency scores are used as an additional index of information processing. Also recorded is the adequacy of the child's initial identification of the 12 objects (scored 1-4 as correct label, appropriate label, descriptive, incorrect), and the number of different classification categories used.

Scores included in the present analyses were: (1) total number of grouping responses, (2) total number of nonscorables, (3) average latency to first response (log transformation), and (4) sum of initial correct labels given to the objects. For the structural analyses, nonscorable responses were differentiated into total number of nonscorable verbal responses (e.g., when asked why the objects go together, the child only repeated E's instructions--"they go together") and total number of nonverbal responses (e.g., non-sorts, or the child said "don't know" or refused to respond when asked for a reason). Given the low frequency of various types of grouping responses at this age, these were not analyzed separately at this time.

Score Properties

Reliability of the Sigel task using a test-retest procedure by combining responses from object and picture presentations under the active sort condition is reported as .69 for grouping responses and .71 for scorable responses (Sigel and Olmsted, no date). These suggest relative stability of performance for these scores over time with a six month test-retest interval. For the present sample, estimated reliability (coefficient alpha) and split half reliabilities (odd-even items) for the above scores were as follows:

Response	Alpha	Split-half
Total Grouping	.91	.91
Total Nonscorable	.94	.95
Total Nonscorable Verbal	.84	.94
Total Nonverbal	.85	.95
Latency	.77	.79
Total Correct Labels (only those scored 1)	.62	.69

The intercorrelations among these scores are given in Table 1.

Table 1

Intercorrelations Among Selected Scores (N = approx. 1090)

	Score	1	2	3	4	5
1	Grouping					
2	Nonscorable	-.70				
3	Latency	-.10	.15			
4	Correct Label	.25	-.16	-.19		
5	Nonscorable Verbal	-.34	.48	-.07	-.05	
6	Nonverbal	-.46	.66	.22	-.13	-.34

Sample Characteristics

Overall distributions for the selected scores were as follows:

Table 2

Percentile Distributions for Selected Sigel Scores

Score	N	Mean	S.D.	Percentiles 10	25	50	75	90
Grouping Responses	1090	3.31	3.80	.00	.07	1.57	6.18	9.30
Average Time to First Response	978	6.49	4.08	2.38	3.56	5.37	8.32	12.62
Initial Correct Labels	1091	8.77	1.76	7.05	7.68	8.93	10.17	10.73
Nonscorable Verbal Responses	1091	2.75	3.76	.00	.23	1.57	3.96	10.34
Nonverbal Responses	1091	3.50	4.38	.00	.00	1.58	6.74	10.79
Nonscorable Responses	1090	6.25	4.70	.12	1.53	5.96	11.06	12.00

Although most children were able to label the objects correctly, few were able to give appropriate verbal rationales for their sorts. Children who repeatedly gave no response when a rationale was requested were frequently incorrectly discontinued by the examiner. Also, the task was automatically discontinued in cases where four consecutive "piling" responses or "no sort" responses were given. Therefore, the frequency of nonscorable verbal and non-verbal responses is actually underrepresented in this table.

Girls consistently gave slightly more grouping responses across sites, but sex differences for this and other scores were negligible. Composite data showed grouping responses to increase with age, but there were no consistent age patterns for this or other scores across sites.

Future analyses will investigate: (1) relationships between the child's cognitive environment (particularly maternal linguistic styles) and his ability to provide verbal rationales for sorting; (2) mode of categorization as a "marker" of level of intellectual functioning; and (3) developmental trends in the child's preferred response mode.

Spontaneous Numerical Correspondence

Purpose

This instrument is intended to assess development of what Piaget (1952) describes as "intuitive approximations" to the understanding of the correspondence of number. As such, it centers on changes within the preoperational period. The instrument attempts to construct a measure that could be repeated, intact, over several years of the study, becoming combined eventually with measures of the conservation of number. Thus, in the third and fourth years of the study, Correspondence/Conservation would assess changes associated with the shift from preoperational thinking to concrete logical operations. Insofar as one-to-one correspondence is basic to primary school mathematics, the measure may tap an important precursor of later "academic" abilities.

Task Description

The task is an adaptation of a procedure described by Piaget in his book on number (1952). In the present version, the tester sets out an array of ceramic tile (1" x 1") and, providing the child with his own 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 10 tiles. In three of the presentations, the tester arranges the tiles in a straight line; in one presentation, the tiles are set out in a designated "random" arrangement.

Scoring

For each of the four items, the tester records the number and color of tiles put out by the child. A graphic record, depicting the configuration of the child's arrangement is also made. From this information, several scores are derived, two of which have been included in most of the data analyses to date. (a) "Configuration Matching." The configuration constructed by the child is coded into one of four categories ("Straight Line," "Random," "Pattern," "Restricted"). The response is scored as "matching" if the configuration is of the same type as that set up by the tester. The range for possible total matching scores is 0 to 4. (b) Deviation Score. This score is derived from the arithmetical difference between the number of tiles put out by the child and the number set out by the tester:

Score

0 = No difference (Correct)
1 = Deviation of ± 1 or ± 2
2 = Deviation of ± 3, or ± 4, or ± 5
3 = Deviation of greater than 5

Scores for any given item range from 0 to 3; the range for the total test is 0 to 12. It should be stressed that for this particular scoring system, lower scores indicate a more accurate performance in matching number.

Sample Characteristics

The distributions in Table 1 indicate considerable spread in total scores, with evidence that the task was generally a difficult one, as assessed by the deviation score. The decline in mean deviation score with age indicates that accuracy in matching number improves within the age range of the study. Matching the configuration was considerably easier than matching the number. While 38% of the children matched configuration on all four items, only 2% of the children matched the number of tiles on all four items (in other words, obtained a total deviation score of 0).

Table 1

Spontaneous Correspondence: Distribution of Total Deviation Score by Age*

Age	N	Mean	S.D.	Percentiles 10	25	50	75	90
42-44 mo.	64	7.30	3.06	3.30	5.13	7.37	9.75	11.44
45-47 mo.	267	7.26	2.95	3.53	5.29	7.13	9.62	11.46
48-50 mo.	301	7.21	3.10	2.92	4.91	7.17	9.94	11.33
51-53 mo.	335	6.64	3.34	2.54	3.89	6.42	9.48	11.32
54-56 mo.	251	6.06	3.40	1.31	3.56	5.86	8.88	10.92
57-59 mo.	56	5.82	3.62	1.16	2.50	5.70	8.75	11.18
Total	1274	6.79	3.25	2.43	4.32	6.73	9.50	11.27

*Range = 0-12.

