1

44431

NATIONAL LIBRARY
OTTAWA BIBLIOTHÈQUE NATIONALE
OTTAWA

NAME OF AUTHOR..... Arlene STAIRS
TITLE OF THESIS..... Local Action Research in an Early Childhood Education
Program: Towards More Adequate Conceptualization and
More Appropriate Methodology
UNIVERSITY..... Carleton University
DEGREE FOR WHICH THESIS WAS PRESENTED..... Ph.D.
YEAR THIS DEGREE GRANTED..... 1979

Permission is hereby granted to THE NATIONAL LIBRARY
OF CANADA to microfilm this thesis and to lend or sell copies
of the film.

The author reserves other publication rights, and
neither the thesis nor extensive extracts from it may be
printed or otherwise reproduced without the author's
written permission.

(Signed).

Arlene Stairs

PERMANENT ADDRESS:

.....c/a Psychology Dept....
Carleton University
.....
.....

DATED..... May 8 19 79

NL-91 (10-68)

Supervisor..... Prof. T. Ryan

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

LOCAL ACTION RESEARCH
IN AN EARLY CHILDHOOD EDUCATION PROGRAM:
Towards More Adequate Conceptualization
and More Appropriate Methodology

by

© Arlene Stairs

Thesis submitted to the Faculty
of Graduate Studies of Carleton University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Department of Psychology

Carleton University

Ottawa, Ontario

February, 1979

The undersigned recommend to the Faculty
of Graduate Studies acceptance of the thesis
"Local Action Research in an Early Childhood Education Program:
Towards More Adequate Conceptualization and More Appropriate Methodology"
submitted by
Arlene Stairs
in partial fulfilment of the requirements for
the degree of
Doctor of Philosophy

.....
Thesis Supervisor

.....
Chairman, Department of Psychology

Carleton University

April, 1979

Abstract

Four cohorts of preschool children from a rural, isolated community were the subjects of longitudinal developmental study during and following a three-summer early educational program. Longitudinal individual assessments focussed on (a) major areas of development and achievement measured using a highly differentiated range of psychological and educational instruments, and (b) child-environment interaction characteristics rated using locally-developed items derived from current competency-based constructs and measures in early childhood research.

An action research paradigm was adopted, involving progressive changes in program and evaluation procedures, and in teacher and community participation and training. Data were analyzed primarily descriptively to characterize the project population in development-achievement and interaction areas, to follow patterns of change in these two areas, and to detect longitudinal relationships between development-achievement and interaction processes. Individual case study, particularly of extreme high- and low-achieving children, was combined with group-averaged analysis. Internal and external comparison data, rather than an experimental control group, was used in interpreting results.

The study revealed a unique pattern of developmental strengths and weaknesses in this population relative to other disadvantaged and mainstream populations studied, and marked individual variation even within this small isolated community. Original assumptions of verbal and auditorily-based, relative to nonverbal and visually-based, deficits were modified in favour of future concern with certain higher-order cognitive processes as the basic learning difficulty of these children, and possibly of other

similar groups. The importance of child-environment interactions, particularly interactions with novel and complex situations, as precursors and concomitants of school success was supported, and further study with several of the interaction characteristics rated was advocated. Considerable evidence accumulated for a community diffusion effect over the three-year program period, and for greater impact each year as the program evolved. Program effect results also revealed wide variation across developmental domains in immediacy of impact, from extremely short term (e.g., visual-motor skills) to several years' lag (e.g., reading achievement). It was argued that the range of conceptual and methodological directions which these results suggest for subsequent programming and study demonstrates the continuing value of local action research projects in early childhood education.

Acknowledgements

For more reasons than can be written down here, I want to
thank --

the children and parents of Grosse Ile, Magdalen Islands,
Quebec,

the teachers at the Grosse Ile school, especially
Geraldine Burke, Elaine Clarke and
Roma Turnbull,

the supporters of the summer preschool-kindergarten
project in Grosse Ile, including the Ministère
de l'Education du Québec and the administration
of the Regional School Board of Gaspésie,

my fellow workers in this early childhood education pro-
ject, including Marion Barnett, Barbara Cram,
Jack Cram, Carol Haney, Minna Trower, and espec-
ially Nancy Sheeran who worked extensively on
language assessment aspects of the study,

my advisor at Carleton University, Tom Ryan,

my typist, Anne Buie;

my husband, Gavin, whose least contribution is the graphic
work in this dissertation

TABLE OF CONTENTS

INTRODUCTION	1
History of Early Childhood Education and its Study ..	2
Educational Programs for Young Children	2
Origins of early childhood education and the nursery school era	2
Beginnings of the preschool era and a behavioural concept of early childhood education	8
Evolution of preschool and a developmental concept of early childhood education	13
Evaluation of Educational Programs for Young Children	27
Some Contemporary Needs and Directions in Early Childhood Education and its Study	47
Conceptualization: Developmental Constructs	48
Methodology: Research Paradigms and Techniques ...	55
Purposes of this Study	75
METHOD	81
Field Setting	81
Chronology of Program and Evaluation	84
Participants	87
Program Content	91
Assessment Procedures	93
Development and Achievement Measures	93
Interaction Process Ratings	96
Data Analyses	97
Development-Achievement Data	97
Interaction Process Data	103
Relationships between Interaction Ratings and Development-Achievement Measures	108
RESULTS	113
Observations of Parents and Teachers	114
Development-Achievement Measures	117
Intellectual Development	117
Review of intellectual development results according to project purposes (1. Characteri- zation, 2. Patterns of change, 3. Program effects)	128
Perceptual-Motor Abilities	132
Visually-based abilities	132
Auditorily-based abilities	138
Motor abilities	147
Review of perceptual-motor abilities results according to project purposes (1. Characteri- zation, 2. Patterns of change, 3. Program effects)	148

Language Acquisition	155
Review of language acquisition results according to project purposes (1. Characterization, 2. Patterns of change, 3. Program effects).....	158
Conceptual-Factual Learning and Academic Achievement	159
Conceptual-factual learning (preschool and kindergarten levels)	159
Academic achievement (grade school levels)	166
Review of learning and achievement results according to project purposes (1. Charac- terization, 2. Patterns of change, 3. Program effects)	173
Interaction Process Ratings	176
Item Usability-Answerability	176
Inter-Rater Agreement	180
Patterns of Rating Distribution	180
Relationships between Interaction Process Ratings and Development-Achievement Measures	185
Rating Item Discrimination and Prediction of Achievement Levels	185
Correlation of Selected Rating Items and Development Achievement Measures	188
Review of Interaction Ratings and their Relationships with Development-Achievement Measures according to Project Purposes (1. Characterization, 2. Patterns of change, 3, Program effects)	194
DISCUSSION	201
Development and Achievement	201
Characterization of the Project Population	201
Patterns of Change and Program Effects	211
Interaction Processes	226
Characterization of the Project Population	226
Patterns of Change and Program Effects	230
Conclusions	240
REFERENCES	246
APPENDICES	
A Development-Achievement Measures	A1-A9
B Interaction Process Rating Items	B1-B3
C Relative Change in Development-Achievement Measures as Determined with Multiple Regression (Group 2)	C1
D Profiles of Development-Achievement Measures for Highest and Lowest Ranked Children (Groups 2 and 3)	D1-D2
E Rating Distributions for Three Selected Interaction Process Items (7, 19, 30)	E1-E3

LIST OF TABLES

Table	Page
1. Peabody Picture Vocabulary Test: Mean IQ Scores and Standard Deviations, by Age-Grade Level, Group and Sex	118
2. Goodenough-Harris Drawing Test: Mean Standard Scores and Standard Deviations, by Age-Grade Level, Group and Sex	118
3. Slingerland Screening Visual Subtests (Pre-reading, A and B Levels): Mean Total Scores and Standard Deviations, by Age-Grade Level, Group and Sex	133
4. Slingerland Screening Visual Subtests (Pre-reading, A and B Levels): Ranks of the Four Visually-Based Skill Areas, by Age-Grade Level, Group and Sex	137
5. Slingerland Screening (A-B Levels): Mean Total Error Scores and Standard Deviations for (a) All Subtests (1-8), (b) Auditory Subtests (6-8), and (c) Subtests Comprising Breakoff Screening Score (3-8)	139
6. Slingerland Screening (A-B Levels): Ranks of All Subtests (1-8), by Age-Grade Level, Group and Sex ..	141-142
7. Circus 7 (How Words Sound): Mean Scores and Standard Deviations of Full (1975) and Reduced (1975, 1976) Item Pool, by Age-Grade Level, Group and Sex .	142
8. Sentence Memory, WISC Digit Span, Engelmann Digit Repetition and ITPA Auditory Sequential Memory, (Verbal Auditory Memory Measures): Mean Scores and Standard Deviations, by Age-Grade Level, Group and Sex	145-146
9. Tapping (Nonverbal Auditory Memory): Mean Scores and Standard Deviations by Age-Grade Level, Group and Sex	146
10. Gross Motor Observations (Missouri): Number of Children Over One Year Below Age Norm, by Age-Grade Level, Group and Sex	150
11. Circus 8 (How Words Work): Mean Scores and Standard Deviations by Age-Grade Level, Group and Sex	150

Table	Page
12. Metropolitan Readiness Listening Subtest: Mean Scores and Standard Deviations, by Age-Grade Level, Group and Sex	156
13. Circus 10-II (Say and Tell): Mean Scores and Standard Deviations for Full (1975) and Reduced (1975, 1976) Item Pool, by Age-Grade Level, Group and Sex	156
14. Circus 2 (How Much and How Many): Mean Scores and Standard Deviations for Full and Reduced (1975, pre-SPS, Group 3 only) Item Pool, by Age-Grade Level, Group and Sex	160
15. Opposites (Verbal Concepts): Mean Scores and Standard Deviations by Age-Grade Level, Group and Sex	160
16. Circus 11 (Do You Know): Mean Scores and Standard Deviations by Age-Grade Level, Group and Sex	164
17. Circus 5 (Finding Letters and Numbers - Capital Letter and Number Sections): Mean Scores and Standard Deviations by Age-Grade Level, Group and Sex	164
18. Early Childhood Inventory Identifications of Body Parts, Shapes and Colours: Mean Scores and Standard Deviations by Age-Grade Level, Group and Sex	164
19. Stanford Achievement Tests - Primary I and II Batteries: Mean Scores and Standard Deviations in Grade Equivalent Units, by Age-Grade Level, Group and Sex	167-168
20. Slosson Oral Reading Test: Means and Standard Deviations in Grade Equivalent Units, by Age-Grade Level, Group and Sex	169
21. Summary of Academic Achievement Results (Slosson and Stanford Tests): Mean Grade Equivalents by Age-Grade Level and Group, with Comparison Data from 1973 (post-Grade 2)	169
22. Five Category Ranking of Interaction Rating Items for Usability-Answerability, Interrater Agreement, Discrimination and Prediction of Achievement Levels	178
23. Interrater Agreement (C) and Reliability (Y) of Interaction Ratings by Age-Grade Level	179

LIST OF FIGURES

Figure	Page
1. Summer preschool-kindergarten, Grosse Ile, Magdalen Islands, Quebec: Chronology of program and evaluation, 1974-1977	85
2. Peabody Picture Vocabulary Test: Mean IQ scores by age-grade level and group, with comparison data from 1973 (grades 1 and 2) and 1977 (grades 4 and 5)	119
3. Goodenough-Harris Drawing Test: Mean standard scores by age-grade level and group	119
4. Peabody Picture Vocabulary and Goodenough-Harris Drawing tests: Mean scores of group(s) assessed at each age-grade level	126
5. Slingerland Screening visual subtests (Prereading level): Mean total scores by age-grade level and group	134
6. Slingerland Screening (A-B levels): Mean total error scores for auditory subtests, visual subtests, and all eight subtests combined, with breakoff screening scores indicated, by age-grade level and group	134
7. Circus 7 (How Words Sound): Mean scores of full (1975) and reduced (1975, 1976) item pool, by age-grade level and group	149
8. Tapping (Nonverbal auditory memory): Mean scores by age-grade level and group	149
9. Circus 8 (How Words Work): Mean scores by age-grade level and group	149
10. Circus 10-II (Say and Tell): Mean scores of full (1975) and reduced (1975-1976) item pool, by age-grade level and group	149
11. Circus 2 (How Much and How Many): Mean scores by age-grade level and group	161
12. Opposites (verbal concepts): Mean scores by age-grade level and group	161

Figure

Page

13. Circus 11 (Do you Know): Mean scores by age-grade level and group 161
14. Circus 5 (Finding Letters and Numbers - capital letter and number sections): Mean scores by age-grade level and group 161
15. Predictive (to post-grade 1) and concurrent correlations (γ) of selected interaction rating items (7, 19, 30) with development and achievement measures 190
16. Predictive correlations (γ) from selected interaction rating items (7, 19, 30) to post-grade 1 achievement (Group 2) 193

INTRODUCTION

This action research project is a longitudinal developmental study of children in an isolated rural community during and following introduction of a three-year summer preschool-kindergarten program. Action research in early childhood education (ECE) involves concepts and methods derived from two rapidly evolving social science fields: (a) child development psychology and its education applications, and (b) evaluation research. Diversity of opinion and approach, due to the multidisciplinary nature of these two fields in addition to their rapid growth, characterizes ECE action research. The context of this study, therefore, is one of active current debate among psychologists and educators. This scope and relative newness suggests that the present example of ECE action research will be better understood if its roots in child development psychology and evaluation research are traced in some detail.

The Introduction first presents the history of ECE program evolution as it corresponds to three phases in child development psychology: (a) the "maturational" nursery school phase, (b) the "behavioural" preschool phase, and (c) the now dominant "developmental" preschool phase. The attempts to evaluate ECE during these three phases of evolution are then reviewed.

Secondly, the introduction summarizes the contemporary needs and directions in ECE and its study which are considered most relevant to current, local action research. This summary comprises both (a) conceptualization, i.e., developmental constructs derived from child psychology and (b) methodology, i.e. paradigms and techniques derived from evaluation research. The specific purposes of this study are then presented.

The History of Early Childhood Education and its Study

Educational programs for young children

Origins of early childhood education and the nursery school era

The first infant school in Britain (1816) was remarkably close in concept and purpose, and in the areas of controversy it generated, to the contemporary ECE movement in North America. Its founder, Robert Owen, was fundamentally a Lockeian environmentalist who envisaged using education to achieve greater social equality. He provided a richly stimulating and informally structured educational environment for the 1- to 6-year-old children of cotton mill workers in his New Lanark model community. Children's ages and interests received careful individualized attention. Stimulated by visits to New Lanark, educators began opening infant schools in London during the next few years. The prime promoter of these schools, Samuel Wilderspin (1792-1866) became hostile towards Owen. He advocated more rigid instruction and discipline, practical demonstration and rote learning, and the dissociation of education and social change (E.G.S. Evans, 1969; Whitbread, 1972).

During these same years, education for young children evolved in continental Europe under the influence of Rousseau's child-centered maturationist ideas (Emile, 1762/1966). In Switzerland, Johann Heinrich Pestalozzi (1746-1827) pursued the conviction that education must harmonize with the child's nature at all ages and for all social classes. Sensory observation and discovery preceded verbal learning of concepts and facts in the "object lessons" which typified his teaching of young children. At the close of the Napoleonic Wars (1815), British educators including Robert Owen were able to visit Pestalozzi's castle school in Yverdon in large numbers. His influence on infant education was felt more slowly among the conservative Anglicans than it was in Ireland (e.g., Richard Lovell Edgeworth and his family; the John Synge school in County Wicklow).

Friedrich Froebel (1782-1852) worked under Pestalozzi for two years before opening his first kindergarten in Keilhau, Prussia in 1837. Froebel believed, as had his first master, in waiting until a child felt the need to learn before teaching basic skills. His Rousseauian emphasis on the value of play as a way of learning and his reverence for the mother-child relationship were romantic almost to the point of mysticism. He provided young children with six symbolic "gifts" (or "occupations" for slightly older children) of education apparatus. When kindergarten was prohibited in Prussia in 1851, Froebel's followers found that their views and methods were readily acceptable to the

romanticism of middle and upper classes in mid-19th century England. Shortly after the civil war in the United States, Susan Blow established a kindergarten in St. Louis directly modelled on Froebel's program (E.D. Evans, 1971). As kindergarten became formalized in the American public schools, Froebel's ideal met those of Dewey (1965a, 1965b), and on a smaller scale those of Montessori (see below), in a debate which anticipated the current instruction-enrichment controversy in ECE (Elkind, 1973).

About the turn of the century Maria Montessori (1870-1952) adopted some of Froebel's exercises for her Casa dei Bambini in a Rome slum-remodeling project. She emphasized (a) self-discipline rather than prizes and punishment, and (b) auto-education via carefully designed didactic materials in an orderly prepared environment. The role of sensory-motor activity in the development of intellect and as a precursor to reading and writing skills was central to Montessori's education. Despite groups of middle class enthusiasts, Montessori had little substantial influence on ECE in Britain or North America until the recent revival and republication of her works in the United States (Montessori, 1909/1964; Rambusch, 1962). The Froebelians reacted against the rejection of fantasy and the didactic rigidity which they perceived in her methods. Her influence was soon very apparent, however, in the design of educational toys for all nursery schools, public and private. Margaret McMillan (1860-1931), creator of the English Nursery school, and Susan Isaacs (1885-1948), key figure in the child study movement, were both to select from

Montessori methods and materials and to continue her concern for the individual child.

McMillan established an open-air nursery in 1913 in response to the very desperate medical and social needs of young children in British slums. She reacted against the intelligence testing which was categorizing these children as retarded. Homelike informality with small groupings, and an emphasis on imagination, use of the hands and speech characterized the McMillan nursery school (McMillan, 1930). Her goals of fostering physical and social-emotional well-being were also fundamental to the rise of nursery schools in the United States. Important model nursery school programs initiated in the late 1920's included the Gesell Child Guidance Nursery (Yale University, New Haven), the Merrill-Palmer Institute (Detroit), Teacher's College (Columbia University, New York) and the Iowa Child Welfare Research Station (University of Iowa).

Susan Isaacs was the first educator to demonstrate the concept of basing ECE on systematic psychological study of the young child (1930, 1937). The value of observation and detailed individual reports were stressed in her work from 1924 at the experimental Malting House School in Cambridge. Perhaps due to its incompatibility with her psychoanalytic training, she rejected the narrow behaviorist psychology fashionable at the time. Isaacs drew attention to the development of reasoning and language via child-initiated exploration of the environment, and to the role of teachers in meeting and facilitating such free inquiry.

Her close links with Piaget's ideas as applied to early education gave her a central place in the later British movement towards informal infant schools (Cass and Gardner, 1965; Central Advisory Council for Education, 1967; Weber, 1971).

William Blatz, leader of the nursery school movement in Canada, opened his St. George's School for Child Study in 1925 as an adjunct to the Department of Psychology at the University of Toronto (Northway, 1973). From its inception, 50 years ahead of current family-centered initiatives in ECE, the school comprised a Parent Education division as well as a Nursery School division. The school's original statement of purpose and program still endures as the theme of Canadian child study, and is of immediate relevance to the present study in anticipating contemporary action research in ECE. There was to be no division between "pure" and "applied" work at the St. George's School. The approach was very fundamentally multidisciplinary, and Blatz has been widely acclaimed for demonstrating what an interdisciplinary institute should be (Smith, 1951 in Northway, 1973). Studies were carried out in natural settings and were designed for both immediate and long-term analysis. Only by the longitudinal method, insisted Blatz, would comprehensive answers be given to problems of human growth and development. Service, i.e., inherent concern for children's well-being, was not to be sacrificed to research per se. The adult teaching program was adapted to practical needs of parents, child care workers, counsellors or other groups being taught in the school setting. Although Blatz acknowledged allegiance to no particular school of psychology, he

was a defender of Dewey and an outspoken critic of Freud and Gesell (Wright, 1974). His nursery school, closely resembling later Piagetian models, created conditions which fostered the child's natural, active tendencies to explore and create.

The establishment of North American ECE on a substantial scale outside the research institutions began during the depression of the 1930's, primarily as a means of teacher employment (E.D. Evans, 1971). Canadian contributions to ECE grew during this period with (a) the spreading influence, here and in the United States, of Blatz' Institute for Child Study and (b) the interdisciplinary, longitudinal study monitoring the development of the Dionne quintuplets following their birth in 1936. Further expansion of ECE resulted from father absence and mother employment during World War II. A group of Canadian nursery teachers, lead by Blatz under the auspices of the Mental Hygiene Committee (Clare Hincks, administrator and funder), were sent to England to train supervisors of preschool centres for children whose parents were in the war industries. This group included Mary Wright (1974, 1975a, 1975b, 1976) and other outstanding current contributors to ECE and its study in Canada.

Caretaking provision for the working class and social-emotional nurturance for the middle class remained the primary objectives throughout the nursery school era. Postwar cooperative nursery schools gave more conscious articulation to the social-emotional objectives (Kamii, 1971). This essentially non-educational child welfare orientation can be seen as the extreme end of a pendulum swing away from the 19th century denial of childhood.

As reviewed, the European pioneers of early education who began the swing were reacting against child labour and the "faculty psychology" school of strictly intellectual education (DiLorenzo, 1969). Only in the last decade has the mainstream of ECE in North America returned to more direct concern for the young child's intellect.

Beginnings of the preschool era and a behavioural concept of early childhood education

Nursery schools and enrolments for nursery-teacher training declined in the 1950's (Sigel, 1972a), perhaps a sign of the decade's relative social stability. In the 1960's a heavily funded and politicized resurgence of ECE in the form of pre-schools took place in the United States. The civil rights movement, particularly in the racially segregated south, was the most visible manifestation of a growing concern for economic and therefore educational equality which fostered the preschool era in the United States. Equal opportunity in American society, to be achieved through the prevention of school failure, was the original preschool rationale and hope (Hechinger, 1966; Hess and Baer, 1968). The acquisition of preacademic skills by children of the lowest socioeconomic status (SES) families - often single-parent, unemployed and financially dependent - were added to the earlier physical and social-emotional ECE objectives for children of more self-sufficient working families. Preschools were seen as education-oriented extensions of school whereas nursery schools had been essentially care-oriented extensions of home (Kamii, 1971).

As with nursery schools, the first preschools were set up

under near laboratory conditions on university campuses. Among the most significant of these early experimental programs were those of M. Deutsch (1964), Gray and Klaus (1965), Kirk (1958), and Weikart, Kamii and Radin (1964). In Canada the most extensive program of experimental preschool and related study was based on the curricular model of M. Deutsch (1964) and directed by Barbara Clark at Dalhousie University (Clark, 1967; Ryan, 1972, Chapter 10). Primarily in response to urgent social and political pressures, preschool was adopted in the United States on a national scale even before the laboratory programs had become fully established and documented. In 1964 the Economic Opportunities Act (Office of Economic Opportunity) was passed and led to the dramatically rapid implementation of Project Head Start one year later (Bureau of Head Start and Early Childhood, 1969). Two hundred million dollars of federal funds were allocated to Head Start, of which 90 percent was to be spent on the 1965 summer preschool programs involving half a million economically disadvantaged children. Additional pre-kindergarten programs were funded concurrently by the United States Office of Education under Title I of the Elementary and Secondary Education Act (ESEA). Head Start was unique among ESEA and other educational programs for the disadvantaged, however, in its Office of Economic Opportunity rather than Office of Education sponsorship until 1969. This administrative context reflected the vast social and economic, not just educational, expectations surrounding the beginnings of preschool in the United States (Educational Testing Service, 1968).

The second force underlying the rise of preschools in the 1960's, interactive with social and economic conditions, was a convergence of certain theoretical and empirical directions in developmental psychology. Perhaps most influential were (a) J.McV. Hunt's (1961) review of evidence demonstrating the unexpected degree of plasticity, beyond given genetic determinants, in the development of man and animals; and (b) Bloom's (1964) analysis concluding that half the variance in measures of human intelligence was determined by 4-years-of-age. Accumulated results from studies of sensory and social deprivation, enrichment, or qualitative variation in stimulation (e.g., Bennett, Diamond, Kretch and Rosenzweig, 1963; Bronfenbrenner, 1968; Dennis, 1966; Fowler, 1962; Hebb, 1949; Skeels, 1966; L.H. Stott & Ball, 1965) provided the basis for these environmentalist conclusions. Although controversial in its application to education, the growing technology and social theorizing of behaviourist psychologists (e.g., Skinner, 1961, 1968) also encouraged faith in the possibility of changing the school-relevant skills of young children and so, eventually, of changing society.

The rush into Head Start did not, however, allow time for widespread consensus on and assimilation of any experimental preschool curriculum based on environmentalist-behaviourist principles. Instead, many programs initially fell back on the well-established "bag of virtues" curriculum of the nursery schools (Burgess, 1965; Kohlberg & Turiel, 1971; Sears & Dowley, 1963). The innovative early educators advocating their various new curricula challenged this inconsistency between structured behavioural conceptualization

and loose maturational practice in the first large-scale implementation of compensatory preschool. The same criticism, regarding both curriculum and teacher training, can be directed towards the preschool programs initiated in 1964 by the Canadian Council of Jewish Women (Ryan, 1972). This preschool project, however, must be commended as the only semblance of a nationwide ECE program to be implemented even yet in Canada.

In retrospect, the behavioural conceptualization of this first optimistic preschool phase was remarkably naive. The concept relied on an essentially mechanistic model of human development as cumulative and externally determined (Kohlberg, 1968; Langer, 1969; Reese & Overton, 1970). Within this model, development consists of an increase over time in behavioural content as skills and strategies are acquired. The process of development is therefore quantifiable and so reducible to its behavioural components, each of which remains constant in meaning or function throughout life. The individual is fundamentally a passive, reactive mechanism motivated by outside forces. A child becomes what she or he is made by the environment. Disadvantage or deprivation is a deficit in culturally-valued skills due to the inadequacy of the home environment in transmitting those skills. The preschool environment and program are seen as antidotes to compensate for this lack (Cole & Bruner, 1972; J. McV. Hunt, 1964). ECE working from this mechanistic perspective thus focuses directly on manipulation of the child and expects immediate observable gains in skills selected for teaching (Bijou & Baer, 1961; Gagné, 1967; Resnick, 1967).

It was easy for compensatory education within the mechanistic model to accept the traditional paradigm of experimental research. The paradigm is ostensibly value free, but in fact implicitly adopts a common value position from the social context of the experiment. At first, easy relatively unexamined assumptions were made concerning a universal preschool objective of providing children with skills defined in terms of their immediate future success in the existing school systems. Experimental control of extraneous sources of variance was also assumed. Preschool educators began with a strong belief in the power of their treatments relative to situational and social factors. It was assumed that input at just one point could break the cycle of disadvantaged children → school failure → unemployment → low income families → another generation of disadvantaged children (Dilorenzo, 1969).

The early preschool era was further marked by extreme brevity of treatment characteristic of classical experimentation, with the anticipation of permanently eliminating cultural deficits. Such a narrow cause-effect paradigm made it possible to ignore long-term sequential patterns of change and of interaction among child, program and situation. Preschool was initially conceptualized as a unitary experimental treatment. Lack of differentiation among program components corresponded to lack of differentiation among developmental areas. Preschool children were treated as if they differed only in overall rate along a single developmental function for which a single index (e.g., IQ) would suffice. The result was an emphasis on certain intellectual-academic aspects

of development for which objectives and measures were most explicit and available, to the relative exclusion of broader cognitive, perceptual-motor, social-emotional and other aspects of the whole child. In summary, the mechanistic model within which preschool began was simplistic in its linear, decontextualized, short-term and undifferentiated cause-effect assumptions. Fallacies in the paradigm have now been repeatedly expounded from both theoretical (Guttentag, 1971; Weiss, 1972b) and empirical (Averch, Carroll, Donaldson, Kiesling & Pincus, 1971; Coleman, 1966) foundations.

Evolution of preschool and a developmental concept of early childhood education

The preschool era has been marked by an increasing range of ECE goals and program variations (Kamii, 1971; Kohlberg, 1968). Some traditional maturationist nursery schools continued to thrive during the years of enthusiasm for more environmentalist-behaviourist compensatory programs (for examples of the latter see Bereiter & Engelmann, 1966; Bushell, 1969; Karnes, 1972; Nedler, 1972). The specifically educational objectives of most early preschools supplemented rather than replaced the physical and social-emotional objectives of the nursery school era. During the past decade a developmentalist line of preschool evolution has also become prominent and has added developmental-process objectives to the earlier behavioural-product and maturational objectives for ECE.

As in any applied endeavour, no single conceptual base is adequate to all the practical aspects and levels of an operating

program (Gordon, 1972) and there are no pure examples of these three evolutionary lines. Understanding the current multitude of hybrid preschool programs in terms of maturationist, behaviorist and developmentalist influences, however, avoids the confusion of "reification" (Banta, 1966b); of considering program models to be unique and internally consistent simply because they have labels (Parker & Day, 1972). Tracing the relative contributions of the three conceptual lines also provides insight into the present evolutionary direction of ECE. It is continued evolution along the developmentalist line which, in the author's view, promises the adaptations now needed by preschool for survival and significance.

Two factors seem to underlie North Americans' increasing adoption of developmentalist elements in conceptualizing preschool. The first is growing dissatisfaction with the adequacy of behavioural conceptualizations to deal with the complex dynamics of ECE. Although a few theorists have rejected outright any compensatory goals for ECE in light of its apparent failure (Jencks, 1972; Jensen, 1969), many others are seeking a conceptual alternative to the simple treatment-outcome behavioural model. The second factor is the increasing influence of Piaget's developmental psychology, of North American psychologists relating their work to that of Piaget (e.g., Bruner, 1966; Elkind, 1970; Kohlberg, 1968; Sigel, 1969), and of the strongly Piagetian British Infant School movement and its advocates in both America and Britain (e.g., Eisner, 1974; Weber, 1971).

The intellectual tradition out of which Piaget's work arose (e.g., Dewey, 1938/1963; Montessori, 1909/1964; Werner, 1948) rejects the heredity-environment dichotomy which distinguishes maturationist and behaviourist concepts of ECE. Development is the result of interaction between structuring tendencies of the individual and the structure of the outside world, rather than reflecting either one directly. The term interaction, signifying active reciprocal causation of individual and environment in the process of development, often becomes associated with this evolutionary line in psychology and education (e.g., "developmental-interactionist", Shapiro & Biber, 1972).. The term cognition, signifying the importance of internal changes in mind or processes rather than just external changes in behaviour or products, is also often linked to the developmental concept of ECE, (e.g., "cognitive-developmental", Kohlberg, 1966, 1968). The preschool curricula which most reflect cognitive-interactive developmental concepts claim a Piagetian (e.g., Kamii, 1972; Lavatelli, 1970; Sprigle, 1970; Weikart, 1971a) or Montessorian basis. A recent review of preschool programs suggests that only these two theorists, remarkably close in their models of development (Elkind, 1967), offer guidelines for a truly comprehensive ECE program (Parker, 1974).

The developmental concept of preschool requires an organismic model of human nature. As expounded by philosophers (e.g., Polanyi, 1968) and biologists (e.g., Bertalanffy, 1962) as well as psychologists (e.g., D.B. Harris, 1957; Reese & Overton, 1970; Sameroff, 1975), two principles distinguish an organismic from a mechanistic

model of development. The first principle is emergent levels of organization. Since developmental change consists of qualitative reorganization there can be no assumption of simple continuity between the present and future function or meaning of a particular behaviour (Beeghly, Washington & Young, 1973; Bridgeman & Shipman, 1975; Emmerich, 1971; Shapiro, 1972; Sigel, 1975). ECE must, therefore, focus on complex interrelated patterns of change in behavioural characteristics which may appear long-term, indirect or curvilinear. Such "nonisomorphic" developmental change has been demonstrated, for instance, in the attention-seeking, socially expressive, competitive and aggressive behaviours of preschool children. Preschool children appear to (a) increase and then decrease in these behaviours in the course of development and/or (b) converge on optimal levels from high or low levels shown on entrance to the program (e.g., B.L. White & Watts, 1973; Wright, 1975a).

The second principle central to organismic development is intrinsic motivation or purpose. The individual is essentially active in the seeking out and ordering of his experience into successive internal structures. In a prepared responsive setting the child will spontaneously initiate activities which are well-matched to her/his existing structures of thought and behaviour (Lay, 1972; Nimnicht, 1972). Developmental preschool educators thus emphasize manipulation of the environment, not simply direct manipulation or teaching of the child. The ideal preschool provides people, materials and activities which react to the child as the child acts on them in a manner which is optimally discrepant

with her/his existing internal processes. Situational factors which are involved in the expression of skills and strategies, not simply decontextualized behaviours, must therefore be considered in the preschool program.

Disadvantage from a contextualized developmentalist perspective can be seen as a cultural difference or disparity rather than a cultural deficiency (Cole & Bruner, 1972). Deficits are considered to be artifacts of conventional psychological theory which confuses the universal biological characteristics of development with cultural characteristics (Bruner, 1972b). The compensatory function of developmental ECE is the extension of a child's existing processes to situations deemed crucial to success in the mainstream culture. Perhaps more profoundly, the focus of compensatory early education is on the child's ability to apply basic cognitive-interactive processes within both the dominant culture and the child's own subculture, and to shift appropriately according to context (Rand Corporation, 1974; Shapiro, 1973; and specifically concerning language see Ervin-Tripp, 1972; Labov, 1970). This move towards contextualization implies a return to cultural and even individual relativism in psychology and its educational applications. Acceptance of relativism is seen by many developmentalists as a sign of maturity in behavioural science which dispels the illusion of abstract generalizable laws for compensatory education (D.E. Hunt, 1975; Shapiro, 1973; Thirton, 1975).

A contextualized concept of development will not tolerate the major oversimplifications derived from a behavioural model.

of preschool intervention. Firstly, a preschool program cannot be abstracted from its larger context of value perspectives. Preschool goals are socio-political, philosophical and ethical as well as scientific and are as numerous as the constituents of any given program, e.g., the individual children, their parents, the subcultures they represent, the program personnel, the funding agencies and/or tax payers, and interested members of the educational and research communities. It may be superficially possible to design a preschool program around objective hypotheses concerning specific school-related behavioural learnings. Goals at the level of developmental process, however, involve the whole child across all behavioural domains, both in and outside of school, and must admit multiple value-laden subjective elements.

Secondly, a preschool program cannot be experimentally insulated from its context of environmental forces. Any ECE program is but one of many powerful social and physical factors significant in the development of an individual child. Likewise, preschool is but one element in the complex of causes and effects underlying collective social problems of poverty and inequality. The DARCEE "program in the child's ecology" (Miller & Camp, 1972) is notable for its early efforts in reaching beyond the "special intervention environment" of the preschool to the promotion of an "optimal maintenance environment" (Cronbach, 1969). Along with DARCEE, Gordon (1973) and Karnes (1969) pioneered the development of curricular roles for the child's primary caretaker and other persons in her or his ecological system, as well as for the child. Among the earliest preschool programs to emphasize

broad-based ecological intervention, including home economics and social work along with preschool education, were those developed in Canada by the National Council of Jewish Women (Ryan, 1972). Several of these projects attempted to differentiate the effects of preschool and family intervention components.

In his review of home-based intervention, Bronfenbrenner (1975) notes that even such broad intensive programs are seriously limited in applicability and effectiveness at the lowest end of the SES distribution. He asserts that the first and most essential requirement for ecological intervention is provision of those conditions necessary for life and for the family to function as a child-rearing system, i.e., adequate health care, nutrition, housing, employment, and opportunity and status for parenthood. The Southwest Educational Development Laboratory (Nedler, 1972) has been exceptional in this respect in preceding the development of program goals and strategies with an explicit, comprehensive context analysis involving family structure, housing, health, income, language usage, child rearing practices, educational attainment and aspiration level.

In a recent research essay prepared as part of Canadian participation in the International Year of the Child, Little and Ryan (1978) present a comprehensive outline of contextual variables relevant to human development. This interpretative review of work from many disciplines, Children in Context: The social ecology of human development, provides an orientation to the extensive interdependencies between children and their ecosystems; interdependencies which are implicated in ECE or any other,

developmental intervention. The early expectation that preschool alone would provide a lifetime inoculation against continuing effects of a poor economic, social and educational environment has now been repeatedly labeled naive and unrealistic (Abelson & Zigler, 1974; Bereiter, 1972; Gotts, 1973; Kohlberg, 1968; Miller & Dyer, 1975; Smith & James, 1975; Stern, 1968; Zigler, 1973). Developmentalists point out that since education is not characteristically a traumatic treatment, it must be part of a broad multivariate and/or relatively long-term input to stimulate profound and lasting change at the personal or social level (Bronfenbrenner, 1975; Eisner, 1972; Weiss, 1972b).

As preschool educators reject an experimental model of intervention, the time span of concern to them increases in various ways. Those who value ECE despite negative short-term results argue that preschool can be justified solely on the basis of the immediate quality of experiences provided for children, their parents and communities (Hodges, 1978; Raths, 1971; Sigel, 1972a). Others construct longer term goals for ECE emphasizing, for instance, (a) the sequential interactions of various preschool and post-preschool educational experiences (Miller & Dyer, 1975), (b) the nonlinear pattern of detectable change in certain developmental characteristics (B.L. White & Watts, 1973; Wright, 1975a), and (c) "sleeper" effects (Weikart, 1971c) such as those recently reported in large-scale follow-up studies of upper elementary level children who had experienced early educational intervention (Shipman, 1976).

Another response to discouraging preschool results is extension of the actual period of participation in compensatory education to include both younger and older children. A wide variety of infant stimulation projects consisting of home visits and/or group programs have evolved. These projects involve parents and the home as a means of providing more intense, continuous and longer intervention (e.g., Caldwell, 1969; Levenstein, 1970; Painter, 1968; B.L. White, 1968). This earlier exposure to compensatory education also responds to observations and theory indicating a younger critical period for developmental intervention. Project Follow Through was established by the United States Congress in 1967 to lengthen intervention into later childhood by providing four years of public school education articulated with Head Start goals and approaches (Gotts, 1973). Not only did Follow Through extend the time span of compensatory education but its "planned variations" design began the explicit differentiation of various component factors in preschool educational experience.

Preschool evolution under the influence of developmental concepts increasingly differentiates domains of child development, characteristics of disadvantaged development, and critical variables in preschool programming. In contrast to the narrow and essentially univariate behavioural concept of preschool, multiple objectives are specified for perceptual-motor, social-emotional and cognitive-academic domains.. Range and detail of these objectives vary according to program designers' theoretical orientations, and particularly according to their understanding

of the nature of cultural deprivation. Among the objectives of well-established developmentally-oriented preschool programs are (a) auditory discrimination in the perceptual-motor domain (e.g., C.P. Deutsch, 1964; Zigmond & Cicci, 1968); (b) self-image in the social-emotional domain (Shapiro & Biber, 1972; Coopersmith, 1968); and (c) social, physical and logico-mathematical knowledge (Kamii, 1972), and language and its psycholinguistic components (Karnes, Zehrbach & Teska, 1972) in the cognitive domain.

Recently early childhood educators have called for even greater differentiation on a lifestyle and even individual basis, claiming that the monolithic "culture of poverty" is a myth (Educational Testing Service, 1976; Kohn & Cohen, 1975; Rand Corporation, 1974). Quantitatively, Herzog, Newcomb & Cisin (1972) point out that low SES is not an all-or-none phenomenon; many levels exist in and above "poverty". Qualitatively, several recent small-scale studies (e.g., Grosse, 1972; Wright, 1975a, 1976) have found that the commonly accepted deviances of low income children from norms of, e.g., (a) rational and communicative language use and (b) motivation to explore and learn, did not accurately describe their particular disadvantaged populations.

Studies of disadvantaged children in their home environments have brought program elements beyond actual curricular content to the attention of preschool educators. Characteristics which have been found to distinguish certain disadvantaged homes from the mainstream culture involve, for instance, the physical and temporal structuring of daily life (Miller & Camp, 1972; Schoggen, 1971); and the adult-child patterns of expectations and reciprocity

(Bruner, 1972), of language and teaching style (Bernstein, 1970), and of reinforcement (Hess & Shipman, 1965). The size of groups and the nature of child-child and child-adult interactions they incorporate, the degree of structure and direction, the sequencing of activities, and the type and availability of toys and materials are among the variables which have been identified and purposely varied in consequence of these home environment studies. Group and individual heterogeneity must be considered in these organizational variables as in curricular content. For instance, an urban child's disadvantage might be related to overstimulation and a need for structured learning while a rural child's disadvantage might be related to understimulation and a need for open-ended enrichment.

To some extent as preschool evolves along developmental lines it is beginning to reintegrate this differentiated spectrum of objectives into a global developmental aim for ECE (Kohlberg & Mayer, 1972). The apparent paradox is consistent with the organismic concept of development as a dialectic process comprising (a) the differentiation of thought-behaviour structures from a global whole and (b) the articulation and integration of originally isolated structures (Sigel, 1972, p. 104; Werner, 1957). Development as a preschool goal is elaborated in terms of child and program processes rather than in terms of behavioural prescriptions and products. The same fundamental developmental processes are considered to underlie all behavioural domains, and thus the same environmental-educational conditions foster positive growth in all domains. Such a process approach to ECE cross-cuts the

conventional and increasingly unsatisfactory perceptual-motor/cognitive/affective trichotomy in child psychology and education (Bloom, Hastings & Madaus, 1971; Walker, 1973). A continuity between action and thought is assumed which leads developmentalist early educators to consider such learnings as size and shape relations (e.g., classification and seriation) to be cognitive as well as perceptual-motor in nature (Kamii, 1971). The apparently specific perceptual deficits of culturally disadvantaged children, suggests Blank (1973a, 1973b), are in fact more pervasive cognitive characteristics involving abstraction, symbolization and categorization.

Social-emotional development is also considered to be inextricable from cognitive development as they proceed together in the course of the child's interactions with the persons as well as the things in her or his surroundings. One of the few pedagogical principles directly stated by Piaget (1969; Flavell, 1963) is the importance of social collaboration in early education. The social necessity of coordinating multiple viewpoints is highly significant in bringing the child out of her/his egocentrism in all spheres of perceiving, thinking and feeling. The young child's intrinsic motivation to engage in appropriately matched activities, her or his evident "pleasure in being a cause" (Central Advisory Council for Education, 1967; McCall, 1972), pervades interactions with both physical and social world from a very early age. Several recent studies have been carried out to empirically demonstrate that young children exhibit interactive strategies, variable in quality and quantity among

individuals, which they generalize to both the physical and the social world (Radin, 1971; Samsky, 1973; Werner, 1964, 1972).

It is also being demonstrated that social-emotional factors account for a significant proportion of the preschool level variance in IQ tests designed to measure predominantly cognitive potential (Dlugokinski, Weiss & Johnston, 1976). The motivational rather than simply intellectual factors in the poor school and test performances of young culturally disadvantaged children are becoming increasingly prominent in the theory and practice of developmentalist early childhood educators (J.McV. Hunt, 1972; Zigler & Butterfield, 1968). The use to which a child puts his repertoire of strategies for relating to and exploring the world, not just her/his skills per se, is the central concern of programs which specify goals such as initiative, self-direction, resourcefulness, social cooperation, and persistence (e.g., Armington, 1968; Biber, 1967).

Hunt's proposition that mental development is the cumulative result of strategies actually used by the child from the earliest years, and the corollary that wide differences in development arise from variations in degree of use, provides one conceptual base for this integrated cognitive-affective goal orientation. Theories of play, of the "Funktionlust" (Bühler, 1934, in Bruner, 1972a) that keeps a child exploring his environment, provide another base. On the basis of extensive experience and study, Miller and Dyer (1975) join others in rejecting the separation of cognitive and noncognitive competencies in designing programs for pre-kindergarten children. They point out that according to

the conventional categorization of developmental domains, the activities and strategies in which young children engage as learners may be thought of as either personality characteristics or learning skills. For more than a decade the school establishment has carefully distinguished between cognitive and affective objectives, as manifest in the two standard taxonomies of educational objectives: I. Cognitive domain (Bloom, 1927), and II. Affective domain (Krathwohl, Bloom and Masia, 1964). Recently a proposal for integrating and balancing the two domains was published in which educational objectives approximate the diagonal rising between cognitive and affective axes (Douglas & Douglas, 1972). Such changes in the public school system perhaps reflect the developmentalist evolution in ECE.

While developmentalist educators insist that social-emotional and cognitive development are utterly inseparable (Kamii, 1972), some are concerned with the lack of systematic attention given to date to social-emotional factors (S. Cohen, 1966; Sigel, 1972b). Certainly the Piagetian principles most fundamental to developmental preschool have been explicated and applied largely in terms of cognitive processes and changes. A number of those who consider cognitive developmental theory insufficiently direct in the affective domain do, however, find it compatible with more explicit sociodramatic play (Smilansky, 1968) and psychodynamic (Biber & Franklin, 1967) concepts. The Bank Street program manifests such theoretical integration in its goal of an open-ended play orientation to intellectual experience; a goal which accepts and encourages integration of subjective and objective,

rational and irrational processes (Shapiro & Biber, 1972).

Bank Street and other leading developmentally-oriented ECE programs owe much to the work of the Isaacs in England. Beginning well before the era of Piaget's influence in English-language education, Susan (1930, 1937, 1963) and later Nathan Isaacs (1961, 1965; Weber, 1971) reconciled a great deal of psychoanalytic and cognitivist thought into a whole-child concept of development which was reflected in staff structure, space use, materials and other "function" elements of ECE programs. Very early in the history of North American ECE, William Blatz developed a Canadian program model which combined the American focus on social learning with the European interest in cognitive development (Wright, 1974). Over four decades Blatz applied this integrated cognitive-affective educational psychology (culminating in his theory of human security published in 1966) to ECE and its study. Preschool based on concepts which reintegrate the perceptual-motor/cognitive/affective trichotomy is, however, just beginning its evolution along the line begun by these avant garde developmentalist efforts. It has been suggested that the logical connectiveness of cognitive and social-emotional factors has yet to be made in developmental terms (Sigel, 1972b); that much more thought and research is needed to make "development" a useful aim and standard for ECE (Kohlberg & Mayer, 1972).