The four test items do not appear to differ markedly in difficulty. Mean deviation scores for the items ranged from 1.61 to 1.79. The alpha coefficient for the total deviation score was .74.

Remarks

Analysis of this measure indicates that for almost all of the children in the particular age group tested, matching the configuration of an array of tile is a necessary but not a sufficient condition for accurately matching the number of tiles. Although the lines of demarcation are not sharp, it is possible to consider four categories of response-type on the task, with three of the four clearly represented in the present sample. First, there are children who respond to the tester's request by putting out a collection of tiles, by sometimes emptying the box or by just taking out a few handfuls. They comply with the request to "put out" tiles, but they generally do not match the tester's arrangement in either number or configuration (low configuration - high deviation scores). Secondly, there are many children who took pains to match the configuration but who nevertheless were not very close in matching number (high configuration - high deviation scores). Thirdly, there are children who clearly matched configuration with a precision that brought with it a close, if not exact, matching in number (high configuration - low deviation scores). Finally, there is a fourth type of performance which is very rare in the present group, but which is common in a considerably older group, in which the correct number is put out but no effort is made to match configuration (low configuration - low deviation scores). These children simply count out the tiles in complying with a request to put out the "same number." The first three patterns of responding are reported by Piaget in his analysis of development within the preoperational period. He interprets such responses as reflecting an understanding of number that is essentially perceptual in character.

TAMA General Knowledge Test*

Purpose

This test was devised specifically for the Longitudinal Study and included for the following overlapping reasons: (a) General knowledge is one of the most prevalent operational definitions of "intelligence," and general knowledge questions are featured in many of the most respected intelligence tests. (b) Certain kinds of general knowledge are essential to functioning in society, and it is one of the major responsibilities of the school and home to see that children acquire these pieces of information. (c) Other kinds of information are important to communication, social interaction, and pleasure. (d) Some general knowledge is important as a base for acquiring other general knowledge. (e) Inclusion of general knowledge measures in the study allows a more comprehensive assessment of explicit school goals than a measurement strategy limited to skills and aspects of personality and social development. (See ETS, PR-68-4, p. C-64 f.) Successively more difficult and comprehensive forms, vertically equated, were planned for use in later years of the study.

Task Description

The TAMA General Knowledge Test for 3 1/2- to 4-year-olds requires S to point to the correct picture among three pictures in response to a question from E. For example, S would be shown pictures of a spoon, a knife, and a broom and asked, "Which of these is not safe to play with?" (78 percent of the sample got that one right and the biserial correlation with total score was .54.) The response mode for the TAMA contrasts with that for the

*TAMA was derived from the last names of the team responsible for developing the test: Masako Tanaka, Scarvia Anderson, Carolyn Massad, and Dolores Ahrens.

Cooperative Preschool Inventory and the Information subtest of the WPPSI where the child must make an oral response.

The 25 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.

Scoring

Each item was scored as follows: right, wrong, child refused to answer, or indeterminate. The following scores were tabulated: total number of items correct, total number of items in which E had to repeat the question, and total number of items to which S gave elaborated responses. Although the items differ in content, only 5 of the 25 biserial correlations fell below .30. Sixteen exceeded .40, and 9 were in the range .50 to .64. This finding supported the decision to report a total score.

Sample Characteristics

For a subsample of 629 cases* across the four sites, the mean total correct score was 13.82 with a standard deviation of 3.90. Reliability (coefficient alpha) was .65.

The chart below summarizes scores and percentile points by age:

Table 1

Percentile Distributions of Total Score by Age

Age	N	Mean	S.D.	10	25	50	75	90
42-44 mo.	47	12.68	3.50	8.40	10.44	12.20	14.95	17.65
45-47 mo.	121	13.17	3.87	8.68	10.58	12.62	16.25	18.39
48-50 mo.	156	13.00	3.89	8.54	10.13	12.63	15.23	18.35
51-53 mo.	164	14.71	3.85	9.92	11.88	14.50	17.38	19.93
54-56 mo.	114	14.48	3.91	9.78	11.87	14.08	17.56	19.46
57-59 mo.	27	15.15	3.26	10.93	12.87	15.13	17.25	19.65
Total	629	13.82	3.90	9.04	10.99	13.49	16.61	19.17

*Prior to keypunching, a portion of the data were misplaced and, as yet, have not been recovered.

Except for a slight reversal in the 51-53 and 54-56 month categories, total score showed a linear relationship with age. There was no difference, however, in the mean performance of boys and girls (Mean=13.96 and 13.66, respectively).

Vigor 1 & 2

Purpose

A child's vigor or level of physical energy may influence his performance on other tasks. The low-vigor child without "energy" to respond might be regarded by a tester or teacher as poorly motivated, and the unusually energetic child might be labeled as aggressive. Two vigor tasks were included in the initial study year to determine the relationship, if any, of motor energy to the cognitive areas under investigation. It was vigor rather than mere muscular strength or motivation to do what E asks which was hopefully being tapped. It should be noted, however, that these measures are indices of "immediate" in contrast to "sustained" energy which would be measured differently and might have different correlates.

Task Description

Two tasks, with two trials each, were administered on different days to obtain some overall measure of physical energy. On the Running Task the child was asked to "run as fast as you can when I say 'go'," a space of 12 feet being marked off on the floor with lines drawn to indicate starting and stopping places.

For the Crank Turning Task the child was asked to "turn the crank as fast as you can until I say 'stop'." It had previously been established that most children could turn the crank without difficulty.

Scoring

For the running task, time was measured to the nearest .2 second. For crank turning, the number of complete revolutions in 15 seconds was recorded.

Score Characteristics

The relationship between the two vigor measures was slight, with r = -.24 between mean scores on each task. The correlations of trials 1 and 2 within tasks were quite high, however, with r = .71 for running and .76 for crank turning. During testing it was discovered that running speed was affected by a number of other factors such as fear of falling, tester differences in where they stood, closeness of walls and doors, etc. Given these confounding influences and the low correlation between tasks, mean number of crank turns was the vigor measure used in the structural analyses.

Sample Characteristics

The mean running time on trial 1 across all 4 sites (N=1429) was 2.48 seconds (S.D.=1.32); on trial 2 it was 2.19 (S.D.=1.15). The trial 2 speed was faster for all subgroups of children (classified by age and sex) across sites, clearly indicating a practice effect.

This practice effect was also evident in the crank turning task. In the 15-second period, children on their first trial (N=1470) scored a mean of 10.59 turns (S.D.=3.48), while the mean number of turns for trial 2 was 11.79 (S.D. = 3.66).

Mean running time and mean number of crank turns are presented in Tables 1 & 2 by age and sex subgroups.

Table 1

Mean Running Time by Age and Sex

Group	N	Mean	S.D.	Percentiles 10	25	50	75	90
42-44 mo.	85	2.97	1.60	1.20	2.07	3.25	4.42	5.12
45-47 mo.	299	2.39	1.05	1.20	1.29	2.84	4.04	4.77
48-50 mo.	336	2.50	1.59	1.00	1.45	2.91	4.09	4.79
51-53 mo.	378	2.34	.99	1.00	1.02	2.71	4.00	4.77
54-56 mo.	266	2.13	.95	1.00	1.00	2.26	3.74	4.63
57-59 mo.	56	2.05	.57	1.20	1.20	2.35	3.75	3.80
Male	754	2.36	1.37	1.00	1.00	2.61	3.96	4.76
Female	664	2.39	.99	1.00	1.35	2.87	4.06	4.77
Total	1418	2.38	1.21	1.00	1.07	2.74	4.01	4.77

Table 2

Mean Number of Crank Turns by Age and Sex

Group	N	Mean	S.D.	Percentiles 10	25	50	75	90
42-44 mo.	84	8.87	2.49	5.87	7.55	8.70	10.31	11.65
45-47 mo.	316	9.87	2.81	6.18	8.00	9.89	11.85	13.52
48-50 mo.	348	10.53	3.15	6.58	8.20	10.18	12.44	14.42
51-53 mo.	379	11.97	3.33	7.81	9.63	11.84	14.23	16.29
54-56 mo.	277	12.67	3.33	8.13	10.13	12.36	14.75	16.83
57-59 mo.	61	13.33	3.31	9.62	11.13	13.42	15.72	17.72
Male	776	11.45	3.43	7.56	8.96	11.48	13.81	16.11
Female	689	10.89	3.27	6.86	8.35	10.41	13.04	15.55
Total	1465	11.19	3.37	7.28	8.60	10.94	13.60	15.86

Score distributions by 3-month age intervals show a clear increase in scores with age, with older children running faster and turning the crank handle more times than younger ones. Boys showed more vigor on crank turning, but sex differences in running were negligible. The age findings suggest that these measures may be more influenced by muscle strength than originally expected.

APPENDIX C

SUPPLEMENTARY TABLES

Table C-1

Means and Standard Deviations for the Total Group

Score*	Mean	S.D.	Score*	Mean	S.D.
1	2.04	1.68	24	0.06	0.24
2	4.05	2.00	25	14.41	5.32
3	3.28	1.94	26	4.26	2.25
4	49.94	8.65	27	1.63	0.19
5	12.77	2.95	28	2.48	3.42
6	26.46	11.20	29	0.60	0.12
7	2.29	1.73	30	0.61	0.30
8	11.19	3.37	31	9.76	8.31
9	6.76	3.26	32	0.43	0.37
10	2.93	1.11	33	1.55	2.97
11	0.00	1.00	34	0.82	0.15
12	0.00	1.00	35	0.32	0.46
13	1.40	0.81	36	12.14	5.55
14	5.46	4.71	37	0.85	0.17
15	3.31	3.80	38	28.78	5.31
16	0.81	0.21	39	28.44	13.15
17	8.77	1.76	40	19.12	8.91
18	0.64	0.48	41	0.25	0.43
19	16.93	4.86	42	0.22	0.41
20	3.10	0.32	43	0.78	2.89
21	1.80	3.72	44	0.76	4.03
22	0.45	0.50	45	5.90	3.51
23	0.07	0.25	46	0.29	0.46

* See attached list for score descriptions

*Score Labels

1. Hess and Shipman Toy Sorting Task: Total Score
2. Hess and Shipman Eight-Block Sorting Task: Total Score
3. Interaction Ratings: Mean (of 1, 2, or 3 tasks) Cooperation Rating
4. Motor Inhibition Test: Average Time, Trial 2, for the Walking and Drawing Subtests
5. ETS Matched Pictures: Total Score
6. Preschool Inventory (Caldwell): Adjusted Total Score (Total Score Minus Scores for Items 52-55)
7. Form Reproduction: Total Score
8. Vigor 2 (Crank Turning): Average Number of Turns
9. Spontaneous Numerical Correspondence Task: Total Deviation Score
10. Spontaneous Numerical Correspondence Task: Total Configuration Matching
11. Massad Mimicry: Nonsense Words, Total Sounds (Standardized by Scorer)
12. Massad Mimicry: Meaningful Word Phrases, Final Sounds (Standardized by Scorer)
13. Risk-Taking 2: Derived Score (0=Toy on Both, 1=Bag on Trial 2, 2=Bag on Trial 1)
14. Picture Completion Subtest: Total Correct
15. Sigel Object Categorization: Total Grouping Responses
16. Sigel Object Categorization: Average Time to Response (Log 10)
17. Sigel Object Categorization: Total Correct Object Identification
18. Mischel Technique: Choice (0=smaller now; 1=larger later)
19. Johns Hopkins Perceptual Test: Total Correct
20. Open Field Test: Mean Play Complexity
21. Open Field Test: Number of Periods Child Talks to Himself
22. Open Field Test: Number of Periods Child Talks to Tester (1=if any, 0=none)
23. Open Field Test: Number of Periods Child Approaches Tester (1 if any)
24. Open Field Test: Number of Periods Child Attempts to Leave Task (1 if any)
25. Open Field Test: Longest Simple Sequence
26. ETS Story Sequence Task: Test Items 1 and 2 (Receptive Language) Total Score
27. Seguin Form Board: Log 10 (Fastest Time for Correct Placement (out of 3 trials))
28. Seguin Form Board: Number of Errors Made During Trial with Fastest Time for Correct Placement
29. Matching Familiar Figures: Mean Log (X+1) of Response Times for Valid Test Items
30. Matching Familiar Figures: Mean Errors Per Valid Test Item
31. Fixation: Mean Recovery Time
32. Fixation: Mean Habituation
33. Brown Self-Concept Task: Number of Items Omitted
34. Brown Self-Concept Task: Self Concept Score - No. Positive (1)/ No. Coded 0 or 1
35. Brown Self-Concept Task: Smiling (1) or not Smiling (0)
36. Preschool Embedded Figures Test: Total Correct
37. Preschool Embedded Figures Test: Average Time for First Response
38. Children's Auditory Discrimination Inventory: Total Correct
39. Peabody Picture Vocabulary Test, Form A: Total Correct to Criterion
40. Peabody Picture Vocabulary Test, Form B: Total Correct
41. Boy-Girl Identity Task: Task 1 (Girl), Item 1 Score
42. Boy-Girl Identity Task: Task 2 (Boy), Item 1 Score
43. Boy-Girl Identity Task: Sum of Task 1 Items 2, 3, 4, & 5
44. Boy-Girl Identity Task: Sum of Task 2 Items 2, 3, 4, & 5
45. Enumeration Task I: Total Correct (Items 1-12)
46. Enumeration Task I: Correct on Item 13 (Counting)