Evaluation of Educational Programs for Young Children

For much of the nursery school era judgements of worth in ECE programs, as in most other educational and social programs,

were based on traditional intuitive standards of "good practice" (Weiss, 1972b). A few exceptional nursery schools attempted more objective evaluation based on quantitative measurement of program effects (McCandless, 1967). The State University of Iowa reported Binet IQ gains, averaged across 22 different nursery school groups, of about 5 points during 6 months attendance (Wellman, 1945). The British wartime nursery schools supervised by Canadian teachers showed a similar rise in IQ scores over 18 months (Northway, 1973). A number of other superficially comparable programs during the late 1930's and the 1940's did not, however, show IQ gains. The inconsistency of these early results has been attributed to such artifacts as variations in practice and in halo effects, unanalysed distinctions among programs (e.g., teacher experience, child SES), and inappropriate statistical analyses (McNemar, 1940). McCandless (1967) summarized the limited ECE outcome data available one decade ago as indicating, "that the gains made in nursery school are permanent, and are positively correlated with later academic performance" (pp. 350-351). As the following discussion will reveal, this conclusion is being continually qualified and revised.

The experimental preschools which directly preceeded Head Start (Gray & Klaus, 1965; Weikart et al., 1964) moved further towards more formal outcome-focussed evaluation. Their reports of impressive IQ gains of 15 points or more in the space of a few months (Bronfenbrenner, 1975) were cited as evidence for political decisions regarding preschool expansion. Scientific confidence in the validity of this first wave of outcome data was limited

however. As heavy investment in and broad social objectives for preschool emerged, researchers and many other interested parties demanded a continuation and extension of accountability. Evaluation was a specified part of Head Start right from its inception, with Educational Testing Service given evaluative responsibility for the first Head Start summer. Further evaluation contracts awarded during 1966 became a network of Child Development Evaluation and Research Centers (Bureau of Head Start and Early Education, 1969). Early IQ-achievement results showed a uniform positive impact followed by a plateau during which non-Head Start children caught up to the Head Start graduates. Only where parents became intensively involved in the program or where children moved from Head Start into a superior and sympathetic school situation did the initial advantage seem to be maintained (Gotts, 1973). It was the findings of Wolff and Stern (1967) in New York City Head Start centers which first drew widespread attention to the transience of apparent gains following preschool experience.

Conclusions were difficult to draw from these Child Development and Research Center evaluations, due to mixed results and incomparability across programs. In 1968 the Office of Economic Opportunity reorganized its evaluation efforts into a large-scale coordinated study conducted by Westinghouse Learning Corporation and Ohio University (Cicirelli, 1969). The evaluation was an ex post facto study of a limited set of child outcome variables in 104 Head Start centers across the United States. Head Start subjects were matched with control subjects from their local

schools. Program variations were only minimally reflected via simple qualitative indicators. It has been suggested that such an evaluation design was acceptable only because crude survey research had become entrenched and normative in the federal system (Gotts, 1973). Except for a few regions, summer Head Start programs appeared ineffective. Full-year programs showed significant gains in readiness but no superiority in later achievement at grade 1 or 2 levels. This failure to show the dramatic results suggested by early experimental preschools was attributed to a lack of structured curriculum in most Head Start programs (Cicirelli, 1969). The positive academic effects which were seen were concentrated in southeast, Black, urban areas. This suggestion of significant variability among individual program situations was a most important contribution of the first national preschool evaluation. The affective instruments used were untried and the results essentially uninterpretable.

At about the same time as the Westinghouse results were made public, more extended evaluations were also becoming available from relatively homogeneous experimental programs (Bereiter & Engelmann, 1966; Caldwell & Smith, 1970; M. Deutsch, 1971; Hodges, McCandless & Spicker, 1967; Klaus & Gray, 1968; Southern & Plant, 1972; Sprigle, Van de Reit & Van de Reit, 1967; Weikart, 1970). While many of these showed more substantial gains in IQ and cognitive skills than were apparent in Head Start studies, most also now showed the classical decline soon after the initial spurt. Results suggested a temporary acceleration in the normal course of development, not any basic changes in level or style of

functioning. As with Head Start studies, no information regarding noncognitive variables was available from these single program evaluations in 1968 (Miller & Dyer, 1975). While these studies did allow for relating outcome to particular program variations, most were conducted by the curriculum developers and advocates themselves and therefore open to bias. Parent involvement was the one program variation which generated a substantial weight of positive evidence (Horowitz & Paden, 1973). Parent-child intervention programs showed somewhat greater maintenance of IQ gains, at least up to grade 1, with the magnitude of gains inversely related to age at entrance (e.g., Karnes, Levenstein and others in Bronfenbrenner, 1975). Children who were involved in an intensive program of parent intervention from 1- to 2-years-of-age appeared to achieve greater and more enduring gains in later preschool or school programs (Gilmour, Miller & Gray, 1970; Gordon, 1973; Radin, 1969). In a regression analysis of his longitudinal data, Weikart (1970) further supported parent involvement by showing home environment to be more important than preschool experience in predicting later achievement.

Criticisms of the first evaluation attempts were numerous and severe (Campbell & Erlebacher, 1970; Cicirelli, Evans & Schiller, 1970; J.W. Evans, 1969; Hellmuth, 1970; Smith & Bissell, 1970; Williams & Evans, 1969). The common negative conclusion drawn from accumulating evaluation results was challenged by Campbell and Frey's (1970) argument that the observed fade-out of positive preschool effects is to be anticipated even from highly successful programs. These analysts present a mathematical

model to illustrate their alternative interpretation of the result curve usually obtained. The steady decline which follows preschool children's immediate growth spurt may represent learning effects which are in fact maintained but which are cumulatively counterbalanced by other effects as the intellectual quality of the environment falls back to its original level. The most important and still continuing (House, Glass, McLean & Walker, 1978) design and analysis controversy surrounds the use of experimental controls. It is argued that neither matching nor statistical analysis (covariance correction) made participant and control children strictly comparable. Preschool programs have been made to appear ineffectual and even detrimental as a result of (a) the characteristic undercorrection of covariance analysis and/or (b) the initial superiority of controls over participants in either variables used as post-treatment outcome indicators, or in unmeasured variables closely and causally tied to outcome indicators. Self-selection and attrition, with their related motivational and broad ecological factors, further complicate the interpretation of comparisons between control and treatment groups. The use of distal but similar communities, national norms, local pretest norms and time-series designs have been both advised and criticized as solutions to the control group problem (Bronfenbrenner, 1975; Campbell, 1967; Vane, 1976). The classical experimental solution of truly randomized assignment to treatment and control groups is most often socially and politically undesirable, impossible or illusory in actual preschool situations. Randomization has, however, been seriously advocated with detailed

practical recommendations by several leading evaluators (Cook & Campbell, in press; and see Boruch, 1974 in Rand, 1974 for a bibliography of randomized field experiments for program planning and evaluation).

The most carefully randomized design can still incorporate the confounding of cross-sectional and even of superior longitudinal designs. The former fails to distinguish developmental-educational changes from cohort and individual differences. The latter can reduce this problem via within subjects analysis but fails to distinguish program effects from historical changes. Either model alone is pre-experimental and unsatisfactory in the eyes of many social science methodologists (Campbell & Stanley, 1963; Schaie, 1965). Evaluation models have been developed which provide alternatives to experimentation by sequencing research approaches in accordance with program evolution and situational needs. In over 50 ECE programs reviewed, however, Goodwin (1974) found very few which make any systematic use of an evaluation framework during all program phases. Perhaps this lack of complete evaluation framework is responsible for the lack of cumulation in ECE results (Rand Corporation, 1974).

Criteria used to assess program effectiveness are also subject to serious criticism. The inadequacy of school success as an evaluative index of the ultimate preschool goal, i.e., real life competence and mobility out of poverty, is now well documented in the United States and Britain (McClelland, 1976). The first and most common preschool outcome index, IQ, appears to be of minor importance in occupational success (Bowles & Gentis,

1972 in Bentley et al., 1973). It has been argued that, regardless of its quality, any single index is an inadequate criterion of program effectiveness (e.g., Messick & Barrows, 1972). The outcome from intervention programs is inevitably multivariate, even from curricula with a unitary focus such as language training. Ryan and Moffitt (1974) suggest that the apparent ineffectiveness of ECE may be due to researchers' reliance on too few dependent variables. The promised development of evaluative instruments focussing on a broad range of domains was abandoned as Follow Through research proceeded, and this failing is one of the strongest bases for critiques of the results obtained and conclusions drawn (House et al., 1978). Assuming adequate multiple indices can be used, it is suggested that comparing disadvantaged preschool children with similarly disadvantaged controls seriously misses the intent of preschool (Abelson & Zigler, 1974; D.K. Cohen, 1970). Rather they should be compared with populations not perceived as needing compensatory education.

The Educational Testing Service - Head Start Longitudinal Study, Young Children and Their First School Experiences (1976), is solving some of these evaluative problems. This Educational Testing Service study, begun in 1969 and still ongoing, exemplifies the movement towards evaluation which is more consistent with developmental rather than behavioural concepts of preschool. Principal aims of the study are (a) identification of the components of early education that are associated with the cognitive, personal and social development of disadvantaged children, and

(b) description and measurement of the environmental processes that moderate these associations (e.g., Bridgeman & Shipman, 1975; Educational Testing Service, 1968; Emmerich, 1971, 1973; Shipman, 1972, 1975). Careful attention is being given to the heterogeneous characteristics of the low income children involved in Head Start and to the structural stabilities and transformations in these characteristics during the course of development and educational experience. Information is being gathered longitudinally in six areas: the family, the teacher, the classroom, the school, the community, and the child. The array of over 30 assessment procedures used includes the comprehensive, modularized set of Circus measures developed from thorough study of existing ECE theory and measures. These procedures take account of extensivity as well as intensivity of preschool effects (Anderson, 1974-1975; Educational Testing Service, n.d.).

Several other studies taking a similarly extensive view of program effects have offered some positive observations overlooked by earlier, narrower evaluations. It has been noted, for instance, that during six years of Head Start-Follow Through programming (a) health professionals have been sensitized to the needs of the low SES population and (b) the roles of schools have changed in the direction of local staff employment and parent participation (Halasa, 1972). Hodges (1978) points out the worth of Follow Through programs in demonstrating a new model for tying educational research to practice, in developing inservice training expertise, and in confirming the possibility of positive change in program content, teacher behaviour, and school and home

atmosphere. Family and community diffusion effects have been documented in several ECE programs where younger brothers, sisters and neighbours of preschool participants showed developmental gains and entered school or preschool at higher levels than their older siblings and friends (Bronfenbrenner, 1975; Gray & Klaus, 1969).

During the late 1960's and early 1970's another set of evaluative results appeared which added the dimension of independent cross-program comparisons to longitudinal study. Five such evaluations were reviewed by Bereiter in 1972 (DiLorenzo, 1969; Erickson, McMillan, Bennet & Callahan, 1969; Karnes, Hodgins, Teska & Kirk, 1969; Miller & Dyer, 1970; Weikart, 1972). As results from longitudinal comparative evaluations accumulated, three incompatible conclusions were drawn regarding differential program effects. One block of evidence seemed to indicate greater short-term gains in Bereiter-Engelmann style direct-instruction programs, but the longer-term loss and slight reversal of this advantage in favour of enrichment style programs (Karnes, 1969; Weikart, 1971c). A second block of evidence suggested little or no overall difference among program models in conventional intelligence and achievement test terms (Gordon, 1972; Rusk, 1968). Recent studies are supporting the conclusion of similar effects by demonstrating that the IQ indicator of gain can be similarly affected by diverse programs emphasizing cognitive, social-emotional or other particular educational objectives (Jacobson, Berger, Bergman, Millham & Greeson, 1971; Zigler, Abelson & Seitz, 1976; Zigler & Butterfield, 1968). Motivation

to succeed in school, considered by Zigler and others (e.g., J. McV. Hunt, 1972) as basic to the early jump in IQ scores of preschool participants, can apparently be taught through either cognitively-oriented or socially-oriented interaction (Koep, 1973). Several studies related any program differences in IQ-achievement gains not to the content and structural variables used to define program types, but to more subtle factors such as teaching style (Beller, 1969) and staff morale (Weikart, 1972). Continuity of curricular style, for instance, seems particularly significant to maintenance and extension of the preschool gains made in several fundamentally different program variations (Horton and others in Miller & Dyer, 1975).

As evaluation techniques expanded to include a broader range of child, program and ecological variables, a third block of evidence emerged indicating qualitative rather than quantitative differences in the effects of various programs. Qualitatively differentiated evaluation began with single program studies which distinguished between traditionally-measured intellectual changes and other effects of preschool. In Halifax, for instance, Clark (1967; Ryan, 1972, Chapter 10) was able to show that over a two year period preschool children lost their early IQ-achievement gains but maintained teacher ratings in academic and social behaviour and in general conduct which were equal or superior to control children. The third conclusion regarding differential program outcome, i.e., that preschool models vary qualitatively in their effects, is being continually refined as evaluations of the eight Head Start and fourteen Follow Through

planned variations establish the actual operational differences among programs.

The first Stanford Research Institute report on Head Start planned variations (Bissell, 1971) supported the growing conviction (Pines, 1967) that structured programs, as opposed to open "whole child" approaches, were most effective in raising the cognitive ability and academic achievement scores of disadvantaged children. Subsequent Stanford Research Institute studies which extended to Follow Through planned variations suggested that this conclusion was at best only half the evaluation story. Certainly it seemed that academic programs characterized by systematic instruction and positive reinforcement schedules produced larger gains in reading, mathematics and in task persistence (Bissell, 1973; Stallings, 1975). On the other hand, discovery-enrichment programs characterized by open and flexible instruction showed greater increases in nonverbal problem solving, lower absenteeism, greater independence from adults and more cooperation among children during school work. Enrichment programs were also related to children's sense of responsibility for their school successes whereas instructional programs were related to a sense of responsibility for failures. In their study of 151 Follow Through classrooms Soar and Soar (1972) found a similar dichotomy. Gains in concrete skills reflected teacher control, drill and corrective feedback while gains in abstract skills and problem solving reflected teaching flexibility. Within just one program variation (Montessori) Banta (1972) too found classrooms functioning on an either/or basis, selectively fostering either

analytical thinking or creativity and innovative behaviours. These results seemed to present ECE policy makers with a value-laden dilemma of choice between structured and open programs (Almy, 1975; Katz, 1974 in Miller & Dyer, 1975).

Other threads of evidence, however, are accumulating to suggest that instructional and enrichment programs are not so clearly dichotomous in their developmental effects over longer time periods. Even the recently publicized conclusion from the extensive Follow Through studies in favour of "basic skills" (as opposed to "cognitive-conceptual" and "affective-cognitive" programs) is being rejected as misleading due to seriously flawed design, analytic procedure and interpretation (No Simple Answer, House et al., 1978). Several years ago, the Soars (1969) observed that greater academic gains during the summer vacations were related to unstructured and individualized teaching during the preceding school year. More recently, a strong association between achievement gains and positive attitudes towards school was seen after several years in enrichment-type but not in academic Follow Through programs (Bissell, 1973). One Follow Through center (Hamden-New Haven, Connecticut) reported that, after four years in the enrichment Bank Street program, grade 3 children were showing superiority over non-Follow Through disadvantaged children not only on intelligence and social-emotional measures but on academic achievement measures as well (Abelson & Zigler, 1974). At the kindergarten and grade 1 levels, readiness and achievement results had been consistent with the Stanford Research Institute national findings in not showing any academic advantage

for the Follow Through children. Similarly Miller and Dyer (1975) discovered that of the four compensatory programs they evaluated (Bereiter-Engelmann, DARCEE, Montessori, & Traditional), the one showing most positive IQ and achievement results at the grade 2 level (Montessori, in the case of boys only) had revealed no immediate impact in early cognitive-achievement tests such as the Preschool Inventory. Inventiveness and curiosity measures had shown gains during the pre-kindergarten year and these positive motivational-behavioural effects were still apparent in grade 2 along with the high IQ, reading and mathematics achievement scores.

One conclusion to be drawn from this outcome is that choice between convergent and divergent teaching methods is neither necessary nor desirable. These results pose the more complex questions of optima in levels of certain enrichment-instruction variables (Soar & Soar, 1972), in curricular balance between enrichment and direct instruction (Kohlberg, 1968), and in longitudinal ordering of enrichment and instruction in ECE programming. One emerging answer to the latter question of ordering, for instance, is that specific skill instruction can and perhaps should be delayed until later preschool or kindergarten. Initial enrichment programs can then allow time for the assimilation of basic adaptive ability in both social and non-social contexts, i.e., for learning how to learn (Beller, 1973; Wright, 1976).

A second conclusion from the overall Miller and Dyer results is that ECE effects in various domains become apparent sequentially over a considerable time period. Evidence such as (a) the effectiveness of working with parents and (b) the more

"middle class" home environment of children who maintained their early educational gains (Strodbeck, 1964) indicates to these researchers that full development of cognitive potential is related to motivations, attitudes and styles of behaviour developed during the preschool years. It is expected, therefore, that program effects will be reflected first in motivational and attitudinal measures, with benefits measured using cognitive and achievement tests becoming apparent over a longer time span. A recent report of the Educational Testing Service longitudinal study (Bridgeman & Shipman, 1975) concludes that "short-term impacts of preschool programs may be observed in different levels of achievement motivation and task orientation upon entering grade school; however, academic skills mediated by these behaviours may not be evident until later grades" (grade 3 in the cited study, pp. 81-82). Bronfenbrenner's (1975) in depth analysis of evaluations from family-centered preschool interventions suggests it may be the integration of social-emotional involvement with cognitively challenging tasks which underlies motivational-attitudinal changes observed as the earliest detected effects of successful enrichment-oriented preschool.

This interpretation leads to the third and perhaps most germane conclusion to be drawn from the Miller and Dyer and other contemporary ECE evaluations. It is the developmentalist concept of child-environment interaction processes which now offers the most promising focus for meaningful evaluation of immediate program effects. These interactive processes have been variously labelled "behavioural" (Hess, Kramer, Slaughter, Torney, Berry &

Hull, 1966), "noncognitive" (Miller & Dyer, 1975), "non-achievement" (Stallings, 1975), "horizontal" (Bussis & Chittenden, 1973), or "other" (E.D. Evans, 1974) developmental characteristics. Changes in these process characteristics are appearing to be the most enduring as well as the earliest detectable effects of compensatory ECE (Beller, 1972; Miller & Dyer, 1975; Weikart, 1971b; Wright, 1975b).

A recent Canadian study (Wright, 1976) uses the developmentalist framework of interaction processes to explain summer holiday gains made by low SES children following their first preschool year. Teacher ratings indicated gains during this preschool year in self-direction, mastery behaviour, self-management, curiosity and imagination. In neo-Piagetian terms (Case, 1975), it appears that these children enlarged their repertoire of task-related and effective executive schemes which enabled them to support their own continued development in a less stimulating environment. Similar developmental variables are described by more behaviourally-oriented educators as prerequisites to learning from the environment and functioning in a classroom (Gagne, 1967, 1968; Nedler, 1972; Resnick, 1963, 1967). These prerequisite capacities, normally learned in the course of development from infancy to school age, include paying attention, responding, following directions, freely and accurately exploring the environment, and functioning appropriately in relation to others (Hewett, 1967).

Developmentalists conceptualize these processes not simply as prerequisite learnings but as a higher order of behavioural

abstraction than any given act or skill itself (Fitzgibbon & Morris, 1975; McClelland, 1976). As a child develops, these underlying processes evolve and become focussed on particular ability areas and interests which can be more easily observed and differentiated in the concrete manner of traditional outcome testing. While changes in interactive processes themselves can still be evaluated in older persons, it is especially important to evaluate development at a basic process level with young children in which these processes are not yet "canalized" (D.H. Stott, Personal communication, January 28, 1977), i.e., not yet manifest in particular achievement domains. Factor analytic studies of children's autonomous behaviour, for instance, show that not until the early primary school period is this characteristic differentiated into the two dimensions of (a) social independence and dominance and (b) intellectual creativity and purposiveness (Baumrind, 1975). The recent Office of Economic Opportunity-commissioned Rand Corporation proposal (1974) for the next phase of Head Start evaluation asserts the value of basic process-level assessment, and advises a special substudy to pursue hypothesis generation and instrument development. The proposal uses the labels "metacognition" and "metalanguage" for such interactive processes, comprising strategies for using cognitive abilities, for seeking needed information, for selecting appropriate problem-solving approaches from a repertoire, for searching memory, and for communicating.

Questions concerning qualitative differences in program effects become even more complex as evaluators acknowledge the

significance of individual-program interactions. Early evaluations are accused of "washing out" any effects for particular children in particular programs by reporting results for "the average child" in "the average program" (Rand Corporation, 1974). Educators increasingly acknowledge the limited relevance to any single individual of findings based on large, random research samples (Kratochwill, 1976). In a reanalysis of all evaluation data up to 1973 for Head Start planned variations, Featherstone (1974) demonstrates that the interactions between program type and child characteristics, i.e., initial ability, SES, ethnicity, sex, and particularly response style and previous educational experience, account for more variance than do main effects for program type. Variations from site-to-site within Follow Through program models, and even more so within the three categories into which the major evaluators (Abt Associates) grouped the models, are comparable in magnitude to overall differences in effect among models (House et al., 1978). This result provides strong evidence for moderating variables, whether child-centered or contextual, which may override program model variables and make cross-model comparisons illusory.

The moderating effects of SES have been of specific concern to evaluators for some time. Some conclude that while preschool enrichment is appropriate for middle class children, structured instructional programs are most effective for lower class children who initially lack external and internal controls (Blank, 1973a; Camp, 1973; C.P. Deutsch, 1964; Elkind, 1973; Schoggen & Schoggen, 1971). Others interpret available information as indicating that

only high-achieving children fit differentially into open or traditional classrooms; low achievers experience similar difficulties in all educational settings (Winnett & Edwards, 1974). A recent comparison of Mexican and Black American preschool children revealed ethnic and cultural factors which were of greater significance to program effects than were curriculum factors such as "language development" and "autonomy" program variations. These two racial groups showed different patterns of gain on a verbal intelligence measure (Peabody Picture Vocabulary Test), and different relationships between this measure and process variables such as innovative and exploratory behaviours, curiosity, reflectivity and field independence (Kuzma & Stern, 1972).

Sex as a powerful moderator of program effects has been demonstrated in studies showing, for example, (a) that boys respond more to reinforcement inherent in preschool activities themselves, whereas girls respond more to a teacher's social reinforcement (Koep, 1973); and (b) that the social development of boys, but not of girls, is stimulated by enriched play equipment (Busse, Ree & Gertride, 1970). Such differences may be related to within-program differences in implementation for the two sexes, e.g., more instructional teacher contact with girls but more disciplinary contacts with boys, as well as to any differences in responses of the two sexes to identical experiences (Appleford & Ryan, in press; Biber, Miller & Dyer, 1972). In the ongoing Educational Testing Service study, sex was found to interact with urban-rural differences in moderating longitudinal relationships between achievement and certain interaction process variables.

Achievement motivation and self-esteem were significantly better early predictors of grade school achievement for rural than for urban children, especially with boys (Bridgeman & Shipman, 1975).

Individual differences in actual program "treatment" may be particularly significant in enrichment programs where, to a considerable degree, each child defines his own unique set of experiences (Karlson, 1972 in Stodolsky, 1972; Van Alstyne, 1932). The apparent superiority of highly-structured teacher-directed programs may be simply due to the greater homogeneity of their participants' experiences, and thus the greater effectiveness of objective-articulated evaluative measures in relating outcome to specific treatments (Stodolsky, 1972). Evaluators involved in developmental preschool programs must therefore, suggests Stodolsky, come to terms with diversity in education and in children. Bentley and others (1973) propose that diagnostic rather than actuarial evaluation is necessary to judge the educational progress of young children. They further advocate that this evaluation begin with baseline data gathered even before program planning. Rather than total reliance on a program model individual differences can then be taken into account with models used eclectically as resource pools to solve particular curricular problems. The educational implementation of such diagnostic evaluation corresponds to D.E. Hunt's proposal of individual study in terms of "accessibility characteristics" (1975; and see earlier work of Cronbach & Snow, 1969). Students can be described and then provided with appropriately adapted environmental structure.

Of the eight original Head Start variations being evaluated

by Stanford Research Institute, the one appearing most longitudinally successful in maintaining child gains in IQ-achievement as well as in noncognitive areas of development is the program most consciously oriented to individualization, to process-level sequential changes, and to the complementary benefits of enrichment and instructional preschool situations. The Learning to Learn program (Sprigle, 1969, 1970; Van de Reit & Resnick, 1976) begins with individual developmental diagnosis of each child and the initial planning of a personalized curriculum. Two educational environments are provided. One is a large heterogeneously-grouped setting richly equipped for child-directed activity. The other is a small homogeneously-grouped setting for teacher-directed instruction and social interaction. The proportion of small to large setting experience and the content of small group sessions is geared to individual needs. Instruction is sequenced according to Piagetian principles and focuses on integrated cognitive-affective learning strategies rather than on specific content. Primary goals for the preschool years involve stimulating the child to become active in her or his own learning through human interaction and creative exploration of the environment.

Some Contemporary Needs and Directions in Early Childhood Education and its Study

In 1972 Sigel looked back carefully and critically at ECE and its study during the preschool era.

From the research perspective, the expectations were unrealistic and even naive. We know our conceptions were inadequate, our measures crude,

our research designs fraught with error. In spite of sophisticated techniques and data analytic methods, the initial effort should be viewed as a pilot project. There were some exceptions, but even they are full of conceptual and methodological shortcomings (Stanley, 1972). (Sigel, 1972a, p. 102).

Perhaps today, building on accumulated learning during that pilot project period, early childhood educators are entering another era characterized by the exploration of alternative and potentially more adequate concepts and methods.

Conceptualization: Developmental Constructs

Conceptually, the sum of evidence reviewed suggests a contemporary phase of ECE programming and evaluation directed toward basic developmental processes which reintegrate the cognitive and affective domains. A need is implied for well-developed constructs and related measures which reflect the contextualized, individualized and sequential aspects of child-environment interactive processes. An extensive but fragmented literature exists concerning several relevant constructs and their manifestations in overt physical and social interactions: (a) creativity (e.g., Torrence, 1963), (b) curiosity (e.g., Maw, 1967), (c) locus of control (e.g., Rotter, 1966), (d) self-concept (e.g., Wylie, 1961), (e) cognitive style and particularly impulsivity-reflectivity (e.g., Kagan, 1965; Kagan, Moss & Sigel, 1963), and (f) achievement motivation and similar motivational constructs (e.g., Bruner, 1972b; J.M. Hunt, 1972; McClelland, 1976). Child developmentalists vary, overlap and conflict in their definition and use of these constructs and offer no consensus in relating them to classical perceptual-motor/cognitive/affective schemata (Walker, 1973). Self-

concept has been particularly cited by recent reviewers for the limitless collection of behaviours considered descriptive of various degrees and modes of self-concept (Coller, 1971 in Walker, 1973; DiLorenzo, 1969). Curiosity has been used in ECE study both as a summary variable comprised of novelty-seeking plus exploration plus problem-solving (Pick, 1974), and as a link in the circular relationships among exploration, self-image and concept formation (Minuchin, 1971). One solution to this confusion among constructs which vary in dimensions and extent has been to bypass explicit conceptualization in specifying process level goals for ECE. Education Development Center, for instance, provides teachers with an intuitive list of questions to ask themselves as they watch children from day to day, e.g., Does she make things? Does he like to think of variations in ways of doing something? Does she cooperate with other children in trying to solve a problem? (Duckworth, 1970).

Other preschool educators, however, are pursuing the conceptualization problem as a top priority for ECE. Robert White's The Concept of Competence (1959) underlies much of this work. The word competence is chosen by White to indicate the common property of all behaviours whereby an animal or child learns to interact effectively with its environment, e.g., visual exploration, manipulating the surroundings. Competence is characterized by "effectance motivation", seen most unambiguously in the playful and investigatory behaviours of young animals and children. The significance of these behaviours is lost if one arbitrarily breaks into their circular nature and declares cognition or active effort

alone as the real goal or source of satisfaction. The moderate but persistent motivation to make contact with the environment and to generate changing and interesting feedback leads to competent interactions in general, not simply in association with the immediate needs served by classical psychological drives.

Six groups of researchers stand out for their work in developing competence-related constructs of interactive developmental processes and applying these constructs to ECE.¹

1. D.H. Stott and colleagues (J.D. Sharp & D.H. Stott, 1976; D.H. Stott, 1961; D.H. Stott & Albin, 1975; D.H. Stott, Williams & Sharp, 1974) have tried to translate R. White's concept of competence into systematic, verifiable observations. Based on the work of Murphy ("coping", 1962), Rotter ("locus of control", 1966), and Wenar ("executive competence", 1964), as well as White, they propose and demonstrate a general factor of "effectiveness motivation". Their Effectiveness Motivation Scale for preschool children measures five broad categories of "nonorganic" motivation: recognition, discrimination, completion, control, and exploration effecting change. These five aspects characterize the general relationship of effectiveness which the child seeks to establish and maintain with her or his environment.

¹ These six conceptual contributions are generating a large number of observational process measures appropriate for use in early childhood. A review of these measures, as proposed in the dissertation prospectus, is in progress as a extensive separate project. This review includes the following information for each of approximately 35 measures: variables/subscales, original context, construction method, application and research, procedure and special features.

2. B.L. White and colleagues (B.L. White & Watts, 1973) observed several hundred children to generate lists of social and non-social abilities which distinguish (a) those who cope in superior fashion with anything they meet in daily life from (b) those free from gross pathology but generally of very low competence. Social aspects of "competence" involve, e.g., getting and maintaining adult attention, using adults as resources, leading and following peers, and showing pride in accomplishments. Non-social aspects include executive abilities in carrying out multistep activities and in effectively using resources, and attentional abilities such as dual focussing. Other than a general foundation in Piagetian developmental psychology, White adopts an empirical, anthropological stance in the development of his competency description and related observational scales.

3. Banta (1964, 1966b, 1967, 1970) has been pursuing the developmental construct of "autonomy" from a comprehensive eclectic foundation in such earlier constructs as creativity, cognitive style, and independence. Autonomy is defined by Shapiro and Biber (1972) as a corollary of competence. Banta's Tests for the Evaluation of ECE: The Cincinnati Autonomy Test Battery (CATB) comprises fourteen aspects of autonomy; of the "self-regulating behaviours that facilitate effective problem-solving" (Banta, 1970, p. 424). These include curiosity, exploratory behaviour, persistence, resistance to distraction, control of impulse, analytic perceptual processes and innovative behaviour. Several subscales of the CATB have been the basis of recently-published longitudinal findings which promise to redirect ECE

programming and evaluation efforts (Miller & Dyer, 1975; Stallings, 1975).

4. Murphy and colleagues (Moriarity, 1961; Murphy, 1962; Murphy & Moriarity, 1976) are refining their construct of "coping", encompassing "what a child does, actively or passively to handle, organize, accept or influence environmental (Coping I) or internal (Coping II) forces" (Moriarity, 1961, p. 13). These researchers focus both on the motivation to deal with new, unmastered aspects of the environment and on the strategies used in such coping, and are particularly concerned with the development of individual differences in coping during the preschool years. Their Comprehensive Coping Inventory lists coping devices derived from clinical, home and excursion observations as well as from psychological, psychiatric and pediatric examinations and histories. An instrument for the observation of defined coping behaviours in preschool children was presented in a recent dissertation (Silvernail, 1974). In her review of social-emotional developmental measures for young children, Walker (1973) values the coping construct for forcing one to look at the interface of all aspects of development, thereby avoiding the false cognitive - noncognitive dichotomization so prevalent in current psychological literature.

5. Factor analytic studies of classroom behaviour carried out by diverse groups of researchers over more than a decade are converging on a three-dimensional conceptual model for "classroom adaptation and competence" (Baumrind, 1975; Baumrind & Black, 1967; Behar & Stringfield, 1974; Becker & Krug, 1964;

Black, 1965; Emmerich, 1971, 1973; Kohn & Cohen, 1975; Kohn & Rosman, 1972a, 1972b; Mirante & Ryckman, 1974; Peterson, 1960; Rubenstein & Fisher, 1974; C. Schaefer, Baker & Zawel, 1975; E.S. Schaefer, 1971, 1975; E.S. Schaefer & Aaronson, 1976).

Generally a two dimensional circular model is derived for social-emotional behaviours, consisting of extraversion/introversion and considerateness/hostility factors. More recently the third dimension of task-orientation/distractibility, orthogonal to the circumplex, has been added to produce a spherical model. Significant correlations between academic achievement and the positive poles of these three factors (extraversion, considerateness, task-orientation) have been demonstrated (E.S. Schaefer, 1975).

To date the most detailed work with this structural-developmental model has been done by Emmerich (1973). His research has identified (a) behaviours characterizing various points within the model and (b) multiple interactions among moderators of the routes along which developmental change occurs in the preschool classroom context. While young children usually move from introversion toward extraversion, for example, Emmerich illustrates

a more complex curvilinear relationship between extraversion and task-orientation. Maximum polarization of task-versus-person orientation seems to occur at moderate levels of extraversion. Very recently E.S. Schaefer and colleagues (1976) extended the spherical construct into a four dimensional model which integrates many clinical psychology and personality studies into a comprehensive guide for sampling children's adaptation to the classroom. He uses a hierarchical presentation for ease in communicating the model's four dimensions, as follows:

6. Zigler's (1973) original suggestion of the concept of "social competence" as the success criterion for Head Start has been followed up in recent efforts to design a new national evaluation of early compensatory education (Anderson, 1974-1975; Rand Corporation, 1974; Washington, 1975). This new evaluation would focus broadly on a child's everyday effectiveness in dealing with her or his environment and would comprise health, intellectual ability and social-emotional development. An Educational Testing Service conference, convened by the United States Office of Child Development to define social competence in young children, produced 29 competency statements of which 22 are incorporated in the Circus assessment instruments (Anderson & Messick, 1975). Even after reducing the definition of social competence to "effectiveness in the role of pupil", Rand (1974) questioned the degree of goal consensus among different social groups and program constituents. The planners concluded that a national evaluation using a common set of outcome measures was an inappropriate strategy for assessing the social competence outcomes of Head Start. While Rand notes that a careful explication of social competence is badly needed, they see this as a major concept formation task which is difficult to conduct in the context of designing a national evaluation. The task is equally difficult in

the context of a small-scale study, but must be attempted in some degree if one is not to fall back entirely on earlier concepts now proven to be wholly inadequate for the evaluation of ECE. Despite their diversity, the six conceptual contributions reviewed do offer some increase in the precision and integration of process-level developmental constructs.

Methodology: Research Paradigms and Techniques

The most promising contemporary responses to the methodological shortcomings of the preschool era involve broadening the range of acceptable research paradigms. Seven basic characteristics are proposed by Patton (1975) to distinguish the classical experimental paradigm from an equally idealized naturalistic alternative: quantitative versus qualitative techniques, reliability versus validity emphasis, objectivity versus subjectivity, distance from versus closeness to the data, component versus holistic analysis, outcome versus process orientation, and generalization versus uniqueness. A continuum rather than a dichotomy between experimental and naturalistic paradigms is proposed by Willems (1969). Ideal experimental or analytical research involves a high degree of (a) investigator manipulation of antecedent conditions and (b) imposition of units on the data. Ideal naturalistic or descriptive research involves a low degree of manipulation and imposition of units. Most research falls somewhere between these two extremes.

In accepting paradigms with increasingly naturalistic characteristics, a new dimension is added to the options involved

in matching method to research problem (Guttentag, 1971). This dimension of paradigm choice has particular relevance to the study of such highly interdependent or "coupled" systems (Weaver in Siever, 1968) as those implicated in an educational program. The vast uncontrolled variability, the program fluidity, and the sheer multivariate complexity of ECE situations impose special and extreme limitations on established experimental research designs (Carter & Wharf, 1973; Hyman & Wright, 1967). All program evaluation evidence is very fundamentally situation bound; it is never "pure" measurement of treatment effects. It is becoming more widely recognized that much of our past knowledge concerning development and educational effects was based on child behaviours in artificial, unfamiliar and unnaturally short-lived treatment situations (Bronfenbrenner, 1976). The highly significant distortions produced by such experimental "untying" of variables, i.e., separating them from other variables with which they consistently covary in a child's normal life, is being demonstrated (Willems & Willems, 1965). Apparently the rush to a physical science model of research often caused educational and psychological investigators to bypass description and so to lose touch with the phenomena they sought to study (Sigel, 1972a; Tinbergen, 1951).

As ECE researchers become dissatisfied with the scientific monopoly of the experimental paradigm they are beginning to recognize the importance of ecological (Brunswik, 1955; R. Evans, 1975) or social context (Fairweather, 1967; Little & Ryan, 1978) validity. The quality and interpretation of educational evaluation

results depends as surely on these real-world sources of validity, as on scientific control and quantification. As well as ecological criteria, utilization criteria such as use mode of results, importance to participants and understandability to the evaluation audience (Cherns, 1969; Patton, 1975; Stake & Denny, 1969; Weiss, 1972c) are competing with the traditional sources of scientific validity and reliability (Bernstein & Freeman, 1974) as standards for evaluation excellence.

Rather than definitive "crucial" experiments, new paradigm alternatives favour an ongoing stream of research (Baumrind, 1975; Riecken, Boruch, Campbell, Caplan, Glennan, Pratt, Rees, & Williams, 1975). Longitudinal studies based on (a) explicit stepwise theoretical models of intended program effects and (b) replication of effects in similar or disparate situations are replacing single controlled experiments as means towards scientific conclusions concerning the nature and range of program outcome (Cronbach, 1963; Fitzgibbon & Morris, 1975; Kratochwill, 1976; Suchman, 1970; C. Weiss, 1972b). Sequential theory-based evaluation is particularly appropriate in situations such as ECE where ultimate goals are by nature long-term and often intangible (Kamii, 1973; Kohlberg & Mayer, 1972). Rather than frequently unrealistic non-treatment control data, baselines are sought out from archival, normative or comparison data. The internal validity of interpretations in context is a more immediate goal than eventual externally-valid generalizations (Thirton, 1975). As such research proceeds, data are added piecemeal to gradually reduce and refine plausible alternative hypotheses concerning program dynamics.

In line with this changing viewpoint, the United States government is being advised that a series of small-scale focussed studies, in place of further large-scale national evaluations, is the most effective expenditure of the funds available for the study of ECE (Rand, 1974). These studies would be of three types: methodological, naturalistic-substantial, and manipulative-experimental. Similar conclusions regarding localized ECE evaluation research were reached at a recent conference reviewing European efforts to evaluate ECE (Council of Europe, 1975). Historically ECE research has progressed from (a) small fragmented studies, to (b) attempts at experimentally-modelled, coordinated study on a vast scale, to (c) the present reinstatement of limited differentiated study in a new cumulative perspective (Rentfrow, 1975; and see Educational Research Service, 1976 for a review which comprehends both small projects in distinctive local settings and large mainstream ECE studies). Campbell (1969; 1972) views the present phase in terms of "means idealism"; of evaluation practices which are desirable as political processes, not justified merely in terms of scientific end-products. He asks whether society and scientists are methodologically ready for the "monster of measurement and experimentation" foreshadowed by the national evaluation studies. Prominent critiques of recently completed Follow Through evaluations (House et al., 1978) conclude that massive social-educational experiments with narrow outcome measures are bad investments and are inappropriate in a pluralistic society; in short that "evaluations like this are no longer needed"

(p. 158). For both social and scientific reasons, small-scale locally self-sufficient evaluation alternatives appear to have a continuing place before, beside and to a considerable degree in place of the recently dominant highly-centralized program evaluations.

The growing speciality of evaluation research has produced several long-term, comprehensive models of differentiated and/or staged evaluation adapted to various types and evolutionary phases of educational programs (e.g., Alkin, 1969; Carter & Wharf, 1973; Provus, 1969; Stufflebeam, Foley, Gephart, Guba, Hammond, Merriman & Provus, 1971). Representative of the differences among programs which determine suitable forms of evaluation are size, scope, duration, clarity and specificity of input, complexity and time span of goals, and innovativeness (Stufflebeam, 1970; Tripodi, Fellin & Epstein, 1971; C. Weiss, 1972b). For instance a descriptive evaluation model, rather than a judgemental model based on prematurely defined success criteria, is advocated in situations where clear statements of strategy, anticipated causal relationships and intended outcome are not yet possible. Such descriptive or exploratory evaluation is appropriate for basic conceptualization, and for the determination of effect via correlational and case study or other qualitative techniques. This evaluation approach incorporates many elements of the naturalistic research paradigm and is gaining respectability in situations where variables lack definition but are relevant to a clearly focussed theoretical and/or practical purpose (de Groot, 1969). Judgemental or demonstration evaluation, on the other hand, is appropriate for well-established

criterion-outcome comparisons, and for the determination of effectiveness via more controlled and quantitative techniques. This evaluation approach more closely follows the experimental research paradigm.

Successions of evaluative stages have been described which can be modified and repeated as programs move through stages of planning, exploratory implementation and improvement, and demonstration. The early application of criterion-outcome evaluation can short-circuit the natural states of program development and often fails to provide the information necessary for decision-making during the dynamic stages of program growth (Provus, 1969). Conceptualization of goals and program processes necessarily change during a longitudinal program of research and require progressive adjustments and reinterpretations (e.g., DiLorenzo, 1969). Particularly in early exploratory as opposed to later demonstration projects, many evaluators are advocating leaving both educational and evaluative procedures relatively free and even encouraging them to evolve over time (Carter & Wharf, 1973; Suchman, 1970). Increasingly researchers are accepting a social development role which implies a "how to make it better" rather than an exclusively "is it good" evaluative function. Evaluation is becoming a continuous process which serves program improvement as well as program assessment (Guba, 1969; Provus, 1969). Program and evaluation become combined into an action research unit in which program delivery and controlled investigation are balanced priorities.

In addition to service and research goals, Alkin and Kosecoff (1973), Cronbach (1963) and others suggest that program evaluation

should also serve individual-diagnostic and personnel-training functions. An overriding goal of these ECE researchers is to systematize informal observations and so to reinforce spontaneous independent self-evaluation by teachers and other program participants (Thirton, 1975). From the action research perspective, even small-scale studies of ECE can and should serve both society and science (Bronfenbrenner, 1975; Campbell, 1969, 1972). Fairweather asserts that action research, or "experimental social innovation" in his terms (1967), is based on a combination of scientific and humanitarian thought. The researcher involved in ECE evaluation must consider various "middleman" roles which allow her or him to act as a link between society and science. Role models from several fields have been put forward for consideration and trial. These include: (a) the general "professional model" involving an ongoing practice-research-practice cycle (Cherns, 1969); (b) the "medical" model involving a systems approach to side effects and context (Messick, 1969); (c) the "counsellor" or "attorney" model (Guba, 1969; Levine, 1974); (d) the "social engineer" model (Cherns, 1969); (e) the "scientific management" model (Suchman, 1969); and (f) the "boundary role" derived from systems theory (Schulberg & Baker, 1968).

Evaluation methodologists have contributed the terms "summative" and "formative" to distinguish the classical experimental model of program evaluation from action research. These terms are helpful in establishing evaluation-program relationships as long as the prevalence of mixed forms, e.g., hybrid or monitoring evaluation (Scriven, 1967), is recognized. Idealized models of summative

evaluation and formative evaluation may differ on at least five dimensions: (a) the administrative versus the programmatic level of the user or decision-maker; (b) the centralized, external versus the local, internal locus of study and result utilization; (c) the coarse-grain, whole-program versus the fine-grain, program-component focus of the study; (d) the terminal timing of a single assessment versus continual sequences of evaluative feedback; and (e) the comparative versus the non-comparative (often related to the quantitative versus the qualitative) nature of the study (Bloom, Hastings & Madaus, 1971; Carter & Wharf, 1973; Johnson, 1970; Scriven, 1967).

Evaluation anywhere along the summative-formative continuum (Klein, 1972) can be centered on several different aspects of an educational program. Stake (1967) presents a very useful methodological model which enables small-scale or sequential ECE studies to define their limits and place themselves appropriately in the larger context of related action research projects. Programs are separated into three components: antecedents, transactions, and outcomes. Each component can then be studied descriptively and judgementally. Descriptive evaluation comprises study of (a) logical contingencies among intended antecedents, intended transactions and intended outcomes, (b) empirical contingencies among observed antecedents, observed transactions and observed outcomes, and (c) congruencies between intents and observations for each of the three components. Judgemental evaluation comprises (a) relative comparisons between the six-celled descriptive matrices for two program alternatives and (b) absolute comparisons between the

descriptive matrix of a program and a matrix of standards of excellence.

Associated with the growing acceptance of the naturalistic paradigm in ECE action research are advances in research technique which make differentiated, individualized, sequential and contextualized evaluation (Yawkey & Wilvern, 1975) increasingly possible and practical. The primary function of this expanding methodology is to legitimize and deal effectively with observational and subjective data. A lengthening list of dissatisfactions with standardized, normative psychometric tests as measures of developmental effects underlies the adaptation and creation of more naturalistic measures by ECE researchers. Traits and domains being measured are poorly conceptualized in many traditional tests (Chazan, 1975), and items are chosen empirically on the basis of age discriminability rather than on the basis of developmental theory. The fallacies these tests incorporate, such as assumptions of a linear development of social play (Moncada, 1973) or a polarity of dependency and independence (Beller, 1955), are being exposed through observational study of children (Emmench, 1971, 1973). Confused results from tests based loosely on developmental constructs such as self-esteem and task orientation underline the need for more tightly defined and finely differentiated models of measurement (Bridge-man, 1976). The use of multifactored assessment in educational evaluation and placement, incorporating multiple procedures and data sources, has recently been legislated in the United States (see Engin, Leppaluoto & Petty, 1977, for a review of Public Law

94-142 and bibliography of appropriate evaluation models and measures).