Table C-2

Correlations Among the "Cognitive" Subset of Scores (Set A)

Score	1	2	4	5	6	7	8	9	11	12	14	15	17	19	26	27	28	30	33	36	38	39	40	45	46
1		.49	.24	.30	.49	.31	.17	-.16	.24	.21	.36	.31	.19	.27	.30	-.35	-.18	-.28	-.20	.20	.28	.35	.39	.22	.23
2	.49		.30	.29	.51	.35	.17	-.11	.19	.25	.38	.37	.17	.30	.30	-.36	-.23	-.32	-.19	.21	.35	.41	.43	.25	.23
4	.24	.30		.25	.43	.33	.22	-.07	.16	.21	.35	.31	.15	.25	.29	-.34	-.19	-.29	-.15	.18	.32	.40	.40	.22	.23
5	.30	.29	.25		.39	.26	.17	-.07	.16	.22	.31	.26	.14	.25	.25	-.30	-.17	-.25	-.15	.17	.33	.38	.34	.20	.23
6	.50	.51	.43	.39		.54	.30	-.23	.37	.38	.59	.48	.33	.39	.46	-.51	-.30	-.47	-.34	.30	.51	.62	.62	.33	.38
7	.31	.35	.33	.26	.54		.26	-.21	.26	.23	.45	.34	.17	.33	.32	-.47	-.27	-.40	-.17	.31	.34	.44	.42	.34	.33
8	.17	.17	.22	.17	.30	.26		-.04	.20	.10	.26	.19	.12	.13	.18	-.24	-.07	-.24	-.21	.24	.23	.27	.23	.21	.16
9	-.16	-.11	-.07	-.07	-.23	-.21	-.04		-.01	.00	-.13	-.07	-.07	-.12	-.14	.17	.05	.09	.03	-.12	-.03	-.08	-.11	-.14	-.11
11	.24	.19	.16	.16	.37	.26	.20	-.01		.56	.25	.23	.20	.25	.25	-.27	-.18	-.30	-.14	.08	.31	.33	.33	.21	.22
12	.21	.25	.21	.22	.38	.23	.10	.00	.56		.26	.24	.18	.23	.23	-.26	-.15	-.27	-.13	.10	.30	.33	.38	.12	.21
14	.36	.38	.35	.31	.59	.45	.26	-.13	.25	.26		.38	.25	.30	.33	-.42	-.30	-.40	-.24	.29	.40	.56	.56	.30	.29
15	.31	.37	.31	.26	.48	.34	.19	-.07	.23	.24	.38		.25	.41	.37	-.34	-.20	-.41	-.17	.19	.35	.42	.43	.17	.21
17	.19	.17	.15	.14	.33	.17	.12	-.07	.20	.18	.25	.25		.26	.25	-.25	-.19	-.26	-.14	.07	.22	.31	.36	.09	.09
19	.27	.30	.25	.25	.39	.33	.13	-.12	.25	.23	.30	.41	.26		.43	-.41	-.24	-.52	-.08	.20	.28	.33	.35	.22	.18
26	.30	.30	.29	.25	.46	.32	.18	-.14	.25	.23	.33	.37	.25	.43		-.38	-.17	-.43	-.20	.16	.31	.41	.43	.22	.19
27	-.35	-.36	-.34	-.30	-.51	-.47	-.24	.17	-.27	-.26	-.42	-.34	-.25	-.41	-.38		.59	.45	.22	-.34	-.35	-.47	-.50	-.33	-.30
28	-.18	-.23	-.19	-.17	-.30	-.27	-.07	.05	-.18	-.15	-.30	-.20	-.19	-.24	-.17	.59		.25	.15	-.18	-.23	-.30	-.34	-.20	-.18
30	-.28	-.32	-.29	-.25	-.47	-.40	-.24	.09	-.30	-.27	-.40	-.41	-.26	-.52	-.43	.45	.25		.16	-.22	-.35	-.46	-.42	-.24	-.23
33	-.20	-.19	-.15	-.15	-.34	-.17	-.21	.03	-.14	-.13	-.24	-.17	-.14	-.08	-.20	.22	.15	.16		-.22	-.27	-.32	-.32	-.23	-.14
36	.20	.21	.18	.17	.30	.31	.24	-.12	.08	.10	.29	.19	.07	.20	.16	-.34	-.18	-.22	-.22		.35	.36	.17	.36	.20
38	.28	.35	.32	.33	.51	.34	.23	-.03	.31	.30	.40	.35	.22	.28	.31	-.35	-.23	-.35	-.27	.35		.61	.48	.36	.34
39	.35	.41	.40	.38	.62	.44	.27	-.08	.33	.33	.56	.42	.31	.33	.41	-.47	-.30	-.46	-.32	.36	.61		.68	.32	.30
40	.39	.43	.40	.34	.62	.42	.23	-.11	.33	.38	.56	.43	.36	.35	.43	-.50	-.34	-.42	-.34	.17	.48	.68		.24	.29
45	.22	.25	.22	.20	.33	.34	.21	-.14	.21	.12	.30	.17	.09	.22	.22	-.33	-.20	-.24	-.23	.36	.36	.32	.24		.36
46	.23	.23	.23	.23	.38	.33	.16	-.11	.22	.21	.29	.21	.09	.18	.19	-.30	-.18	-.23	-.14	.20	.34	.30	.29	.36	

Table C-3

Correlations Among Remaining Scores (Set B)