It is being further demonstrated that test data depend very heavily on the methods used in obtaining them, and that vast differences occur in estimates of learning potential made under different conditions of performance (Bortner & Birch, 1970; Moffitt, 1974). Wide discrepancies in performance between one-to-one and group situations reveal the danger of measures which ignore setting contingencies (Sigel, 1975). For both cognitive and non-cognitive reasons, young children may perform poorly in test situations outside any rational context (Sigel, 1974). On the other hand, behaviours which children have learned to express in test situations may not be assimilated into their total cognitive or linguistic systems (Cazden, 1972). Maximal or minimal rather than typical performance can easily become the focus of test-centered evaluation. The novelty factor of psychometric testing is often counterproductive in discovering what a child really knows and can do. Many tests with norms based on the majority population clearly discriminate against the subcultures predominant in compensatory ECE programs (Rand Corporation, 1974; Vane, 1976). Achievement tests have been shown to predict the academic success of middle class children better than of low income children (Rand, 1974). Diagnostic tests tend to categorize and label children and depend on highly skilled outside professionals rather than on program personnel. IQ tests particularly are of limited program development usefulness, and when any standardized test is used for detailed planning the future psychometric quality

of the test is threatened. Tests developed according to good psychometric principles which maximize individual differences are poor tools for evaluating educational growth over time within the individual (Carver, 1974 in Rand 1974). Such tests are designed to be resistive rather than responsive to changes of the sort concerning ECE evaluators (Vane, 1976). Finally, formal tests of all types inevitably create often misleading and undesirable practice and reactivity effects on the part of teachers as well as children.

Qualitative observation is being justified by its responsiveness to many of these criticisms, i.e.; by its role in validating or "grounding" and effectively communicating quantitative data; in avoiding evaluation research which is narrowly limited to preconceived program objectives and developmental pathways; and in locating significant variables for further descriptive or quantitative study (Glaser & Strauss, 1970; Parlett & Hamilton, 1972; Ryan, 1976; Walker, 1973; B.L. White & Watts, 1973). Subjective data are being justified on similar bases, and by arguments concerning the validity of common human response tendencies and constructions of reality (Sigel, 1974, 1975; Steinnitz, 1975). Ethologists are demonstrating generally high agreement on the subjective meanings of larger behaviour sequences derived from independent, low-inference observational studies (Jones, 1972). "Introspective ethology" is advocated by Charlesworth as a highly effective way of learning about children (in Sigel, 1975). Methods such as Bayesian statistics are being used to accept and estimate human biases rather than attempting to circumvent them through

"blind" objective measurement (Edwards, Guttentag & Snapper, 1975; Plutchick, Platman, & Fieve, 1969; Riecken et al., 1975). Leaders in the field of evaluation research (Scriven, 1972; Stake, 1969; Messick, 1969) are demonstrating that subjective data from the multiple value viewpoints within and beyond an immediate program context are amenable to social science methodology. Joint evaluations by professionals and program clients epitomize this acknowledgement of diverse subjective perspectives in evaluation (Carter & Wharf, 1973; Mendelsohn, 1969). Legal advocacy procedures have been proposed as appropriate to the divergent interpretations of program evaluation results (Edwards et al., 1975; Levine, 1974).

ECE evaluators attempting to use observational and subjective information are looking to many other fields in the social sciences, humanities and philosophy for data-gathering and data-analytic techniques. Contributing fields include: (a) ethology and naturalistic observation (Fairweather, 1967; Jones, 1972; Willems & Raush, 1969); (b) ecological psychology (Barker, 1968; Gump, 1975); (c) social-psychological systems for direct observation and categorization of classroom behaviours (Coller, 1972); (d) anthropological and sociological forms of qualitative analysis (Hein, 1975; Filstead, 1970; Lofland, 1971); (e) art criticism (Eisner, 1974); (f) systems, dramaturgic, political, historical and biographical models of analysis (Carini, 1973; R. Weiss & Rein, 1972); (g) clinical empiricism and the exploratory clinical methods of Piaget (Eisner, 1974); (h) documentation (Engel, 1975); (i) phenomenology (Carini, 1975); and (j) the

philosophy of "verstehen" (Strike, 1972).

Child-centered naturalistic techniques recognized for particular potential in ECE research include exploratory clinical interaction (Kamii, 1971), unobtrusive measures (E.D. Evans, 1974; Webb, Campbell, Schwartz & Sechrest, 1971) and direct focus on children's productions in place of test results (Barth, 1969; Clegg in Eisner, 1974). These techniques extend the range of developmental processes and outcomes assessed and allow for evaluation of a hierarchy of objectives from abstract to concrete (Klein, 1972). Recently considerable specific attention has been given to teacher ratings which incorporate both observational and subjective information regarding child behaviour and development. The use of well-designed rating scales by experienced teachers is gaining favour as a valid complement to, and often a replacement for, conventional testing (E.D. Evans, 1974; Sigel, 1975). In a recent small-scale Canadian study, teachers rejected testing in favour of ratings on the basis of both information quality and practicality (Wright, 1976). If ECE is to be evaluated comprehensively at the local level, relatively simple naturalistic procedures such as rating must be added to testing, clinical assessment and even elaborate observational schemes.

Brandt (1972, 1973) points out that rating is the most prevalent source of social science data and can, with suitable precautions, be a powerful and reliable technique. He advocates (a) the use of illustrative operational manifestations, (b) the rating of each dimension separately across all subjects, (c) the presentation of forced discrimination choices, (d) the use of

ranking (with small samples) and Q-sort (with larger samples) to improve rating quality. Studies of the personal-social behaviour (Emmerich, 1973) and self-direction (Sansone, 1974) of preschool children have demonstrated the high degree of inter-rater reliability which is obtainable using well-developed instruments. The predictive value of behavioural ratings by teachers and parents, in comparison with predictions based on standardized testing or clinical diagnosis, has been shown in a large number of recent studies using both psychometric and naturalistic measures (e.g., Colligan, 1976; Flapan, Gunn & Neubauer, 1970; Harris & Wagoner, 1973; Hess et al., 1966; Huberty & Swan, 1974; Kohn & Rosman, 1972b). In one instance a relatively simple teacher evaluation of pre-kindergarten children in the areas of verbal, attentive and disruptive behaviours was as effective as a complex observation schedule, and compared favourably with extensive test batteries, in predicting end-of-grade 1 achievement (Forness, Guthrie & Hall, 1976). The investigators advocate teacher training in observational methods as an implication of their results. Perhaps the most extensively used rating measure in ECE has been the facesheet of the Stanford-Binet as adapted by various researchers (Banta, 1970; Hertzog, Birch, Thomas & Mendez, 1968; Hess et al., 1966; Miller & Dyer, 1975). The facesheet items have been analyzed into three factors -- achievement motivation, confidence in ability, and activity level -- which correlate significantly with a wide variety of other behavioural and cognitive measures (Hess et al., 1966).

Within the classroom environment, variables external to the child have been identified and evaluatively related to outcome through the use of observational dimensions such as defined by Stalling's Classroom Observation Instrument (1975), and by earlier work concerned with classroom interactions (e.g., Brandt, 1973; Ezrin, 1971; Medley, 1969; Messick & Barrows, 1972) and with teacher behaviour particularly (e.g., Nuthall, 1970; and see Hoge & Luce, n.d., for a critical review of research relating classroom behaviour to academic achievement). Outside the program situation, changes in families, institutions and communities have been used as criteria of effectiveness (Bronfenbrenner, 1975; Gilmer, Miller & Gray, 1970; Gray & Klaus, 1969; Halasa, 1972; Messick, 1969; and see Hammond, 1970 for a model of context evaluation in local school districts). The adoption of interview and survey techniques have made parent, teacher, child and general public aspirations and evaluative judgements scientifically acceptable and usable as additional standards for program reach and effectiveness (Baumrind, 1973; Campbell, 1972; Riecken et al., 1975). Early use was made of this class of methods by Bissell's (1971) questions to Head Start parents "What difference has Head Start made in your life this year?" and "What are the things you liked most about Head Start?". A comprehensive conceptualization for ECE observation is presented by Gordon and Jester (1973) involving interrelationships among three sets of variables: (a) presage (individual and context attributes), (b) process (program content and procedures) and (c) product (cognitive-affective and situational effects).

The vast quantity of data generated through naturalistic study has led to new models of multivariate analysis appropriate to the interpretation of lifetime developmental change (C.W. Harris, 1967; Nesselroade & Reese, 1973; Nunnally, 1972, 1975; Schaie, 1965; Wohlwill, 1973). The use of partial correlations and multiple regression are advised as the basis of large-scale quantitative analyses of the determinants of change (O'Connor, 1970). Multiple regression techniques are favoured because of greater flexibility in (a) incorporating both categorical and continuous variables, (b) allowing considerations of interactions among input variables, and (c) providing partitioning of variance beyond simple *t* tests for one effect at a time, through stepwise procedures (see e.g., Kerlinger & Pedhazur, 1973). Particular focus on sequential patterns of outcome is neatly demonstrated by Dyer and Miller's (1974) reanalysis of data from inner city Head Start children. Using a cross-lagged panel-correlation technique they showed that, contrary to common assumptions, early achievement scores predicted later IQ better than the reverse.

Perhaps more useful to small-scale longitudinal studies, where applicable statistics are unrealistically static and over-simplified, are demonstrations of narrative-descriptive, graphical and configurational data presentation methods (Baumrind, 1975; Jackson & Messick, 1967; Sigel, 1972a). One limitation of conventional statistical analysis, particularly as applicable to small sample situations, is its concern with only the central tendency or level of the experimental results. Clinical researchers are leading the exploration of analytical alternatives

in which level is only one of several inferential methods applied to diagnostic and intervention data. Clinical neuropsychology, for instance, has elaborated systematic techniques for objectivizing and teaching its decision rules; techniques which incorporate (a) level of performance, (b) specific deficits or signs of pathognomic significance, and (c) differential patterns of ability scores. (Reitan & Davison, 1974, pp. 31-32, 107, 360). These methods are directly relevant to developmental evaluation which seeks individualized, differential, sequential and often nonlinear interpretations of its findings.

Ipsative analyses which rely on profile or pattern display of observations and/or test scores for single individuals and groups have been used in recent evaluations to study multivariate educational effects over considerable time periods (e.g., Sigel, 1974). One approach to diversified idiographic evaluation is the intensive, contextualized behaviour analysis of each individual (Guralnick, 1973; Payne, Mercer, Payne & Davison, 1973). Building on existing observational instruments and anthropological techniques Angrist and Bourke (1974) have developed a flexible evaluative system which is able to focus on either situational and group contexts or on individual child and teacher behaviour. Their system, based on Kamii's (Kamii & Radin, 1970) interpretation of Piagetian developmental domains, is considered simple enough for continual use by teachers themselves to provide their own evaluative feedback. Intensive study of extreme cases to highlight significant variables, as a supplement to large N approaches, is demonstrated in a recent phase of the Educational Testing Service

longitudinal study by Shipman and colleagues (1976). Two groups of grade 3 children were involved in this comprehensive, differentiated analysis: (a) those individuals significantly above or below the average performance of similar children, and (b) those individuals significantly deviant from predictions based on pre-academic skills at four years-of-age. A recent state of the art presentation of statistical tests for $N = 1$ research describes this as an evolving field now entering established training programs and being advocated in psychological assessment, developmental research, learning disability study and general educational research (Kratochwill, 1977).

Some methodologists are suggesting that sophisticated statistical procedures are seldom necessary or desirable in the practice of evaluation research (Carver, 1970). Action researchers advocate a return to simple descriptive statistics, i.e., means and variances, percent and even actual raw figures, as the most useful form for reporting data to a diversified evaluation audience (Scriven, 1969) and for educational research generally (Cronbach & Snow, 1977). Results presented only in terms of significance levels bypass value judgements concerning the criteria for meaningful change and so for program "success". In presenting the argument that statistical significance tests are unnecessary, Kratochwill (1977) distinguished between experimental and clinical criteria. The statistically testable experimental criterion concerns the degree to which behaviour during intervention diverges from projected behaviour change in the absence of intervention. The clinical criterion, on the other hand, involves judgement as to

whether an intervention reaches levels of change significant to functioning in society.

A similar but more sweeping and highly convincing case against statistical significance testing, based on both mathematical and logical arguments, is made in a recent milestone article by Carver (1978). He considers that "statistical significance testing uses a corrupt form of the scientific method" (pp. 397-399) in placing the null hypothesis before the research hypothesis. Rather than decisive "accept" or "reject" interpretations, data should be examined for their degree of support for or disconfirmation of research hypotheses; for up or down adjustments in the probability of truth in a research hypothesis in the manner of Bayesian analysis. A search for accumulated ("replicated" in its broadest sense) evidence and for an effect-size context against which to compare observed magnitudes of difference and change should take priority over statistical significance. Particularly with small samples, Carver suggests that large effects should increase belief in a research hypothesis whether statistically significant or not (and conversely, that small statistically significant effects in large samples should often not be so interpreted), rather than discarding useful and important information. Such realistic contextual interpretation of evidence, rather than testing against a null hypothesis, should be the first step in analyzing research results. Evidence supportive of a research hypothesis can then be considered against alternative hypotheses, including the null hypothesis, as the second analytical step.

Technical reservations are being added to these logical reservations concerning conventional statistical analysis in action research. New questions are being raised as to the robustness of t and F tests of evaluation data necessarily based on small intact natural units and showing skewed distributions of most program characteristics (Hawkrige, 1970). The two assumptions involved in the use of t or F distributions, normal distribution in the population and homogeneity of variance, cannot be safely ignored with very small samples and particularly with such samples of unequal sizes (Hayes, 1963, p. 321). Unless the population distribution is known exactly, suggests Sigel (1953), only non-parametric statistics can be used with sample sizes as small as $N = 6$. A sensitivity to differences or changes other than central tendency provides a further rationale for non-parametric rather than parametric methods in much exploratory small sample research. Non-parametric techniques have become accepted in recent years as highly appropriate for small-scale research which violates the assumptions of parametric statistical methods (Fairweather, 1967). From a clinical action research perspective, Maxwell (1961) expounds the suitability of non-parametric associative analytical techniques for social science generally in the early stages of development, i.e., in searches for basic variables and general relationships. Given the present diversity of opinion regarding statistical analysis, the advice of Follow Through evaluation critics (Wisler, Burns & Iwamoto, 1978) to use as many methods of analysis as resources will permit is probably sound.

Purposes of this Study

This study is a response to the widely expressed need for more adequate conceptualization and more appropriate methodology in ECE and its study. The study is motivated by the need of many relatively isolated and unique rural communities in Canada (McPhee, 1975) for ongoing development of ECE programs on a small local scale. While findings here should bear directly on the summer preschool-kindergarten program being studied, they should contribute concepts and methodology relevant to other local ECE action research situations. The ultimate purpose of such contributions is to facilitate ECE programming and evaluation which is increasingly consistent with a developmental model of ECE, i.e., programming and evaluation which is differentiated, individualized, sequential and contextualized in even the most limited program settings.

The specific conceptual purpose of this study is to integrate the process-level characteristics of child-environment interaction into an enhanced developmental model for ECE research. Firstly, this requires multivariate, longitudinal study of the project population using a range of current psychometric and educational measures in order to

- (a) characterize these children as thoroughly as possible regarding their patterns of development-achievement strengths and weaknesses, and their conformity to, and/or divergence from mainstream North American culture,
- (b) describe the nature and sequence of development-achievement changes occurring during and following their program participation, and

(c) document the evidence for direct and indirect program effects on these development-achievement patterns and changes.

Secondly, well-differentiated observation of interaction processes is required to provide parallel information concerning interactive characteristics of these children vis-à-vis their physical and social environments. The course of gathering such information should further the definition and description of these interaction constructs which are emerging as highly significant in the developmental evaluation of preschool children. Thirdly, the interaction-process observations must be related both concurrently and predictively to the more traditionally studied development-achievement characteristics of the project population. Any relationships detected between the two sets of multivariate longitudinal data would refine the meaning of interaction variables and make their use in program-evaluation projects more practicable.

As reflected in much of the recent literature reviewed, small-scale action research falls more appropriately towards the naturalistic rather than the experimental end of the continuum of paradigm alternatives. The methodological purpose of this study is to implement many elements of a naturalistic research paradigm. This requires a study with the following methodological qualities:

- (a) responsiveness to the evolutionary stages of educational and evaluative procedures, with relatively little constraint or manipulation of the program context,
- (b) progressive involvement of local personnel and community in the ongoing self-evaluation of their own program, including

- the development of local measures, norms and expertise,
- (c) use of a wide range of measures to provide well-differentiated, multifactored patterns of child characteristics and their changes over time, including a few widely used indices providing for comparisons with other ECE settings;
 - (d) description of outcome in individualized and qualitative terms -- including interrelated patterns of characteristics, non-linear changes and clinical-diagnostic case study -- as well as analyses in group-averaged and quantitative terms,
 - (e) reliance on alternatives to unrealistic experimental control groups in the interpretation of results; such alternatives including the use of accumulated comparison data (norms, past local records, concurrent results in other settings) and the comparative combination of longitudinal and cross-sectional data from this study.

These methodological qualities should pervade all aspects of the study in its attempt to integrate interaction processes into a more adequate conceptual model for ECE. The very elementary state of instruments which focus on a basic process level, and so cross-cut the traditional cognitive-affective dichotomy in psychometrics, imposes an additional specific methodological requirement on the study (Rand, 1974). This is the continued development and quality assessment of a measure for child-environment interaction variables. The measure should be consistent with the overall naturalistic goals of the study, i.e., directed to behaviour in context rather than formal testing; relevant and practical for local use. Teacher rating seems best

suited to these goals. The six competency-based constructs reviewed, and the measures they have generated to date, appear to be the soundest available theoretical starting point for deriving useful rating items in the area of interaction processes.

This small local project is considered as part of an ongoing stream of research rather than as a single definitive study. As such it must place itself in relation to other categories and stages of ECE research. In the terms of leading evaluation researchers, the present study focuses on child-centered outcome to the relative exclusion of outcome in other aspects of the program context, and of program antecedents or transactions (Stake, 1967). It falls into two of the three categories proposed by Rand (1974) for small-scale focussed studies, naturalistic-substantial and methodological, but not into the third category of experimental-manipulative studies. For the most part the study is descriptive or exploratory (deGroot, 1969; Stake, 1967; Stufflebeam, 1970) rather than judgemental. The intention is to detect multivariate patterns and sequences of developmental change which may be useful in directing (a) evaluators planning studies in diverse program situations over various timespans and (b) teachers seeking ongoing feedback from their preschool children as a monitor, in part, of program effects. Rather than a demonstration of ideal ECE action research, this study seeks some refinement in the observation and measurement of child-centered outcome which may improve subsequent ECE evaluative research, whether descriptive or judgemental.

In rejecting a strictly experimental research model, this

study rejects a formal hypothesis-testing design. The study thus intends to strengthen or modify, rather than prove or disprove, those assumptions and expectations with which the project began. Such assumptions and expectations are based on past experience with the school and the children in this community, and on the accumulating results of published ECE evaluations. With respect to development-achievement characteristics of the study population, it was initially assumed that verbal abilities and auditorily-based skills were inferior to nonverbal abilities and visually-based skills (Blank, 1973a, 1973b; Cram, 1974; Deutsch, 1964). It was further expected, however, that with detailed study a relatively unique development-achievement profile would emerge here as in many other distinctive local ECE settings (e.g., Educational Testing Service, 1976; Featherstone, 1974).

With respect to rating of interaction processes, the following characteristics were most expected to show considerable developmental changes and relationships to conventional development-achievement measures: task-orientation, self-direction, flexibility, inventiveness, use of adults as resources, leading and cooperating with peers (Miller & Dyer, 1975; F.S. Shaefer, 1975; Stallings, 1976; White & Watts, 1973; Wright, 1975a, 1975b). It was assumed that some aspects of child-environment interaction would show curvilinear developmental patterns, e.g., aggression, social expressiveness. Concerning sequential relationships between interaction processes and development-achievement measures, it was expected that interaction ratings would show the most immediate change. Changes in specific knowledge and skills would follow, followed later still by changes in general ability measures

such as IQ (Bridgeman, 1976; Miller & Dyer, 1975; Shipman, 1976). Any immediate change in broad IQ-type developmental indices were expected to be transitory and due to practice or other extraneous effects.

(. . Evidence for program influences on the pattern and sequence of the developmental changes being studied is the only judgemental aspect of this ECE evaluation, and is secondary to descriptive evaluation goals. With each successive year of the program, greater and more lasting gains in development-achievement measures and more extensive changes in ability profiles and interaction characteristics were expected. These expectations were based on the increased direct and indirect input received by each successive cohort, i.e., longer pre-grade 1 educational experience, more program exposure and formal training of regular classroom teachers, greater continuity between the ECE program and regular school curriculum, more immediate parent involvement in the ECE program (Bronfenbrenner, 1975; Gotts, 1973; Horowitz & Paden, 1973; Miller & Dyer, 1975). It was also expected that a substantial community diffusion effect (Bronfenbrenner, 1975; Gray & Klaus, 1969) would be seen, i.e., that even before ECE experience, children of a given age would in some respects be ahead of children of the same age in previous years.

METHOD

Field Setting

The site of this summer preschool-kindergarten project was Grosse Ile in the Magdalen Islands, Quebec. The Magdalens are located in the Gulf of St. Lawrence, closer to the Maritime provinces than to mainland Quebec (Gaspé, 160 miles west). The English-speaking population of Grosse Ile and Old Harry at the northeast end of the island chain, together with the 250 anglophones of Entry Island 50 miles to the south, comprise about 10 percent of the 14,000 island inhabitants.

Two months of lobster-fishing (May 10 through July 10) is the primary economic activity of the anglophone communities. Work has recently begun on a mining-shipping enterprise in the Grosse Ile area which promises substantial change of uncertain benefit to the residents. These communities have, until recently, existed in relative isolation from both mainland culture and the majority francophone culture of the islands. In the fall of 1973 English-language television was introduced, although reception remains unreliable and of poor quality. Contact with the mainland is being increased by the provincial regionalization of medical and social services. In 1971 the Regional School Board of Gaspésie

assumed responsibility for English-language education on the Magdalen Islands. Grosse Ile-Old Harry and Entry Island thus became linked at many levels to an external network of programs, administrators and consultants. The Adult Education Department has become involved in the economic problems of the communities. The Guidance and Special Education Department, in cooperation with McGill University, is developing in-service teacher training and ongoing consultative programs.

Since becoming part of the regional school board, education for Grosse Ile and Old Harry residents has been extended to grades 10 and 11 (from an elementary level program), a special education program has been developed, and kindergarten was instituted for the first time in 1975. During these five years, the Grosse Ile School of approximately 120 students has seen three different outsiders as headteacher. This year for the first time both principal and head teacher are natives of Grosse Ile. Drop-out rate is high and attendance is poor, particularly among older students during fishing and hunting seasons.

During the 1974 school year, funds for a three summer preschool-kindergarten project were sought and obtained from the Quebec Department of Education. The project would consist of an educational program and its evaluation, evaluation being a condition of the funding. The appropriateness of a language-oriented preschool program as an educational response to the needs of this population was suggested by several parties (e.g., head teachers, Regional School Board of Gaspesia Director of Special Education, diagnostic consultants), and community parents expressed

interest. Previous study (1969-1974) of standardized achievement, aptitude and IQ test profiles of anglophone children in northern and eastern Quebec (Sept Iles-Schefferville, Côte Nord, Gaspésie-Magdalen Islands) revealed a pattern of specifically verbal skill deficiencies in relation to the standards of mainstream North American culture (Cram, 1975). An analysis of 490 Wechsler Intelligence Scale for Children protocols from 22 English-language schools in northeastern Quebec illustrated this pattern. A mean IQ of 81 was found, with Information, Comprehension and Vocabulary subtests primarily responsible for the 19 point decrement relative to the standardization mean. Scores were normally distributed about this mean, approximately one standard deviation below the standardization mean (Cram, 1974).