Score	3	10	13	16	18	20	21	22	23	24	25	29	31	32	34	35	37	41	42	43	44
3		-.17	.01	-.01	.04	.05	-.04	-.05	-.04	.11	-.05	-.02	-.04	-.02	-.14	-.09	.00	-.08	-.03	.06	.05
10	-.17		.05	-.05	-.08	-.04	-.08	.01	-.01	-.02	.04	-.07	-.03	.01	.09	.01	-.07	.03	.00	-.02	-.01
13	.01	.05		-.05	.05	-.02	-.02	-.03	.00	.01	.03	-.02	-.04	.00	.01	-.02	-.03	-.01	-.02	-.01	.05
16	-.01	-.05	-.05		-.02	.11	.07	.08	.06	.02	-.04	.47	.15	.09	.02	.04	.23	.00	-.02	.00	-.02
18	.04	-.08	.05	-.02		.02	.04	.02	-.03	.00	-.01	-.01	-.01	-.02	.02	.00	.05	.00	.03	-.01	.01
20	.05	-.04	-.02	.11	.02		.15	-.01	-.02	-.08	.31	.09	.07	.09	-.02	.01	.08	-.01	.00	.01	.04
21	-.04	-.08	-.02	.07	.04	.15		.19	.09	.02	.09	.11	.06	.07	.03	.08	.10	.05	.06	-.05	.00
22	-.05	.01	-.03	.08	.02	-.01	.19		.21	.18	-.12	.13	.07	.06	.02	.16	.10	.11	.05	-.06	-.06
23	-.04	-.01	.00	.06	-.03	-.02	.09	.21		.07	-.11	.07	.03	.01	.01	.11	.02	.01	.01	-.01	.00
24	.11	-.02	.01	.02	.00	-.08	.02	.18	.07		-.25	.05	.03	.04	.02	.06	.01	.02	-.02	-.01	.01
25	-.05	.04	.03	-.04	-.01	.31	.09	-.12	-.11	-.25		-.04	-.01	.00	.02	-.03	.00	.00	.04	.02	.06
29	-.02	-.07	-.02	.47	-.01	.09	.11	.13	-.07	.05	-.04		.12	.06	.02	.01	.19	-.02	-.05	-.04	-.04
31	-.04	-.03	-.04	.15	-.01	.07	.06	.07	.03	.03	-.01	.12		.53	.10	.00	.05	.04	.06	.00	-.03
32	-.02	.01	.00	.09	-.02	.09	.07	.06	.01	.04	.00	.06	.53		.09	.02	.03	.03	.04	-.03	-.01
34	-.14	.09	.01	.02	.02	-.02	.03	.02	.01	.02	.02	.02	.10	.09		.09	.03	.05	.03	.01	-.02
35	-.09	.01	-.02	.04	.00	.01	.08	.16	.11	.06	-.03	.01	.00	.02	.09		.06	.10	.06	-.05	.02
37	.00	-.07	-.03	.23	.05	.08	.10	.10	.02	.01	.00	.19	.05	.03	.03	.06		.05	.07	-.02	-.01
41	-.08	.03	-.01	.00	.00	-.01	.05	.11	.01	.02	.00	-.02	.04	.03	.05	.10	.05		.45	-.02	.02
42	-.03	.00	-.02	-.02	.03	.00	.06	.05	.01	-.02	.04	-.05	.06	.04	.03	.06	.07	.45		-.01	.04
43	.06	-.02	-.01	.00	-.01	.01	.05	-.06	-.01	-.01	.02	-.04	.00	-.03	.01	-.05	-.02	-.02	-.01		.00
44	.05	-.01	.05	-.02	.01	.04	.00	-.06	.00	.01	.06	-.04	-.03	-.01	-.02	.02	-.01	.02	.04	.01	

Table C-4

Correlations Among Set A and Set B Scores

Score	3	10	13	16	18	20	21	22	23	24	25	29	31	32	34	35	37	41	42	43	44
1	-.25	.18	.03	-.08	-.02	-.08	.01	.09	-.01	-.01	.00	-.05	.11	.10	.12	.07	.02	.13	.10	.01	-.02
2	-.33	.15	.04	-.05	-.06	-.03	.02	.11	.02	.02	-.01	-.03	.06	.01	.16	.09	.03	.17	.13	-.01	-.05
4	-.20	.15	-.02	.00	-.07	.00	.04	.09	.01	-.04	.05	.04	.09	.06	.12	.10	.08	.16	.09	.03	-.02
5	-.18	.12	-.04	-.02	.02	.00	.01	.14	.03	.02	.04	.00	.03	.02	.12	.10	.06	.12	.05	.00	-.04
6	-.31	.25	.05	-.06	-.05	-.08	.03	.18	-.02	.01	.05	-.04	.05	.08	.23	.16	.05	.22	.14	-.08	-.05
7	-.25	.25	.00	.01	-.01	-.05	-.03	.08	.00	-.03	.09	-.02	.05	.05	.19	.12	.03	.14	.07	-.01	-.03
8	-.18	.06	.00	.01	.01	.06	.05	.08	.01	-.01	.03	-.03	.07	.06	.15	.13	.06	.09	.07	.02	.00
9	.09	-.43	-.09	.06	.01	.06	.09	.01	.07	.06	-.05	.10	.03	.02	-.05	-.01	.03	-.03	-.04	.03	-.03
11	-.20	.10	.04	-.06	.00	-.09	-.01	.10	-.02	.03	-.01	-.07	-.03	-.01	.13	.12	.07	.11	.06	.02	.00
12	-.09	.04	.01	-.08	-.01	-.04	.01	.11	.00	.06	-.03	-.07	.02	.01	.11	.11	.08	.15	.12	.01	.00
14	-.27	.18	-.02	.04	-.03	.00	.05	.10	.00	-.01	.05	.05	.10	.08	.23	.12	.12	.18	.09	-.01	-.03
15	-.18	.13	-.01	-.07	.01	-.05	.02	.14	.03	.02	-.01	-.02	.10	.08	.19	.14	.01	.17	.13	-.01	-.06
17	-.09	.07	.02	-.20	-.01	-.04	-.04	-.02	-.04	-.08	-.01	-.08	.03	.01	.14	.08	.03	.08	.05	.01	-.10
19	-.15	.11	.02	-.19	.05	-.07	.03	.11	-.01	.01	.02	-.15	.05	.03	.13	.07	-.01	.13	.13	-.03	-.05
26	-.16	.19	.03	-.29	.00	-.07	.00	.10	-.03	.00	.05	-.18	.01	.02	.12	.08	.04	.12	.09	-.03	-.03
27	.25	-.20	-.04	.08	-.03	.03	.03	-.12	.05	.04	-.09	.13	-.03	-.08	-.16	-.11	.04	-.20	-.12	.04	.17
28	.14	-.06	.00	-.01	-.01	.05	.01	-.11	.00	.04	-.07	.00	-.06	-.01	-.16	-.06	-.01	-.10	-.07	.06	.12
30	.21	-.14	-.02	.11	.00	.06	-.01	-.10	-.01	.01	-.05	.02	-.08	-.05	-.18	-.11	-.07	-.13	-.10	.03	-.01
33	.20	-.10	.02	.05	.03	-.01	-.05	-.12	-.01	.03	-.07	-.01	.07	-.03	-.13	-.11	-.07	-.11	-.02	.02	.05
36	-.19	.12	-.03	.07	-.02	.05	.04	.09	.04	.02	.07	.03	.10	.11	.15	.09	.01	.12	.03	.02	-.06
38	-.23	.10	-.06	.03	-.06	-.05	.04	.20	.04	.04	.00	.04	.10	.06	.18	.14	.17	.20	.09	-.02	-.07
39	-.27	.16	-.02	.06	-.02	-.03	.03	.20	.03	.02	.00	.04	.11	.06	.27	.14	.15	.23	.10	.00	-.07
40	-.23	.12	.07	-.14	-.04	-.11	-.01	.12	-.04	.00	.03	-.05	.05	.03	.22	.11	.09	.20	.12	.01	-.04
45	-.28	.18	.00	.00	-.03	-.04	-.02	-.01	.03	-.03	.02	-.03	.05	.02	.14	.05	.00	.11	.05	.00	-.03
46	-.18	.12	-.02	-.01	-.06	-.06	.02	.05	.04	-.01	.04	.00	.04	.01	.08	.05	.04	.10	.08	-.01	-.06