Grosse Île was originally considered a desirable site for educational intervention research relevant to the larger northeastern Quebec region due to (a) its extreme degree of the anglophone isolation presumed responsible for the educational deficits in the region generally, and (b) its anticipated stability of population and socioeconomic conditions. These assumptions of generality were abandoned early in the program when it became apparent that the figures of Grosse Île-Old Harry children were rapidly moving beyond prediction and that "verbal skill deficiencies" was a simplified and inaccurate guide to educational intervention with this population.

Chronology of Program and Evaluation
(see Figure 1)

The project was of 3-year duration, from summer 1974 through summer 1976. A 3-week preschool (4- to 5-year-olds) and a 6-week kindergarten (5- to 6-year-olds) took place during the first two summers. Following implementation of full-year kindergarten in September 1975, summer kindergarten was dropped in favour of including a younger group in the preschool program. During the third summer all 3- to 5-year-olds in Grosse Ile were involved in a preschool program of 6-weeks' duration.

Children were assessed on entering the program, again one year later (at the end of the school year or the beginning of the next summer's program), and similarly at the end of subsequent school years to date. Individual anecdotal reports and, in the last two project years, rating forms were also prepared by teachers following each summer program. The rating forms were completed early in the fall for older students not involved in summer programs. Additional more detailed follow-up assessments were carried out in 1975 with children involved in the first summer's programs in order to examine certain newly introduced measures of component abilities. The final assessment for all children involved in summer preschool-kindergarten was carried out in June, 1977. Follow-up by the author is intended for several more years, at least until these children complete elementary school.

In July, 1974, two weeks prior to the first project sessions, the project administrator and the local Red Cross nurse met with

2

Figure 1. Summer preschool-kindergarten, Grosse Ile, Magdalen Islands, Quebec: Chronology of program and evaluation, 1974-1977.

all the parents of the preschool-kindergarten children. The goals of the project, discussed previously with the parents on their initiative at the time of funding application, were reviewed. Assessment procedures were explained and permission obtained to carry them out. During the last week of the 6-week summer kindergarten, the parents of the kindergarten children were interviewed individually by the kindergarten teacher. They were asked to report any changes, good or bad, seen in their children since the beginning of "summer school". The teacher also provided some simple creative materials for each child to take home (crayons, scissors, blank books).

Similar parent meetings and interviews occurred during the following two summers. Preceding the final summer program, in which 3-year-olds were to be involved for the first time, the director-teacher visited each home. She reassured parents concerning the nature of "school" for such young children, and left each child a Polaroid snapshot to match one to be found over her or his personal storage cubicle at school on the first day. A follow-up visit was made in October, 1976 to observe the kindergarten children and their teachers in the regular school program.

Several major events during the course of the project should be noted. A summer Learning Center, a set of practical courses in special education under the auspices of McGill University, took place in Grosse Ile School coincident with the 1975 project sessions. All the regular teachers in Grosse Ile School participated in the Learning Center and were joined by colleagues from Entry Island and the Gaspé peninsula. This influx of outsiders

including several instructors, consultants, and other visitors from time to time, was a unique event in the community. Of particular importance was the involvement of Grosse Ile's grade 1 teacher who did her practicum work in the summer kindergarten. This provided full continuity of program between summer session and regular 1975-1976 school year.

Parent involvement was added to teacher involvement during the 1976 session. The preschool director-teacher conducted a McGill University course in ECE which involved the local teachers and several parents in practical work in the preschool as well as in lectures and discussions. Parents also met one day a week with the preschool teachers to discuss the program and child development generally. Several parents assisted during the last three weeks of the preschool program.

It is noteworthy that the present head teacher in Grosse Ile participated in the summer 1976 ECE course and preschool program. Previous head teachers had been primarily oriented toward secondary level education programs. The initiation of full-year kindergarten in 1975 is also significant for changes it will bring to the educational history of the children involved as well as to the rest of the elementary curriculum.

Participants

There were 5 groups of children involved in the project (see Figure 1). The average age at the beginning of grade 1 was 6 years, 1 month (6-1) for Groups 1, 2, and 3; 6-2 for Group 4;

and 6-1 for Group 5. The present study focuses on the first 4 groups only. The youngest children, Group 5, will be studied subsequently as longitudinal data become available.

Group 1 consisted of 7 children, 3 girls and 4 boys, with an age range of 13 months. They entered the program in summer 1974 with 6-week kindergarten.

Group 2 consisted of 11 children, 7 girls and 4 boys, with an age range of 11 months. They entered the program in summer 1974 with 3-week preschool followed by 6-week kindergarten in summer 1975.

Group 3 consisted of 12 children, 4 girls and 8 boys, with an age range of 7 months. They entered the program in summer 1975 with 3-week preschool, followed by the newly instituted full-year kindergarten in 1975-1976.

Group 4 consisted of 5 children, 2 girls and 3 boys, with an age range of 5 months. They entered the program in summer 1976 with 6-week preschool followed by full-year kindergarten in 1976-1977.

In summary, each group had more early educational experience than the preceding groups. Before entering grade 1, Group 1 experienced just 6 weeks of summer kindergarten. Group 2 experienced 3 weeks of summer preschool plus 6 weeks of summer kindergarten; Group 3 experienced 3 weeks of summer preschool plus a full year of half-day kindergarten; and Group 4 experienced 6 weeks of summer preschool plus full-year kindergarten.

Exceptions to the above included (a) the youngest Group 1 girl who repeated the grade 1 program rather than proceeding

through grade 2 in 1975-1976, and (b) the three youngest Group 2 girls who entered kindergarten rather than grade 1 in September 1975. The Group 1 girl was retained in grade 1 due to slow academic progress as well as age. The other three exceptions were due primarily to government age regulations (children must be 6-years-old by September 30 to enter grade 1). These special cases were only included in those aspects of the study noted below (see Results). For most data gathering and analysis steps, $N = 6$ for Group 1 and $N = 8$ for Group 2 due to the exclusion of these four children.

All the children were born on the Magdalen Islands and only a few have visited the mainland. Unstable home situations, complicated by alcoholism, existed for several of the children. Except for two mothers of three of the children, all of the parents were native Magdalen Islanders. Parents averaged a few years of elementary education. Most expressed positive feelings about the project and were concerned over their children's progress.

The project administrator, an Associate Professor of Education, organized, funded and staffed the project, and attended generally to community and Department of Education liaison. One teacher (A) taught preschool during the first and second summer sessions and kindergarten during the first summer session. She was an experienced elementary teacher, specialized in language and reading, with one year's sabbatical experience in British schools immediately preceding the project. A second teacher (B) joined the project and took responsibility for the kindergarten program during the second summer. She worked in a nearby mainland

town with individual children as a psychomotor specialist for a year prior to the project. A third teacher (C), director for a university laboratory preschool, joined teacher A during the third summer and directed both the preschool program and the related teacher-training course and parent meetings. The grade 1 teacher (D) in Grosse Ile School assisted on an informal, part-time basis with the 1974 summer sessions. She participated in the 1975 summer kindergarten as a practicum student in conjunction with Learning Center courses. All three teachers (D,E,F) now working in kindergarten and grades 1, 2, and 3 with the children who have attended summer preschool-kindergarten sessions participated in the 1976 summer preschool and ECE course.

A language, speech and hearing specialist (one in 1974, another in 1975-1976) carried out assessment and recommended language program content. She also arranged medical attention for the several cases of chronic ear-infection discovered, and began speech therapy for several children with articulation difficulties. A school psychologist planned the evaluation, carried the individual developmental assessments, introduced teachers to the group assessment procedures, and collected and analyzed evaluation outcome. She also conducted a follow-up McGill University education course which involved the kindergarten through grade 3 teachers in an advanced practicum in diagnosis and assessment related to the evaluation project.

Program Content

The Magdalen Islands preschool-kindergarten program offered a wide range of experiences and was eclectic, not extreme or unique, in relation to the program variations included in comparative research. Practical limitations to the study precluded formalized observation of the program. The following description is based on consultants' visits to the classroom and the teacher's (A) report following the first project summer. Generally, the program was a hybrid of traditional and British Infant School structure, with an initial emphasis on language experience. The classroom was set up in interest centers: reading corner, music corner, games center, art center, listening post and television. Daily activities for the first three weeks of the kindergarten program included story reading, nursery rhymes, skipping and ball games, self-directed art activities, visual-motor games and puzzles, and structured cutting, drawing and pasting activities. During the second three weeks of kindergarten, prewriting and prereading activities were introduced. Social interaction was encouraged through gradual reduction in group supplies such as paint pots. Speech was encouraged by discussion of the children's own activities and of Sesame Street videotapes, and by using play telephones and cassettes to record the children's own stories. Towards the end of the session creative drama and puppet plays were introduced.

The preschool program was similar to the first three weeks of the kindergarten program, with the appropriate selection of

younger-aged stories, puzzles and other material. During the second summer, somewhat greater structure and more directed instruction in content areas (vocabulary, general information, number concepts) and perceptual skills was provided while maintaining the basic Infant School combination of interest centers and group activities.

During the third summer a more directly Piagetian developmental basis was articulated by the program director. Her position as a preschool director and her summer role as instructor in ECE for the local teachers contributed to more explicit conceptualization of educational goals and a greater range of materials. Third summer preschool activities were divided somewhat differently into the five interest centers in reflection of the director's developmental orientation. The creative area included colour, form, discrimination, classification, and conservation activities (painting, building); the task area included problem-solving and symbol manipulation activities (toys, puzzles); the language arts area included listening and word-object activities (books, tape, documenting); the dramatic area included representational and role-taking activities (dress-up, house); and the block area included size, measurement, counting, balance and representational activities. Despite these changes, the essential program structure and content remained very similar across the three project summers.

Assessment Procedures

Development and Achievement Measures

Originally, an individually-oriented form of assessment was attempted with a view to individualizing the program for each child. Thus primarily psychometric individual measures were used during the first project summer. These were selected to indicate general developmental level and to reveal strength and weakness patterns in basic "readiness" abilities. Assessment priorities shifted during the second year to include measures which would also (a) be suitable for classroom-teacher use, (b) provide program planning guidelines for the group as a whole and (c) allow for evaluation of both individuals and the group program.

Numerous preschool assessment instruments were surveyed in the course of revising assessment procedures. Both teachers and consultants participated in defining the areas to be included in assessment, in selecting the second year measures, and in adding and deleting measures for the third year evaluation on the basis of their first and second years' experience. Selection reflected the original premise and content of the program to the extent of focussing carefully on several aspects of verbal skill development. Measures of receptive language (structural-functional and content message comprehension) and expressive language (usage), and of many perceptual-motor abilities considered to underlie acquisition of spoken and written language, were chosen to narrow in on the apparent verbal skill deficiency of the study population.

Additional measures were chosen to explore broadly the developmental changes occurring during the course of children's participation in the program. At a more immediately content-tied level, characteristic of achievement rather than diagnostic measurement, was assessment of factual and conceptual learning and, at later ages, of academic skills. At a more comprehensive level was the assessment of intelligence in both verbal and nonverbal terms.

All development-achievement measures used formally during the project are described in Appendix A.² For purposes of analysis and communication of results, the measures are grouped into four categories familiar to early educators: (a) intellectual development, (b) perceptual-motor abilities, (c) language acquisition, and (d) conceptual-factual learning and academic achievement. The first three categories have often been labelled "development" and the fourth labelled "achievement". The maturation versus learning distinction responsible for this division is, however, giving way to a multiplicity of more complex educational models which vary with such factors as child age and task familiarity.

The schedule of measures used is presented in Figure 1. While the selection of evaluative procedures changed as the program evolved and the children grew older, an attempt was made to use two measures consistently for purposes of external and internal comparisons. These were the Peabody Picture Vocabulary Test and

² Classroom teachers, the psychologist and the language specialist occasionally made informal use of several unlisted achievement and clinical-diagnostic measures during the course of the project. These measures will be introduced only where relevant to individual cases.

the Goodenough-Harris drawing test. Several other measures (e.g., Slingerland and Circus subtests) were used frequently enough to provide for segments of longitudinal (within-group) and cross-sectional study. At each assessment occasion preceding summer preschool-kindergarten sessions, the psychologist and language specialist prepared individual diagnostic reports incorporating the results of measures used with each child. Preliminary to this reporting, these consultants each met with every child and one of the parents (in most cases the mother) for a session of about 45 minutes. During these sessions (a) time was taken for mutual familiarization of child and parent with the project staff, (b) the individual assessment measures and screening for severe disabilities were carried out and (c) any parental concerns, project related or otherwise, were discussed.

When new measures were introduced in 1975, the psychologist demonstrated these to project teachers who assisted in their administration. At the end of each project session, teachers also prepared individual anecdotal reports for each child and recorded a summary of session end parent interviews. The full final assessments in June, 1977 were carried out more independently by the children's teachers in consultation with the psychologist. All standardized achievement testing (Stanford), including subtest selection, was carried out by regular classroom teachers under auspices of the school administration rather than of project personnel.

Interaction Process Ratings

In addition to the selection of development-achievement measures, a companion teacher-rating form was developed. This measure focussed on the integrated cognitive-affective processes of interaction with the social and physical environment. A preliminary trial of process-rating items derived from the work of B.L. White & Watts (1973), Banta (1970), Educational Testing Service (e.g., Anderson, 1974-1975) and others (see Introduction) plus input from teachers and project personnel, was carried out during 1975. The items concerned child interactions with adults and peers, exploration of surroundings, and attentional control and use of language in interactions with both people and things. A revised set of ratings was used in 1976 and 1977. The current form of these locally-developed rating items is included as Appendix B.

All teachers of the project children, either in SPS-SKG programs (teachers A, B, C) or during the regular school programs to date (teachers D, E, F), have participated in the interaction rating (see Figure 1). Before any teacher used the rating items for the first time, she discussed them thoroughly with the researcher, and with the other teachers who had used them previously and/or would be using them concurrently. Teachers were given the following statements as interpretation of any rating omissions: "I have not observed this child in situations where this behaviour might occur", or "I have not noticed relevant behaviour".

Teachers were also asked to comment on the items and report those most difficult to rate.

Data Analyses

Analytical procedures were selected on the basis of their appropriateness to the local scale and the naturalistic, exploratory purpose of this action research project. As argued and demonstrated recently by numerous methodologists (see Introduction), procedures particularly suited to such a study include visual-graphic presentations, simple descriptive and nonparametric statistics, the inclusion of ~~some~~ intensive case study, and the use of multiple criteria and inferential methods. This predominantly descriptive analytical approach respected the methodological limits of the study as well as its exploratory goal. For instance, small numbers of subjects with no truly comparable controls made it impossible to distinguish conclusively in this single study between idiographic and normative phenomena, or between maturational (or historic) and program effects. Analysis here was directed primarily towards establishing rather than testing hypotheses.

Development-Achievement Data

For each measure, results are presented in the following sequence of descriptive and analytical steps: (a) cross-sectional among groups within age-grade levels and, where relevant, within assessment dates, (b) longitudinal, across assessment occasions, within groups, (c) comparative, in relation to test norms and any available local-regional comparison data, and (d) by sex. Bar graphs of results for the major measures provide for visual

inspection of cross-sectional differences and longitudinal changes. Where group sizes permitted ($N > 5$), the statistical significance of group differences was tested using (a) the ~~Mann~~-Whitney U test for two groups or (b) the Kruskal-Wallis one-way analysis of variance by ranks (H) for three or four groups. The statistical significance of changes over time for groups of $N > 5$ was tested with (a) the Wilcoxon matched-pairs signed ranks test (T) for two assessment occasions or (b) the Friedman two-way analysis of variance by ranks (χ^2) for three or four occasions.

An interim multiple regression analysis was carried out with data from the 14 assessment measures used preceding and following the second summer kindergarten (SKG; Group 2, 1975, see Figure 1). This large set of measures was employed and analyzed in the middle year of the project to help evaluate and select development-achievement measures for future use. From input vectors for individuals and occasions, a squared multiple correlation coefficient (R^2) for each of these measures was derived indicating the proportion of variance attributable to assessment occasion (the multiple regression of development-achievement results on individuals and occasions minus the multiple regression of results on individuals alone, i.e., R^2 Development-achievement measure. Individual; Occasions - R^2 Development-achievement measure. Individuals). Supplementary multiple regressions were carried out using age and sex input vectors in place of vectors for individual children. This analysis included data from all 11 children involved in the second year summer kindergarten, i.e., eight Group 2 children plus the three younger children excluded

from most subsequent analyses. Despite these three additional children, the major R^2 results are presented here for their relevance to comparisons among areas of development and achievement rather than among groups or individuals. Graphic representation of relative change in the measures used pre- and post-second year summer kindergarten as determined by the multiple regression analysis is included as Appendix C.

It should be noted that the two measures of intellectual development, the Peabody Picture Vocabulary Test (PPVT) and the Goodenough-Harris Drawing Test (GH), were the only measures used in the study for which scores are age-corrected, i.e., scores are comparable across all ages for which the test is standardized (see Appendix A). A number of other measures provide normative scores for various ages, but the scores themselves are not adjusted for age. Deviations from the norm in PPVT and GH scores were considered and compared in terms of standard deviation units and quartiles. The two intelligence measures are similar in normative structure, i.e., $M = 100$, $SD = 15$ for each age level. This structure allowed for comparisons between concurrent scores on the PPVT verbal measure and the GH nonverbal measure, as well as comparisons involving either measure across age levels. Comparative longitudinal profiles of the two IQ measures, from pre-SPS to post-grade 3 age-grade levels, were constructed from the average of group mean scores at each level. An exception was made for the Group 2 pre-summer preschool (pre-SPS) PPVT score, which was not included in the overall pre-SPS average. This score differed by over 10 points from other scores at the same age-grade level, while

the range of all other group-averaged scores on both measures was within 5 points.

The following sets of local-regional IQ data were located for comparison with IQ scores during this project: (a) PPVT scores of grades 4 and 5 in Grosse Ile School in 1977, (b) PPVT scores of grades 1 and 2 in Grosse Ile School in 1973, (c) Lorge-Thorndike group IQ scores for grades 4 to 9 children in Grosse Ile School in 1971, and (d) WISC scores for 490 anglophone grade-school children in the Northeastern Quebec region in 1973-1974 (Cram, 1974). Students t was used to test the statistical significance of differences between mean PPVT scores of (a) all project children at last assessment (kindergarten to grade 3) and the two preceding age-grade cohorts (grades 4 and 5), and (b) project groups tested at grade 1 and 2 levels and the two preceding cohorts at these levels. These comparisons with external data represent the only use of parametric significance testing in this study.

To interpret the Slingerland Screening results it must be noted that the division of four areas of visually-based perceptual-motor abilities into the first five subtests differs with test level. Visual discrimination is assessed with two subtests (1 and 2) at the Prereading level but with one subtest (4) at A-B levels. Visual-motor skill is assessed with one subtest (4) at the Prereading level but two subtests (1 and 2) at A-B levels. Visual discrimination with memory and visual-motor skill with memory are assessed by subtests 3 and 5 respectively at all levels. Kendall's coefficient of concordance (W) was used to test the statistical significance of score rankings across the four areas

of visually-based perceptual-motor ability in all Slingerland administrations considered together, and in Prereading and A-B level assessments considered separately. At A-B levels the final three auditorily-based Slingerland subtests (6,7,8) were also administered, and the statistical significance of score rankings across all eight subtests was tested again using Kendall's W. The three auditory subtests plus three of the five visual subtests (3,4,5) combine to produce a screening breakoff score. The ratio of auditorily-based to visually-based errors comprising this score was calculated and compared across groups and age-grade levels. Formal separate assessment of the gross motor aspect of perceptual-motor abilities was dropped after two years as teachers came to feel more proficient in assessing this area observationally and to find quantitative assessment useful only in special individual cases.

Several of the perceptual-motor and language measures used during the second project summer were modified or rejected following 1 year's experience and the interim multiple correlation study. The language specialist reduced the Circus 7 measure of auditory discrimination by half on the basis of local dialect. Comparable shortened scores were obtained for the first summer's assessment by item recount. Different measures were used each year in a continuing search for valid and reliable assessments of auditory memory as related to language acquisition and academic achievement. Existing auditory memory measures appear to have serious problems of construct validity, i.e., differentially involving a number of as yet poorly defined factors. The Metropolitan Readiness

Listening subtest was dropped after 1 year because of the very short age-grade range between basal and ceiling scores and the spurious importance of visual-motor skills (picture interpretation and marking) apparent in this measure of language comprehension. Measures of both receptive and expressive language (Circus 8 and Circus 10-II) were modified for third summer assessment on the basis of local language usage. Circus 8 was prorated for 2 omitted items; Circus 10-II was reduced by about one third and the scoring simplified. An item recount of Circus 10-II results from the second summer provides a shortened score for comparison with third year scores.

The language specialist noted a marked correlation of the expressive language (Circus 10-II) and auditory discrimination (Circus 7) measures with age, especially among the youngest children involved in the third program summer. The Goodman and Kruskal γ (gamma) rank correlation statistic (association in ordered classes; see Hayes, 1963, p. 655) was used to quantify the relationship between age and scores in this data which contains several tied ranks.

The Circus 2 measure of quantitative concepts was selectively shortened (by about one quarter) for its first use with preschool level children (second summer). Items involving digit recognition were removed in response to reactivity effects noted during the first project summer, i.e., parents' overconcern with rote alphabet and number learning by their small children who would be starting "school" and answering "tests". This caution was felt unnecessary in the third summer. In all administrations of the Circus 11 measure

of general information, scores were prorated for two specifically U.S. items. Only two of the three subparts of Circus 5, Finding Letters and Numbers, were used -- capital letters and numbers, but not lower case letters -- again for brevity and de-emphasis of very early rote alphabet learning. After 1 year's trial, the Early Childhood Inventory identifications (body parts, shapes, colours) were dropped due to the low ceiling of these items for the project population.

The academic achievement scores for grade-school children (Slosson and Stanford tests) are reported in grade-equivalent terms. While most meaningful, especially to teachers, grade equivalent units cannot be considered equal in size at different levels (Stanley and Hopkins, 1972, pp. 366-371). For this reason, differences and changes in academic achievement were not subjected to tests of statistical significance. Comparisons are reported and judgements as to the significance, for instance, of half a grade equivalent at the grade 1 or the grade 3 level, are left to the reader. The following sets of local achievement data were located for comparison with the achievement to date of project children: (a) Stanford Achievement Tests for grades 2, 4 and 5 in 1975, (b) Canadian Tests of Basic Skills for grades 4 to 7 in 1971. As with the Slingerland subtests of perceptual-motor ability areas, comparisons among achievement areas was made wherever possible

Interaction Process Data

Since teachers were instructed to omit rating items concerning behaviour which they had no opportunity to observe or had

otherwise not noticed, item usability (Stanley & Hopkins, 1972) - answerability is here defined solely in terms of rating omissions. The proportion of items omitted was calculated for each rater across her uses of the 30 rating items as a teacher, a teaching assistant, or an observer. Where the proportion of a rater's total omissions was substantial ($>.10$), the significance of change in proportion of omissions from her first to her second use of the items was tested with χ^2 (association, with Yate's correction). The significance of variations in proportions of omissions from child to child within each group was tested using χ^2 (goodness of fit to flat distribution) wherever possible, i.e., where expected frequency of omissions-per-child was 1 or more. The significance of variations among items in proportion of omissions, across all uses to date of each item, was also tested with χ^2 (goodness of fit to flat distribution). The distribution of omissions-per-item ($\bar{M}, \text{SD}$) was used to rank the interaction rating items into five categories based on standardized (z) usability-answerability scores. Category divisions reflect a particular interest in identifying the few high-omission items.

Reliability, in the general sense of consistency or stability of measurement, was of concern to this study in relation to individual items only rather than to the full set of 30 items. Split-half or internal consistency indices of instrument reliability will only become relevant in subsequent work with groups of the items being studied singly here. Since short-term change was of great interest and was anticipated for at least certain rating items, rate-rater reliability was also irrelevant to the present

study. Concurrent agreement between raters was selected as the most appropriate index of stability for the interaction rating items.

A distinction must be made between interrater reliability and interrater agreement (Tinsley & Weiss, 1975). Interrater reliability represents the degree to which ratings of different judges are proportional, i.e., the similarity of order relationship among rated individuals regardless of any differences in the absolute ratings used to express this relationship. In effect, interrater reliability concerns the agreement among standardized (deviation) ratings of different judges, and is usually reported in terms of correlations or proportions from an analysis of variance. Interrater agreement represents the degree to which different judges assign ratings of the same absolute value to a subject, and is important where points on the rating scale are meaningfully defined. Statistical indices of interrater agreement, particularly those which quantify the seriousness of various disagreements, are relatively recent.

Different raters' orderings of children on the various rating items (reliability) was considered secondary to their perceptions of children's interactive levels in actual values as defined by the four scale points (agreement). Cicchetti's (1972) new measure of agreement between rank-ordered variables (C) was selected as the best available index to quantify the similarity of rater perceptions in this study. C represents the proportion of agreement between two raters, and always varies directly as a function of the closeness of individual pairs of ratings. The γ (gamma)

statistic of Goodman and Kruskal -- mathematically equivalent in the absence of ties to Kendall's τ (tau), the most commonly selected reliability statistic for ordinal variables (Cicchetti, 1972) -- was used here to quantify the similarity of rater orderings. Only systematic differences in order (negative γ values), however, were of specific interest.

$\underline{C}$ and γ were calculated for each of the four occasions where two raters work concurrently, and for the one occasion with three raters, i.e., three different rater combinations (Group 4 post-SPS; ratings by F on this occasion, when she was observer rather than teacher, were not included due to high proportion of omissions). Only one set of γ correlations was calculated for the ratings of Groups 1 and 2 combined following the third summer's programs since γ involves order relationships and a single pair of teachers in their shared classroom rated these two groups together. The four children who did not advance in grade level with their original groups were included in interrater agreement and reliability calculations since the focus was on items rather than on groups or individuals.

Proportions of agreement ($\underline{C}$) were averaged over all 30 items to obtain a mean level of interrater agreement for each of the four ~~age~~-grade levels rated by two or more teachers (see Figure 1). The absolute values of negative γ correlations at each level were totalled. $\underline{C}$ was averaged over all seven paired ratings for each item, and the items were ranked into five interrater agreement categories based on the $\underline{C}_{\text{average}}$ distribution ($\underline{M}, \underline{SD}$). As with omissions-per-item, categories reflect a particular interest in

isolating the few low agreement items.

Patterns of rating distribution for each item are presented descriptively only. (Tables for each item of the proportions of ratings at each scale point for each rating occasion are available from the author; three representative tables for the selected rating items subjected to correlational analysis [7,19,30 - see below] are included as Appendix E). Only two of the four post-SPS Group 4 raters were considered in describing rating distributions. Rater C was excluded due to her high degree of specialization in ECE and experience with ratings relative to the other teachers involved. Rater F was excluded due to her high proportion of rating omissions as an observer on this occasion.

Description of longitudinal change in rating patterns focuses on the location and concentration (proportion of total ratings) of the modal rating (1, 2, 3, or 4) at each occasion. Four aspects of mode change during the preschool and early grade school years were considered particularly noteworthy: (a) the number of rating scale steps in mode change, (b) the direction(s) of those changes, (c) the linear or higher order trend of such mode changes, and (d) any changes in proportion of ratings on mode (e.g., peaked vs. flat distributions). Items which showed the clearest patterns of longitudinal change in one or more of these aspects are described in some detail, both for their own sake and to exemplify the qualitative nature of changes to be sought in more complex longitudinal patterns. Quantitative trend analysis can only be meaningfully applied in larger scale studies or accumulated replications involving any rating items for which this descriptive study

suggests significant or particularly interesting patterns of change.

Cross-sectional comparisons among groups of the same age-grade level in different program years (and thus having different degrees of pre-grade 1 educational experience) could be made with particular attention to teacher D's ratings. D has taught the youngest age-grade levels of the regular school program in Grosse Ile School since before the SPS-SKG project, and has been associated with the summer program since its initiation -- longer than any other local teacher. D has used the interaction process ratings more frequently than other raters, including two post-SPS groups (3 and 4), two pre-SKG/post-kindergarten groups (3 and 4), and two post-SKG/pre-grade 1 groups (2 and 3). Three cross-sectional comparisons of interaction rating distributions were possible, therefore, based on the perceptions of a single individual. As with any age differences emerging from longitudinal change patterns, any apparent program effects emerging from cross-sectional comparisons must be considered merely as suggestive directions for further research.

Relationships between Interaction Ratings and Development-Achievement Measures

Analysis of the discriminative and predictive power of interaction rating items in relation to development and achievement indices was limited to the two largest groups in the study (2 and 3). Analysis concentrated on the extremes of achievement in these groups to date as advocated by Shipman and colleagues

(1976) in the Education Testing Service-Head Start Longitudinal Study (and see Anastasi, 1968, pp. 108-109 regarding contrasted groups validation method). As in that longitudinal project, three extreme groups were selected for study here -- high, low, and most changed children. Within Group 2 and Group 3, children were first ranked for developmental level at initial assessment (pre-SKG for Group 2, pre-SPS for Group 3), and for both development and achievement levels at most recent assessment (post-grade 2 for Group 2, post-grade 1 for Group 3). Developmental ranks were based on the average of ranks in PPVT verbal IQ measure, GH nonverbal IQ measure, and Slingerland perceptual-motor screening (pre-SPS, when only the visually-based Slingerland subtests were used, auditory memory measures used concurrently were also included in calculating average developmental ranks). Achievement ranks were derived from the average of ranks in Slosson Oral Reading, Stanford Arithmetic and the additional most content-oriented Stanford Achievement Test administered to the group (Science for Group 2, Paragraph Meaning for Group 3).

A point system for multiple criteria was used to quantify the degree to which each item discriminates the two highest achievers and the two lowest achievers in each group. The points were tallied separately for each achievement extreme, as follows: 2 points if the highest (or lowest) child was off the modal rating, 2 additional points if less than .25 of the group received the same rating as the high (or low) child, and 1 additional point if the high (or low) child was the only individual in the group to receive his or her off-mode rating; 1 point if the second-highest (or

-lowest) child was off the modal rating, 1 additional point if less than .25 of the group received the same modal rating as the second-highest (or -lowest) child, and 1 additional point (only tallied when the highest or lowest child was also off mode) if the second-highest (or -lowest) was the only individual in the group (or the only other besides highest or lowest child) to receive his or her off mode rating.

Based on the distribution of points-per-item ($\bar{M}$, SD of total points considering high and low extremes together), items were ranked into five categories of overall discrimination and five categories of prediction (based on point totals from the first rating occasion only -- pre-SKG for Group 2, pre-SPS for Group 3). The category divisions reflect a particular interest in identifying those items with the highest discriminative and predictive power. As well as overall discrimination, points were tallied separately for high and low achievers and the three most discriminative and predictive items for each extreme were identified and described. The point system was used also to identify those items which best discriminate and predict the two children most changed in rank, in either direction, from first developmental assessment to latest achievement assessment.

Three interaction rating items - selected for distinctive longitudinal patterns of change and a range of interactive behaviour, as well as for relatively high levels of usability-answerability, interrater agreement, discrimination and prediction - were included in a rank correlation analysis (Goodman and Kruskal's γ [gamma] was again used to deal with tied ranks). Correlational

analysis was limited to Group 2, the largest group in the project, the only group to participate in both a preschool and a kindergarten level summer program, and the group rated most frequently by a pair of raters (see Figure 1). The analysis is derived from Dyer and Miller's (1972) cross-lagged panel correlation of intelligence and achievement data for a group of Head Start children.

Predictive correlations from each past rating occasion (pre-SPS, post-SPS, post-SKG, pre-grade 1) to post-grade 1 achievement were calculated for the three items. Similar predictive correlations from rating to post-grade 1 developmental ranks were calculated and compared. For just the initial rating and assessment occasion (pre-SPS), predictive correlations from development and achievement measures to post-grade 1 achievement were calculated and compared for predictive power with the rating items. For the initial and most recent occasions, concurrent correlations among development and achievement ratings on the three selected items were also calculated for comparison with predictive correlations. Developmental ranks pre-SPS and post-grade 1 and achievement ranks post-grade 1 were those used for the discrimination and prediction analysis. Achievement ranks pre-SPS were based on three measures of specific skill and knowledge rather than of general ability: the Circus 2 measure of quantitative relationships, the Circus 10 measure of expressive language, and the Circus 11 measure of general information. It must be noted that small group size limited this analysis to correlation, rather than association using χ^2 or a related index of

predictive association (e.g., λ [Goodman and Kruskal's lambda]). Only monotonic relationships between ratings and development or achievement could be detected here, therefore, as opposed to possible future associative study with larger numbers which would detect relationships of any form between ratings and development-achievement measures.

Development-achievement and interaction rating data were combined with psychological reports, teacher anecdotal records and any other information available in longitudinal case studies of four children -- the highest and lowest achievers at last assessment in Groups 2 and 3. From the wide array of data produced by this largely descriptive research project, "an intensive study of extreme cases should serve to highlight significant factors (and/or combinations and sequences of factors) ... that covary with school performance" (Shipman et al., 1976, p. 3). Development-achievement profiles were constructed for each of the four children at each assessment occasion. The profiles are in standard score form (based on M and SD of group scores at each assessment date) to facilitate comparison among measures at each occasion and among strength-weakness patterns across occasions. Indication is made wherever one of these children was highest or lowest in the group, alone or tied with one other child. Individualized interaction ratings for these children, in relation to the concurrent rating distributions of their respective groups (as outlined above for discrimination and prediction analysis of ratings), were compared longitudinally with their development-achievement profiles.

RESULTS

Results of the study are presented as follows:

1. Observations of parents and teachers (informal, anecdotal).
2. Development-achievement measures --
 - (a) results for each measure described cross-sectionally, longitudinally, comparatively, and by sex;
 - (b) review of results for each of the four development-achievement categories (intellectual development, perceptual-motor abilities, language acquisition, and learning-achievement) according to the three major purposes of the project -- (i) to characterize the project population, (ii) to describe the nature and sequence of changes during the project, and (iii) to document evidence for project effects.
3. Interaction process ratings (usability, interrater agreement, and patterns of rating distribution).
4. Relationships between interaction ratings and development-achievement measures --
 - (a) results for individual rating items as they discriminate, predict and correlate with development-achievement levels;
 - (b) review of interaction ratings in relation to other project data according to the three major project purposes (as for development-achievement categories above).

Due to the descriptive exploratory nature of the study, two essentially parallel presentations of development-achievement measures, and of interactions ratings in relation to other project data, are included: (a) detailed description by result type or source, and (b) review of results according to project purposes. The purpose-oriented review may be more meaningful to most readers. The preliminary detailed descriptive presentation provides for alternative emphases and interpretations on the part of any interested party as advocated by naturalistic research methodologists and evaluators (see Introduction). The review presentations of the development-achievement categories and the interaction ratings incorporate findings from the four case studies and other relevant individualized data.³

Observations of Parents and Teachers

Following the first project summer, all parents responded positively to the question, "Have you seen any change in your child, good or bad, since the beginning of, 'summer school'?". Among the most frequently mentioned changes were (a) less destructive use of manipulative materials (e.g., colouring books), (b) more task orientation (e.g., longer periods of directed play), (c) greater interest in books, drawing and new words (spoken and written), (d) improved articulation and self-expression (e.g.,

³ Individualized data are included in Appendices D and E. Full separate case studies for four project children (highest and lowest achievers in the two largest groups) are available from the author. Further longitudinal study of individual project children is in progress.

child talks better, thinks first), and (e) better social control (e.g., positive response to "no").

The first summer project teacher (A) described Grosse Ile children, both preschool and kindergarten levels, as most dramatically lacking in those skills prerequisite to phonic reading instruction. Poor listening in the program setting was particularly noted, although attention to outside environmental sounds, e.g., fishing boat motors, seemed well developed. Teacher A also made the following unexpected observations concerning differences between SPS and SKG children during the first summer. Compared to the SKG group (1), she found the younger SPS (Group 2) children (a) more willing to try new tasks, more attentive to structured tasks, more cooperative with each other, less verbally and behaviourally aggressive, and (b) equal or superior in listening skills (e.g., following directions, rhyming), language use (including grammatical structures) and articulation. Emphasis on social and auditory skills was the end-of-summer recommendation by teacher A. She particularly suggested that settling into school routines at the beginning of the school year be deliberately and slowly paced, and that at least half the curriculum be devoted to language arts activities for the first year of school.

In contrast to both first year groups, teachers A and B reported that children entering preschool in the second summer had all used pencils and scissors, were willing and able to express themselves and spoke clearly except for one individual. They noted an increase in the amount of expressive language during the second summer session. Social skills were again found lacking,

however, particularly "sharing". All project staff noted that parents appeared more often in the school during the second summer, seemed more at ease there and in their conversations with teachers, stayed longer, and evidenced interest in materials and in their children's progress. Teachers A and B stated their conviction that increased parent awareness had caused the advance of this second SPS group (3) over both groups entering the program during the first summer.

Perhaps the most emphasized teacher observation during the second summer was the vast individual differences seen among these children from a small and superficially homogeneous community. Differences were noted most dramatically in the children's awareness of the outside world, negating an initial staff assumption that the children would all share about the same level of knowledge about what goes on around them.

Following the third project summer, parents generally expressed enthusiasm about the SPS-SKG program and reported improved play patterns, more resourcefulness and better listening on the part of their children. Teacher-director C noted numerous signs of serious parent and entire family involvement: a high level of participation in parent meetings and in program assistance; a greater interest in books and reading, and in creative-developmental toys; and some initiatives towards a local parent cooperative preschool.

Teacher C and her assistants reported increasing associative play among children during the third summer program, related to increases in (a) social interactions -- both aggressive and

cooperative, and in (b) verbal interactions -- more frequent use of language, but no noticable increases in vocabulary or language structures. Two particularly noted changes were the satisfaction and pride expressed in self-initiated achievements and the greater time spent attending to a task. After a fall follow-up visit, C reported that many children continued to show interest in new books, in their creations from the previous summer, and in their teacher.

Development-Achievement Measures

Intellectual Development

Preceding summer preschool, and so preceding any school experience on the part of any child, group average Peabody Picture Vocabulary Test (PPVT) scores (see Table 1 and Figure 2) differed significantly across the three project summers ($H(2) = 7.43$, $p < .05$). Almost all of this difference was in the superiority of Groups 3 and 4, which entered SPS during the second and third program summers, over Group 2, which entered in the first summer.

The three groups (1, 2 and 4) which were assessed at one year older preceding SKG again differed significantly ($H(2) = 15.56$, $p < .001$). Almost all of this difference was in the superiority of Group 4, which entered the program in its third summer, over Groups 1 and 2 which entered in the first summer. It is noteworthy that Group 4 had experienced 6 weeks of SPS plus a full school year of half-day kindergarten, while Group 1 had experienced no earlier schooling and Group 2 had experienced 3 weeks of SPS.

TABLE 1

PEABODY PICTURE VOCABULARY TEST: MEAN IQ SCORES AND STANDARD DEVIATIONS
BY AGE-GRADE LEVEL, GROUP AND SEX

	Entire Group			Girls			Boys		
	<u>M</u>	<u>SD</u>	<u>N</u>	<u>M</u>	<u>SD</u>	<u>N</u>	<u>M</u>	<u>SD</u>	<u>N</u>
Pre-SPS									
Gp 2(74)	78.9	9.9	8	84.2	4.3	4	73.5	11.5	4
Gp 3(75)	90.2	9.2	12	86.7	11.4	4	92.0	8.1	8
Gp 4(76)	90.0	13.1	5	94.5	20.5	2	87.0	9.8	3
Pre-SKG/Post-KG									
Gp 1(74)	92.0	12.9	6	101.0	2.8	2	87.5	14.0	4
Gp 2(75)	92.9	10.6	8	86.7	9.5	4	99.0	8.5	4
Gp 4(77)	96.6	7.9	5	99.0	5.7	2	95.0	10.0	3
Post-SKG/Pre-Grade 1									
Gp 1(74)	93.7	12.8	4	112.0	0.	1	87.7	5.0	3
Gp 2(75)	93.7	9.9	8	87.7	6.4	4	99.7	9.6	4
Post-Grade 1									
Cp 3(77)	93.7	8.0	12	92.2	7.0	4	94.4	8.9	8
Pre-Grade 2									
Gp 1(75)	95.5	3.7	6	97.5	3.5	2	94.5	3.9	4
Post-Grade 2									
Gp 2(77)	96.7	8.9	8	98.2	12.0	4	95.2	5.8	4
Post-Grade 3									
Gp 1(77)	94.0	9.9	4	100.0	7.1	2	88.0	9.9	2

TABLE 2

GOODENOUGH HARRIS DRAWING TEST: MEAN STANDARD SCORES AND STANDARD DEVIATIONS
BY AGE-GRADE LEVEL, GROUP AND SEX

	Entire Group			Girls			Boys		
	<u>M</u>	<u>SD</u>	<u>N</u>	<u>M</u>	<u>SD</u>	<u>N</u>	<u>M</u>	<u>SD</u>	<u>N</u>
Pre-SPS									
Gp 2(74)	82.4	8.7	8	80.5	5.8	4	84.2	11.6	4
Gp 3(75)	81.2	13.6	12	86.5	6.0	4	78.6	15.9	8
Gp 4(76)	86.6	3.9	5	84.5	2.1	2	88.0	4.6	3
Pre-SKG/Post-KG									
Gp 1(74)	85.8	11.9	5	75.0	7.1	2	93.0	7.9	3
Gp 2(75)	88.7	11.9	8	87.5	16.7	4	90.0	6.8	4
Gp 4(77)	87.6	4.3	5	85.5	6.4	2	89.0	3.0	3
Post-SKG/Pre-Grade 1									
Gp 2(75)	93.6	7.2	7	92.2	9.4	4	95.3	4.0	3
Post-Grade 1									
Gp 3(77)	100.0	19.1	11	104.0	16.1	4	97.7	21.6	7
Pre-Grade 2									
Gp 1(75)	94.7	14.7	6	86.0	1.4	2	99.0	16.9	4
Post-Grade 2									
Gp 2(77)	98.2	12.1	8	92.0	7.1	4	104.5	13.7	4
Post-Grade 3									
Gp 1(77)	92.2	5.6	4	96.0	5.7	2	88.5	2.1	2

Figure 2. Peabody Picture Vocabulary Test: Mean IQ scores by age-grade level and group, with comparison data from 1973 (grades 1&2) and 1977 (grades 4&5).

Figure 3. Goodenough Harris Drawing Test: Mean standard scores by age-grade level and group.

Immediately following SKG and preceding grade 1, Groups 1 and 2 were still comparable in mean PPVT score. Group 3 attained this same mean score at one year older following grade 1, and also following a year of half-day kindergarten not available to Groups 1 and 2. Subsequently, while groups were not assessed at the same grade levels, an emerging superiority of Group 2 over Group 1 was indicated. Following grade 2, Group 2 PPVT scores were higher than either the pre-grade 2 or post-grade 3 scores of Group 1.

Until the most recent assessment date, older children consistently attained higher PPVT IQ scores than the younger children assessed concurrently. This difference was substantial at the first assessment occasion (13 points, close to 1 SD) but small subsequently (< 5 points). At last assessment (summer 1977), however, the PPVT scores of the youngest children assessed, those completing kindergarten (Group 4), were above the scores of children 1 and 3 years older, completing grades 1 and 3 respectively, and equal to the scores of children 2 years older completing grade 2.

Changes in mean PPVT scores for Group 1 from first assessment to fourth assessment were small and statistically nonsignificant, both overall and over the period of largest gain. The small score changes from pre- to post-SKG were positive. Scores fell back slightly at post-grade 3 assessment, although remaining above the initial pre-SKG level. To date only Group 1, the oldest group and only group not to experience preschool as well as kindergarten, has shown any decline following earlier gains in PPVT scores.

Overall changes across the four PPVT assessments of Group 2 fell short of statistical significance, although the large jump between the first two assessments was significant when considered alone ($T = 5$, $p < .05$). This gain of approximately 1 SD (14 points) from pre-SPS to pre-SKG was by far the largest increment between any two assessment occasions for any of the four groups. Subsequently the PPVT scores of Group 2 children continued to increase, but increments were small and not statistically significant. The multiple correlation analysis of pre- and post-SKG scores showed the proportion of variance due to assessment occasions to be non-significant while variance due to individual children was significant ($R^2_{\text{PPVT.Individuals}} = .940$, $F(10,10) = 15.62$, $p < .05$).

The gain in the mean PPVT score of Group 3 between pre-SPS and post-grade 1 assessments was small (+3.5 points) but statistically significant ($T = 11.5$, $p < .05$). Between the two PPVT assessments of Group 4, pre-SPS and post-kindergarten, there was a somewhat greater gain of 6.6 points. This change represented substantial increases by 3 children (+19, +15, and +10 points) and lesser decreases by 2 children (-6 and -5 points; Group 4 was too small for the statistical test being used). These gains made by groups entering the program in its second and third years were notably smaller than the 1 SD gain made by Group 2 over a similar span (1 year longer for Group 3) but beginning in the program's first year.

Except for the youngest group on the first assessment occasion (Group 2 pre-SPS), all mean PPVT scores were within 10 (.67 SD) points of the standardization mean, i.e., in the 3rd quartile.

PPVT scores of grades 1 and 2 children are available for 1973, one year before the SPS-SKG project began (see Figure 2). Both these scores were more than 1 SD below the standardization mean, with grade 2 lower than grade 1 (grade 1 N = 8; M = 81.2, SD = 10.7; grade 2 N = 6; M = 76.3, SD = 6.1). The grade 1 mean score in 1973 was 12.5 points below pre-grade 1 scores of Group 1 in 1974 ($t(10) = 1.69$, $p < .10$), of Group 2 in 1975 ($t(14) = 2.27$, $p < .025$), and below post-grade 1 scores of Group 3 in 1977 ($t(18) = 3.57$, $p < .005$). The grade 2 mean PPVT score in 1973 was 19.2 points below the pre-grade 2 score in 1975 (Group 1; $t(19) = 5.49$, $p < .0005$) and 20.4 points below the post-grade 2 score in 1977 (Group 2; $t(12) = 5.10$, $p < .0005$). While in 1973 the mean PPVT score for grade 2 was lower than for grade 1, pre-grade 2 scores in 1975 (Group 1) and post-grade 2 scores in 1977 (Group 2) were slightly higher than concurrent grade 1 scores (Groups 2 and 3 respectively). At the most recent assessment occasion, children from kindergarten to grade 3—all of whom had participated in the SPS-SKG program—attained a PPVT average of 95.1 (group means ranged from 93.7 to 96.7). At the same date, grades 4 and 5 who had not experienced preschool or kindergarten (but whose teachers had participated in the training aspect of the project) were 10 points lower in mean PPVT score (N = 16; M = 85.7, SD = 9.2; $t(43) = 3.38$, $p < .005$).