Table C-5

Varimax Six-Factor Solution*

Score	1	2	3	4	5	6
1	0.50**	-0.10	0.09	0.15	0.08	0.07
2	0.55	-0.04	0.02	0.14	0.08	0.10
3	-0.38	-0.05	0.01	-0.20	0.03	0.01
4	0.49	0.05	0.04	0.07	-0.01	0.08
5	0.45	0.03	-0.01	0.06	0.06	0.05
6	0.80	-0.05	0.00	0.16	0.09	0.09
7	0.60	-0.01	0.00	0.23	-0.06	0.03
8	0.36	0.08	0.03	0.05	-0.03	0.04
9	-0.14	0.22	0.02	-0.49	0.04	-0.06
10	0.20	-0.21	-0.02	0.59	0.03	0.01
11	0.49	-0.07	-0.17	-0.21	0.18	-0.03
12	0.47	-0.06	-0.12	-0.27	0.20	0.06
13	0.01	-0.11	-0.02	0.04	-0.01	-0.01
14	0.67	0.10	0.04	0.11	-0.01	0.03
15	0.55	-0.09	0.10	-0.01	0.11	0.10
16	-0.07	0.67	0.06	0.14	0.02	-0.06
17	0.38	-0.19	0.02	-0.11	-0.01	-0.01
18	-0.02	-0.02	0.00	-0.09	-0.02	0.04
19	0.52	-0.28	0.09	-0.07	0.04	0.10
20	-0.05	0.26	0.10	-0.08	-0.35	0.07
21	0.03	0.22	0.07	-0.09	0.02	0.12
22	0.17	0.24	0.05	0.01	0.50	0.16
23	0.00	0.14	0.02	0.03	0.22	0.02
24	-0.02	0.04	0.04	-0.03	0.36	0.01
25	0.08	0.04	0.00	0.00	-0.48	0.07
26	0.53	-0.28	0.02	0.00	0.05	0.07
27	-0.71	0.15	-0.06	-0.08	0.27	-0.10
28	-0.44	-0.01	-0.03	0.01	0.14	-0.01
29	-0.03	0.56	0.05	0.05	0.09	-0.09
30	-0.63	0.15	-0.09	0.06	-0.02	-0.03
31	0.10	0.15	0.80	-0.03	0.05	0.00
32	0.07	0.09	0.61	0.00	0.02	0.02
33	-0.37	-0.08	-0.01	-0.06	0.01	-0.02
34	0.30	0.04	0.09	0.03	0.00	-0.02
35	0.18	0.09	-0.02	-0.01	0.12	0.09
36	0.40	0.14	0.09	0.20	-0.08	0.03
37	0.11	0.31	-0.02	-0.10	0.05	0.03
38	0.64	0.18	-0.02	0.01	0.12	0.04
39	0.78	0.15	0.00	0.01	0.10	0.04
40	0.75	-0.09	-0.04	-0.07	0.06	0.04
41	0.21	0.06	-0.02	-0.01	0.04	0.76
42	0.12	-0.01	0.04	-0.05	-0.01	0.54
43	-0.02	-0.02	-0.02	-0.03	-0.04	-0.03
44	-0.09	-0.03	-0.03	0.00	-0.01	0.07
45	0.44	0.04	-0.02	0.24	-0.07	-0.01
46	0.43	0.05	-0.04	0.14	0.00	0.01
*****	7.86	1.59	1.15	1.10	1.07	1.03

* Using communalities in the diagonal

** Loadings equal to or greater than .30 in absolute value are underlined

APPENDIX D

PROJECT PERSONNEL FOR THE 1971-72 STUDY YEAR

Appendix D

Project Personnel for the 1971-72 Study Year

Project Director: Virginia C. Shipman

Assistant Director: David R. Lindstrom

ETS Administrative Advisory Committee: Herman F. Smith (Chairman), Robert L. Linn

Administration:

Executive Assistant: Susan Simosko
Administrative Assistant: James Towery
Coding Supervisor: Joan Tyson
Financial Coordinator: Carol McKnight
Contract Consultants: Charlotte Farley, Gretchen Allen

Field Operations:

Lee County, Alabama
Technical Consultant: Ray Phillips
Local Coordinator: Carolyn Tamblyn
Testers and Classroom Observers

Portland, Oregon
Local Coordinators: Norma Hannam, Barbara Kerns
Testers and Classroom Observers

Trenton, New Jersey
Project Liaison: Joyce Gant
Classroom Observers

Research:

Marianne Amarel, Scarvia Anderson, Anne Bussis, Edward Chittenden, Diran Dermen, Walter Emmerich, Robert Feldmesser, Lynn Gilbert, Carolyn Massad, Masako Tanaka, William Ward, Ihor Wynnyckyj

Analysis:

Specialist for Design and Analysis: Albert E. Beaton
Coordinator of Analysis: John L. Barone
Assistants for Analysis: Thomas F. Dwyer, Norma Hvasta, Robert Patrick, Emily White

BIBLIOGRAPHY

Baldwin, A. L., Kalhorn, J., & Breese, F. H. The appraisal of parent behavior. Psychological Monographs, 1949, 63 (4, Whole No. 299).

Beaton, A. The use of special matrix operations in statistical calculus. Princeton, N. J.: Educational Testing Service, 1964, RB-64-15.

Beery, H. E. Developmental Test of Visual-Motor Integration: Administration and Scoring Manual. Chicago: Follett Education Corporation, 1967.

Beery, H. E. Visual-Motor Integration: Monograph. Chicago: Follett Educational Corporation, 1967.

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

Bever, T. G. The nature of cerebral dominance in speech behavior of the child and adult. Mechanisms of Language Development. New York: Academic Press, 1970.

Birch, H. G., & Lefford, A. Intersensory development in children. Monographs of the Society for Research in Child Development, 1963, 28 (5, Serial No. 89).

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

Brown, B. Personal communication, 1967.

Bruner, J. S., Olver, R. R., & Greenfield, P. M. Studies in cognitive growth. New York: Wiley, 1966.

Buktenica, N. A. Relative contributions of auditory and visual perception to first grade language learning. Unpublished doctoral dissertation, University of Chicago, 1966.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Carlson, A. B. Criteria for farm managers (Doctoral dissertation University of Illinois) Urbana, Illinois: University Microfilms, 1968 No. 68-1718.

Carson, A. S., & Rabin, A. Verbal comprehension and communication in Negro and white children. Journal of Educational Psychology, 1960, 51, 47-51.

Cazden, C. Some implications of research on language development for preschool education. In Hess, R. D. & Bear, R. M. (Ed.), Early Education: Current Theory, Research, and Practice. Chicago: Aldine Publishing Co., 1968. 131-142.

Chase, C. I., & Pugh, R. C. Social class and performance on an intelligence test. Journal of Educational Measurement, 1971, 8, (3), 197-202.

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

Coates, S. W. Preschool Embedded Figures Test. 1969.

Coats, S., & Braumberg, P. M. The factorial structure of the WPPSI between the ages of 4 & 6 1/2. Unpublished study.

Cohen, J. The factor 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.

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

Cronbach, L. J. Essentials of psychological testing. New York: Harper & Brothers, 1960.

Damarin, F. L., Jr., & Cattell, R. B. Personality factors in early childhood and their relation to intelligence. Monographs of the Society for Research in Child Development, 1968, 33 (6, Serial No. 122).

DeVries, 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).

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

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.

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.

Emmerich, W. Personality development and concepts of structure. Child Development, 1968, 359, 671-690.

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.

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

Ferguson, G. A. On learning and human ability. Canadian Journal of Psychology, 1954, 8, 95-112.

Ferguson, G. A. On transfer and the abilities of man. Canadian Journal of Psychology, 1956, 10, 121-131.

Flavell, J. H. Developmental psychology of Jean Piaget. Princeton, N. J.: Van Nostrand, 1963.

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.

Frostig, M., Maslow, P., Lefever, D. W., & Whittlesly, J. B. The Marianne Frostig Developmental Test of Visual Perception: 1963 standardization. Palo Alto, California: Consulting Psychologists Press, 1964.

Garrett, H. E. A developmental theory of intelligence. American Psychologist, 1946, 1, 372-378.

Goldberg, S., Godfrey, L., & Lewis, M. Play behavior in the year-old infant: Early sex differences. Paper presented at the Society for Research in Child Development Meeting, New York, March 1967.

Goodenough, D. R., & Karp, S. A. Field dependence and intellectual functioning. Journal of Abnormal and Social Psychology, 1961, 63, 241-246.

Gordon, G. Personal correspondence, 1969.

Golden, M., Birns, B., Bridger, W., & Moss, A. Social-class and differentiation in cognitive development among black preschool children. Child Development, 1971, 42, 37-45.

Guilford, J. P. The nature of human intelligence. New York: McGraw-Hill, 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.

Hess, R. D., & Shipman, V. C. Early experience and the socialization of cognitive modes in children. Child Development, 1965, 36, 869-886.

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.

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

Hunt, J. M. Intelligence and Experience. New York: Ronald Press, 1961.

Hurst, J. G. A factor analysis of the Merrill-Palmer with reference to theory and test construction. Educational and Psychological Measurement, 1960, 20, 519-532.

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

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

Kagan, J., Moss, H. A., & Sigel, I. E. Psychological significance of styles of conceptualization. In J. C. Wright and J. Kagan (Eds.), Basic cognitive processes in children. Monographs of the Society for Research in Child Development, 1963, 28 (2, Serial No. 86), 73-112.

Kagan, J., Pearson, L., & Welch, 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 reflection attitudes. Psychological Monographs, 1964, 78 (1, Whole No. 578).

Karp, S. A. Field dependence and overcoming embeddedness. Journal of Consulting Psychology, 1963, 27, 294-302.

Katz, I. A catalogue of personality approaches to Negro performance with research suggestions. Journal of Social Issues, 1969, 30, 13-28.

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

Keeney, T. Young children's imitation of nonsense syllable strings, word strings, and grammatical sentences. Unpublished paper. Educational Testing Service, 1969.

Kephart, N. C. The slow learner in the classroom. Columbus, Ohio: Charles E. Merrill Books, 1960.

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

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, Cal.: Stanford University Press, 1966. 82-173. (b)

Kohlberg, L. Early education: A cognitive-developmental view. Child Development, 1968, 39, 1013-1062.

Koppitz, E. M. The Bender-Gestalt Test for Young Children. New York: Grune and Stratton, 1964.

Lenneberg, E. H. Biological foundations of language. New York: Wiley, 1967.

Lesser, G. S., Fifer, G., & Clark, D. H. Mental abilities of children from different social-class and cultural groups. Monographs of the Society for Research in Child Development, 1965, 30 (4, Serial No. 102).

Lewin, K. Field theory in social science. New York: Harper and Row, 1951.

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, in press.

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., & 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.

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).

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

Lingoes, J. C. An IBM 7090 program for Guttman-Lingoes small space analyses: -1. Behavioral Science, 1965, 10, 183-184.

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

Lucco, A. The curiosity behavior of four-year-old children: An exploratory study. Unpublished doctoral dissertation, University of Chicago, 1964.

Maccoby, E. E. What copying requires. Paper presented at the meetings of the American Psychological Association, Washington, D. C., September 1967.

Maccoby, E. E., Dowley, E. M., Hapan, 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.

McCartin, Sister Rose Amata, & Meyers, C. E. An exploration of six semantic factors at first grade. Multivariate Behavioral Research, 1966, 1, 74-94.

McGaw, B., & Jöreskog, K. G. Factorial invariance of ability measures in groups differing in intelligence and socioeconomic status. Princeton, N. J.: Educational Testing Service, 1970, RB-70-63.

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

Melton, R. S., Charlesworth, R., Tanaka, M. N., Rothenberg, B. B., Pike, L. W., Bussis, A. M., & Gollin, E. S. Cognitive growth in preschool children. Paper presented at the meetings of the American Educational Research Association, Chicago, February 1968. (Also ETS Research Memorandum 68-13, Princeton, N. J.: Educational Testing Service, 1968.)

Messer, S. Reflection-impulsivity: stability and school failure. Journal of Educational Psychology, 1970, 61, 487-490.

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).

Meyer, W. J. Personal communication, 1971.

Meyer, W. J. The development of representational competence. Paper presented at Convention of International Society of the Study of Behavioral Development, Nijmegen, The Netherlands, July, 1971.

Meyers, C. E., Dingman, H. F., Orpet, R. E., Sitker, E. G., & Watts, C. A. Four ability-factor hypotheses at three preliterate levels in normal and retarded children. Monographs of the Society for Research in Child Development, 1964, 29 (5).

Meyers, C. E., Orpet, R. E., Atwell, A. A., & Dingman, H. F. Primary abilities at mental age six. Monographs of the Society for Research in Child Development, 1962, 27 (Whole No. 82).

Millham, J., Jacobson, L., & Berger, S. Effects of intelligence, information processing, and mediation conditions on conceptual learning. Journal of Educational Psychology, 1971, 62, 293-299.

Mischel, W. Preference for delayed reward and social responsibility. Journal of Abnormal and Social Psychology, 1961, 62, 1-7.

Mischel, W., & Gilligan, C. Delay of gratification, motivation for the prohibited gratification and responses to temptation. Journal of Abnormal and Social Psychology, 1964, 69, 411-417.

Mischel, W., & Metzner, R. Preference for delayed reward as a function of age, IQ, and length of delay interval. Journal of Abnormal and Social Psychology, 1962, 64, 425-431.

Moss, H. A. Sex, age and state as determinants of mother-infant interaction. Merrill-Palmer Quarterly, 1967, 13, 19-36.

Osborn, J. Teaching a teaching language to disadvantaged children. In M. A. Brottman (Ed.), Language remediation for the disadvantaged preschool child. 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.

Pavenstedt, E. A comparison of the child-rearing environment of upper-lower and very low-lower class families. American Journal of Orthopsychiatry, 1965, 35, 89-98.

Piaget, J. The psychology of intelligence. New York: Harcourt, Brace, 1950.

Piaget, J. The child's conception of numbers. New York: Humanities, 1952.

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

Ramsey, P. H., & Vane, J. R. A factor analytic study of the Stanford-Binet with young children. Journal of School Psychology, 1970, 8, 278-284.

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

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.

Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs: General and Applied, 1966, 80, (1, Whole No. 609).

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

Sigel, I. E. Developmental trends in the abstraction ability of children. Child Development, 1953, 24, 131-144.

Sigel, I. E The dominance of meaning. Journal of Genetic Psychology, 1954, 85, 20.-207.

Sigel, I. E. Styles of categorization in elementary school children: The role of sex differences and anxiety level. Paper presented at the meeting of the Society for Research in Child Development, Minneapolis, Minnesota, March, 1965.

Sigel, I. E. The distancing hypothesis: A hypothesis crucial to the development of representational competence. Paper presented at the meetings of the American Psychological Association, 1968.

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. Journal of Negro Education, 1966, 35, 218-229.

Sigel, I. E., & McBane, B. Cognitive competence and level of symbolization among five-year-old children. In J. Hellmuth (Ed.), The disadvantaged child, Vol. 1. Seattle, Washington: Special Child Publications, 1967, 435-453.

Sigel, I. E., & Olmsted, P. P. Modification of classificatory competence and level of representation among lower-class Negro kindergarten children. Report submitted to the Institute for Educational Development on research supported by OEO Head Start Subcontract No. 1410, November 1967. (Mimeographed)

Sigel, I. E., & Olmsted, P. P. The development of classification and representational competence. Paper presented at the meeting of the American Educational Research Association, Chicago, February, 1968.

Sigel, I. E., & Olmsted, P. P. Analysis of the Object Categorization Test and the Picture Categorization Test for preschool children. Mimeo, no date.

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. Language competencies of young children. Young Children, 1966, 22, 44-50.

Stevenson, H. W., Williams, A. M., Coleman, E. Interrelations among learning and performance task in disadvantaged children. Journal of Educational Psychology, 1971, 62, (3), 179-184.

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 (3, 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 III. Mental Measurement of Preschool Children. New York: Harcourt, Brace, & World, Inc., 1931 and 1948. 139-262. Reprinted separately.

Tannenbaum, A. J. Social and Psychological Considerations in the Study of the Socially Disadvantaged. In P. A. Witty (Ed.), The Educationally Retarded and Disadvantaged. Sixty-sixth yearbook, Part I. Chicago, Ill.: National Society for the Study of Education, 1967, 45-63.

Thurstone, L. L. The nature of intelligence. New York: Harcourt, Brace & Company, 1926.

Thurstone, L. L. The nature of intelligence. New York: Harcourt, Brace, 1950.

Tulkin, S. Race, class, family, and social achievement. Journal of Personality Social Psychology, 1968, 9, 31-37.

Wachs, T. D., Uzgiris, S. C., & Hunt, J. McV. Cognitive development in infants of different age levels and from different backgrounds: An explanatory investigation. Merrill-Palmer Quarterly, 1970, 17, (4), 283-317.

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

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

Wasik, J. L., & Wasik, B. H. A comparative study of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) and the Wechsler Intelligence Scale for Children (WISC) for culturally deprived children. Durham Educational Improvement Program, 1970, 3, 157.

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

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

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

Weksel, W. Review of the "acquisition of language." (Monograph of the Society for Research in Child Development, 1964, 29). In Language, 1965, 41, 692-709.

Werner, H. Comparative psychology of mental development. (Rev. ed.) New York: International Universities Press, 1957.

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

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

Witkin, H. A. Personal communication, 1971.

Witkin, H. A., Goodenough, D. R., & Karp, S. A. Stability of cognitive style from childhood to young adulthood. Journal of Personality and Social Psychology, 1967, 7, 291-300.

Witkin, H. A., Patterson, H. F., Goodenough, D. R., & Birnbaum, J. Cognitive patterning in mildly retarded boys. Child Development, 1966, 37, 301-316.

Zigler, E. F. Learning, development and social class in the socialization process. In M. H. Marx (Ed.), Learning: Interaction. New York: Macmillan, 1968.

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