Out of the 12 PPVT group assessments to date, girls were superior in 8, i.e., twice as often as boys. Considering all groups together, girls scored more highly at all age-grade levels except post-grade 1 (Group 3 only assessed at this level). Girls

were consistently superior in Groups 1 and 4, with the superiority being particularly marked in Group 1 (+13.5, +24, +3 and +12 points across the four assessment occasions). In Group 2, girls were superior at the initial (pre-SPS) and most recent (post-grade 2) assessments, but boys showed an even greater superiority (+12 points) at the two intervening assessments (pre-SKG, pre-grade 1). Boys were ahead of girls by a smaller margin (+5.3 and +2.4 points) at both Group 3 assessments (pre-SPS, post-grade 1).

Differences among mean Goodenough-Harris Drawing Test (GH) scores (see Table 2 and Figure 3) of groups entering SPS across the three project summers fell somewhat short of statistical significance ($H(2) = 4.87, p < .10$). The most substantial difference (+5.4 points) was between second and third program summers (Groups 3 and 4), rather than between first and second summers (Groups 2 and 3) as with the PPVT. GH scores of the three groups assessed at 1 year older preceding SKG (Groups 1, 2 and 4) differed even less than did the pre-SPS scores (3 points maximum difference). Group 4, which had received 6 weeks of SPS plus full year kindergarten, did not show the superiority in GH scores that it showed in PPVT scores relative to Groups 1 and 2.

Subsequently groups were not assessed at the same age-grade levels but, as with the PPVT, a growing superiority of Group 2 over Group 1 was indicated. Following grade 2, Group 2 GH scores were higher than either the pre-grade 2 or post-grade 3 scores of Group 1. Unlike PPVT results, GH scores gave a similar indication of superiority for Group 3 over both groups preceding it in the history of the program (Groups 1 and 2). Following grade 1, the mean GH score of Group 3 was above the Group 2 scores both

pre-grade 1 and post-grade 2 and above the Group 1 scores at both higher and lower age-grade levels. At this most recent assessment, Group 3 reached the highest mean score to date in either measure of intellectual development and attained the standardization mean (100). As with the PPVT, older children were ahead of younger children assessed concurrently at project initiation and 1 year later, but were not consistently so at the most recent assessment date. At this last assessment, the youngest group (4) was still lowest in GH score whereas it equalled or surpassed older groups in PPVT score. The second youngest group (3), however, was ahead of both older groups as well as the younger Group 4.

Changes in GH mean score for Group 1 from first assessment pre-SKG to third assessment post-grade 3 were statistically non-significant overall, as were differences in the four PPVT scores over the same period. The increment in GH scores from pre-SKG to pre-grade 2, however, was positive and greater than concurrent PPVT gains (+9 points). This mean GH gain represented increases by four children (+33, +15, +10 and +7 points) and a decrease by one child (-10 points). Between the pre-grade 2 and the recent post-grade 3 assessments, the GH scores of Group 1 fell back just slightly (-2.5 points). As with the PPVT, this is the only group which has as yet shown any decline after earlier gains in the GH measure of intellectual maturity.

Overall changes in mean GH score across the four assessment occasions for Group 2 were at a level of probability similar to concurrent longitudinal changes in PPVT scores ($\chi^2 = 5.96$, $p < .20$).

When just the first two assessment occasions were considered (pre-SPS and pre-SKG) the increment in GH scores, as in PPVT scores, reached statistical significance ($T = 15$, $p < .05$). GH gain during this period (+6.3 points) was, however, only half that of PPVT gain. The subsequent two increments (pre-SKG to post-SKG to post-grade 2) were comparable in size (+4.9 and 4.6 points) but were not statistically significant. Again paralleling PPVT results, the multiple correlation analysis of pre- and post-SKG scores showed a nonsignificant proportion of variance due to assessment occasions but a significant proportion of variance attributable to individual children ($R^2_{GH.Individuals} = .791$, $F(10,9) = 3.41$, $p < .05$).

Change in the mean GH score of Group 3 between pre-SPS and post-grade 1 assessments was statistically significant ($T = 5$, $p < .005$) as was the PPVT gain during this period. The GH increase of over 1 SD (18.8 points) was larger than gains made by other groups over any time period in either intellectual development measure. From pre-SPS to date (end of full-year kindergarten), Group 4 showed less gain in GH (+1 point) than in PPVT scores. The negligible increase in GH mean represented gains by two children (+4 and +9 points), losses by two children (-2 and -6 points) and no change by one child.

Figure 4 presents the longitudinal relationship between verbal (PPVT) and nonverbal (GH) measures of intellectual development. Excepting Group 2, PPVT was initially above GH (+6.7 points) and remained so one year later (+5.8 points). By pre-grade 1 assessment, GH score equalled and by post-grade 1 assessment it

Figure 4. Peabody Picture Vocabulary Test and Goodenough Harris Drawing Test: Mean scores of group(s) assessed at each age-grade level.

Note. Wherever group scores are averaged to produce an age-grade mean score, the range of group scores is 5 points or less.

^aThe group 2 pre-SPS PPVT score is over 10 points below the other groups assessed at this level and thus is indicated separately, not included in age-grade mean.

surpassed PPVT score (+6.3 points). To date, PPVT and GH scores of project children at subsequent age-grade levels have fallen within 2 points of each other. Again excepting the first Group 2 assessment, GH scores relative to PPVT scores showed a greater gain from a larger initial deficit to approximately the same level during the primary grades. Group 2, which entered SPS during the first program summer, was exceptional in having an initial PPVT mean slightly lower than GH mean, and in showing overall PPVT gains to date (+15.8 points) which approximate GH gains (+17.8 points). The PPVT gain was more immediate, however, such that from their second assessment pre-SKG the Group 2 children showed a pattern of verbal and nonverbal intelligence test scores similar to the other groups in the study.

Mean GH scores did not reach the 3rd quartile (within 10 points of the standardization mean) until pre-grade 1 assessment. At lower age-grade levels all groups fell into the 4th quartile (over 10 points below norm), whereas with the PPVT only the youngest group in the first program year (Group 2 pre-SPS) scored in the 4th quartile. At the last assessment date however (summer 1977) the average GH score of all children who had participated in SPS and SKG programs was 95.8 (group mean range from 92.2 to 100) -- almost identical to the comparable PPVT average of 95.1. This level in both verbal and nonverbal measures was approximately 1 SD above any combined verbal-nonverbal IQ data available from earlier years in this community or region, i.e., mean Lorge-Thorndike IQ of all grade 4 to 9 students in Grosse Ile School in 1971 = 79.4; mean WISC IQ of 490 children in northeastern Quebec

schools in 1973-1974 = 81.2 (Cram, 1974).

Out of the 11 GH group assessments to date, boys were superior to girls in 8, reversing the 2:1 ratio of PPVT mean scores in favour of girls. Until the most recent assessment, boys scored above girls at all age-grade levels except post-grade 1 (Group 3). In the oldest group assessment to date, Group 1 post-grade 3, boys were inferior to girls in GH (-7.5 points) as well as in PPVT (-12 points) scores. Boys' GH score was consistently ahead of girls' in Groups 2 and 4, and in Group 1 until the most recent assessment. The largest advantage for boys was seen in initial assessment of Group 1 which entered SKG in the first project summer (+18). In Group 3, girls were substantially ahead of boys at both GH assessments, pre-SPS (+7.9 points) and post-grade 1 (+6.3 points).

Review of intellectual development results according to project purposes

1. Characterization: The project population during the preschool years, excepting the youngest group in the first project summer, was characterized by verbal intelligence scores (PPVT) which well exceeded nonverbal scores (GH). During these early years, nonverbal scores fell in the 4th quartile while verbal scores (again excepting Group 2 pre-SPS) were in the 3rd quartile. In neither of the two large groups (2 and 3) did initial assessment with either intellectual development measure clearly predict the highest and lowest achievers in subsequent years (see Appendix D).

Most recently, average verbal intelligence scores for all project children (kindergarten to grade 3 levels) approximately equalled nonverbal intelligence scores. These averages were close to the standardization mean (within 5 points) and about one standard deviation above past regional and local norms for combined verbal and nonverbal IQ. A dramatically high nonverbal intelligence score relative to the group distinguished the high achievers in both large groups at last assessment (see Appendix D). Only in Group 3, however, where development-achievement ranks have remained stable since program entrance, did verbal intelligence score clearly distinguish highest and lowest achievers. Overall, boys exceeded girls in nonverbal scores while girls exceeded boys in verbal scores. Group 3 was exceptional in consistently reversing this pattern.

2. Patterns of change. Both verbal and nonverbal measures of intellectual development showed positive long-term longitudinal change in project children. Gains in nonverbal intelligence relative to gains in verbal scores were generally (a) slower and more continuous, and (b) greater overall from a lower initial score (except Group 2) to approximately the same level as verbal scores. Verbal intelligence gains occurred more immediately after program entrance than did nonverbal gains. The highest mean verbal IQ to date was attained by the youngest group at most recent assessment. The highest nonverbal score was attained by the second youngest group, reflecting the greater lag in GH relative to PPVT gains. Immediate short-term change during a summer program session (SKG, both with and without previous SPS) seemed

minimal in nonverbal and especially in verbal IQ relative to stable individual differences.

To date only the oldest group, and the only group with no preschool experience, has shown any decline following early gains in intellectual development. This post-grade 3 decline in both verbal and nonverbal scores was very slight. The younger groups with preschool experience have continued to date to increase intellectual development scores. At last assessment, highest and lowest achievers in Group 2 obtained similar verbal IQ scores, as they had initially, after an intermediate stage of extreme high and low scores (see Appendix D). No such convergence was seen in their nonverbal intelligence scores, nor in either intellectual development measure in Group 3 where highest and lowest individuals have continued to diverge.

3. Program effects. Program effect on verbal intelligence levels was evidenced by the recent statistically significant superiority of project children over the two Grosse Ile cohorts preceding them -- both as these cohorts were assessed (a) in early primary grades (1 and 2) in the year preceding the project and (b) in early elementary grades (4 and 5) at latest project assessment date. All three project groups with both preschool and kindergarten experience continued to gain in intelligence scores with each year following program participation, contrasting the decline with age level seen preceding the project. Longitudinal gains in the two largest groups where statistical tests were possible were significant; immediate short-term gains during SKG were not significant.

Greater program effect was apparent for each project year i.e., for each successive group of preschool entrants. This increment in effectiveness was most apparent between Groups 3 and 4 in verbal intelligence scores, and between Groups 2 and 3 in non-verbal scores (see above regarding greater lag in GH score gains). The most extensive intervention to date, 6 weeks SPS plus full year kindergarten for Group 4, showed the most immediate effect on verbal intelligence, bringing it to the highest level attained to date (pre-SKG). This effect was clear since Group 4 and the group preceding it (3) began the program with the same average verbal intelligence score.

The advantage of 3-week SPS in addition to SKG alone (Group 2 versus Group 1) was slow to emerge in either verbal or non-verbal intellectual development measure, only appearing after 3 years. The effect was particularly clear with the nonverbal measure where Group 1 scores never reached the level of subsequent cohorts. In the first project year, preschool alone, and any indirect program effects operating during that year, appeared to significantly increase intelligence scores, particularly on the verbal measure.

Significant indirect, or community diffusion, effects were seen with intellectual development scores. The effect on verbal intelligence scores was seen after just one program year, i.e., children entering SPS in the second summer (Group 3) exceeded those entering the first summer (Group 2). This diffusion effect between first and second years was greater than any longitudinal effect of the program itself in later cohorts. A smaller diffusion

effect became manifest more slowly in nonverbal intelligence scores, i.e., third summer preschool entrants were superior to second summer entrants, but no substantial difference was seen between second and first year preschool entrants.

Perceptual-Motor Abilities

Visually-based abilities

Total scores from the five visually-based Slingerland Pre-reading subtests (see Table 3 and Figure 5) differed among the groups assessed at both pre-SPS and pre-SKG age levels. Children entering SPS in the second summer of the project (Group 3) surpassed those who entered SPS during the first summer (Group 2), but the difference fell just short of statistical significance. At the pre-SKG level, differences among the three groups assessed (1, 2 and 4) were statistically significant ($H(2) = 9.40, p < .01$). Most of this difference was in the superiority of Group 4 over Group 2 ($U = 2.5, p < .005$). The difference between groups beginning SPS in first and second summers (Groups 1 and 2) exceeded the difference between SKG groups (2 and 3) in the first 2 summers. This was true despite the fact that the second summer SKG group (2) had experienced 3 weeks of SPS while the first summer SKG group (1) had no previous education experience.

Subsequently, groups were not assessed at the same age level, but Slingerland scores indicated the continuing similarity of Groups 1 and 2 in visually-based perceptual-motor skills, and the further similarity of Group 3 to both these groups (see Figure 6). At the most recent assessment occasion, Group 1 completing grade

TABLE 3

SLINGERLAND VISUAL SUBTESTS (PREREADING AND A-B LEVELS): MEAN TOTAL SCORES
AND STANDARD DEVIATIONS BY AGE-GRADE LEVEL, GROUP AND SEX

	Entire Group			Girls			Boys		
	<u>M</u>	<u>SD</u>	<u>N</u>	<u>M</u>	<u>SD</u>	<u>N</u>	<u>M</u>	<u>SD</u>	<u>N</u>
PREREADING									
Pre-SPS									
Gp 2(74)	6.2	0.9	8	6.2	1.0	4	6.2	1.0	4
Gp 3(75)	7.9	2.3	11	9.0	3.4	3	7.5	1.8	8
Pre-SKG/Post-KG									
Gp 1(74)	9.3	1.9	6	8.5	2.1	2	9.7	1.9	4
Gp 2(75)	10.0	3.7	8	8.0	2.7	4	12.0	3.7	4
Gp 4(77)	19.0	2.8	5	17.7	4.9	2	18.2	3.8	3
Post-SKG/Pre-Grade 1									
Gp 2(75)	18.0	4.4	8	17.5	5.7	4	18.5	3.4	4
A LEVEL									
Post-Grade 1									
Gp 3(77)	-10.4	9.2	11	-5.2	5.1	4	-13.4	10.0	7
Pre-Grade 2									
Gp 1(75)	-10.7	4.3	6	-11.0	5.7	2	-10.5	4.4	4
B LEVEL									
Post-Grade 2									
Gp 2(77)	-8.4	4.0	8	-8.5	4.2	4	-8.2	4.5	4
Post-Grade 3									
Gp 1(77)	-8.0	5.5	4	-8.0	8.5	2	-8.0	4.2	2

NOTE: Prereading scores are positive from low to high values (perfect score = 25);
A-B level scores are negative error totals from high to low values
(perfect score = 0).

Groups

1 2 3 4

134.

Figure 5. Slingerland Screening Visual Subtests (Prereading level): Mean total scores by age-grade level and group.

Figure 6. Slingerland Screening (A-B levels): Mean total total error scores for auditory subtests, visual subtests, and all eight subtests combined, with breakoff screening scores indicated, by age-grade level and group.

Note. A -- auditory errors; subtests 6,7,8.
 B -- visual errors combined with auditory errors to derive screening score; subtests 3,4,5.
 C -- additional visual errors to produce total score; subtests 1,2.

3 and Group 2 completing grade 2 were still essentially equal in visually-based error scores in the Slingerland B. At this same occasion, Group 3 children completing grade 1 made visual error scores on the Slingerland A comparable to those made two years earlier in the project by Group 1 children beginning grade 2. It was only before children began reading and conventional schooling that older groups were ahead of younger groups assessed on the same occasion. Pre-SKG children scored more highly than pre-SPS children on the Slingerland Prereading visual subtests during first (Groups 1 and 2) and second (Groups 2 and 3) project summers, whereas grade 2 and 3 children (Groups 1 and 2) scored equally on the Slingerland B visual subtests at last assessment.

Group 2 was assessed repeatedly with the Prereading level Slingerland (see Table 3 and Figure 5), and score changes from pre-SPS to pre-SKG to pre-grade 1 were highly significant ($\chi^2(2) = 15.06, p < .0001$). The greatest increase was over the short period pre- to post-SKG. The multiple correlation study over this period showed a statistically significant proportion of variance in Slingerland Prereading scores, about one-third, to be due to assessment occasion (R^2 Slingerland P. Individuals, Occasions = R^2 Slingerland P. Individuals = .348, $F(1,10) = 47.48, p < .01$). Variance due to individual children in results from these two occasions was not statistically significant. Group 2 exceeded the average level for school readiness (15) by early grade 1 following 3-week SPS and 6-week SKG. Group 4 attained and surpassed the readiness level at several months younger following 6-week SPS

and a year of half-day kindergarten.

Excepting Group 3, boys equalled or surpassed girls in the Slingerland visual subtests at all levels, with the largest advantage for boys occurring in Group 2 preceding SKG. In Group 3 girls were substantially ahead of boys at both assessments to date, pre-SKG and post-grade 1. This same reversal in Group 3 of overall sex difference patterns was seen in intellectual development measures (see above).

Ranking among the four areas of visually-based perceptual-motor skill agreed significantly across all Slingerland administrations (Slingerland Prereading $W(3) = .70$, $p < .01$; Slingerland A-B $W(3) = .52$, $p < .05$; Slingerland-all levels $W(3) = .62$, $p < .01$; see Table 4). At both Prereading and A-B age-grade levels, project children scored most highly in the subtest focussing on visual discrimination with memory (e.g., recall and matching). Scores in the area of visual-motor skill (e.g., copying figures or words) were somewhat lower. At A-B level where the visual-motor area is separated into two subtests, the one demanding more sequential processing (subtest 1) was lower. Lower again, third out of the four visually-based ability areas, was visual discrimination of form, orientation and sequencing details with little memory involvement (e.g., direct matching of figures and words). This area is separated into two subtests at the Prereading level and, as with the visual-motor subtests at A-B levels, the one demanding sequential processing ranked lower. Lowest ranking by a relatively large margin was the subtest requiring integration of visual discrimination, visual-motor skill and visual memory.

TABLE 4

SLINGERLAND SCREENING VISUAL SUBTESTS (PREREADING AND A-B LEVELS): RANKS OF THE FOUR VISUALLY-BASED SKILLS AREAS BY AGE-GRADE LEVEL, GROUP AND SEX

Skill area	Entire Group				Girls				Boys			
	D ^a	D-M ^b	VM ^c	VM-M ^d	D	D-M	VM	VM-M	D	D-M	VM	D
Prerdg sub-tests	1-2	3	4	5	1-2	3	4	5	1-2	3	4	5
Pre-SPS												
Gp 2(74)	3.5	1	2	3.5	3	1	3	3	3.5	1	2	3.5
Gp 3(75)	2.5	1	2.5	4	2	3	1	4	3	1	2	4
Pre-SKG/Post-KG												
Gp 1(74)	3.0	1	2	4	2	1	3.5	3.5	4	1	2	3
Gp 2(75)	3.5	1	2	3.5	3	1	2	4	3	2	1	4
Gp 4(77)	2.0	3	1	4	4	2	1	3	3	1.5	1.5	4
Post-SKG/Pre-Grade 1												
Gp 2(75)	2.5	2.5	1	4	4	2.5	1	2.5	2	3	1	4
Overall Rank	3.0	1	2	4	3	1	2	4	3	1.5	1.5	4
(Σ ranks)	(17)	(9.5)	(10.5)	(23.5)	(18)	(10.5)	(11.5)	(20)	(18.5)	(9.5)	(9.5)	(22.5)
A-B sub-tests	4	3	1-2	5	4	3	1-2	5	4	3	1-2	5
A LEVEL												
Post-Grade 1												
Gp 3(77)	2.5	1	2.5	4	1.5	1.5	3	4	2.5	1	2.5	4
Pre-Grade 2												
Gp 1(75)	1.0	3	2	4	2	3.5	1	3.5	1	2	3	4
B LEVEL												
Post-Grade 2												
Gp 2(77)	3.0	1	2	4	1	2.5	2.5	4	3.5	1	2	3.5
Post-Grade 3												
Gp 1(77)	4.0	1.5	1.5	3	4	1.5	1.5	3	4	1	2	3
Overall Rank	3.0	1	2	4	2	3	1	4	3	1	2	4
(Σ ranks)	(10.5)	(16.5)	(8)	(15)	(8.5)	(9)	(8)	(14.5)	(11)	(15)	(9.5)	(14.5)

^a (D) - Discrimination

^b (D-M) - Discrimination & Memory

^c (VM) - Visual-Motor

^d (VM-M) - Visual-Motor & Memory

A major difference between sexes in the overall pattern of scores across areas occurred at the grade-school level but not at the prereading level. In Slingerland A-B levels, boys were ahead of girls in visual discrimination skills involving memory, but girls were ahead of boys in visual discrimination without memory demands.

Auditorily-based abilities

At the most recent assessment data, children in the three groups completing grades 1, 2 and 3 (Groups 3, 2 and 1) obtained essentially equal total error scores on the auditorily-based subtests (6,7,8) of Slingerland levels A and B (see Table 5 and Figure 6). At all three age-grade levels, errors in these three auditory subtests accounted for approximately two-thirds (.64 in Group 1, .65 in Group 2 and .66 in Group 3) of the Slingerland score used in language disability screening. The three visual subtests used in disability screening (3,4,5) contributed only one-third of the error score for each group at last assessment. For Group 1 this recent 2:1 ratio of auditory to visual errors represented a sharp decrease in auditorily-based errors from the 3:1 ratio seen 2 years previously preceding grade 2. At the pre-grade 2 assessment, Group 1 had twice the auditory error total obtained by Group 3 at close to the same age-grade level (post-grade 1) but 2 years later in the history of the project. In contrast, the concurrent visual error scores of these two groups were equal (see above and Table 3).

TABLE 5

SLINGERLAND SCREENING (A-B LEVELS): MEAN TOTAL ERROR SCORES AND STANDARD DEVIATIONS BY AGE GRADE LEVEL, GROUP AND SEX FOR (a) ALL SUBTESTS (1-8), (b) AUDITORY SUBTESTS (6-8), and (c) SUBTESTS COMPRISING BREAKOFF SCREENING SCORE (3-8)

	Entire Group			Girls			Boys		
	<u>M</u>	<u>SD</u>	<u>N</u>	<u>M</u>	<u>SD</u>	<u>N</u>	<u>M</u>	<u>SD</u>	<u>N</u>
<u>SUBTESTS 1-8</u>									
A LEVEL									
Post-Grade 1									
Gp 3(77)	-21.4	15.7	11	-10.5	6.4	4	-27.7	16.3	7
Pre-Grade 2									
Gp 1(75)	-31.8	12.0	6	-30.0	12.7	2	-32.7	13.5	4
B LEVEL									
Post-Grade 2									
Gp 2(77)	-18.7	7.8	8	-18.5	7.1	4	-19.0	9.6	4
Post-Grade 3									
Gp 1(77)	-19.7	8.4	4	-14.5	9.2	2	-25.0	4.2	2
<u>SUBTESTS 3-8</u>									
A LEVEL									
Post-Grade 1									
Gp 3(77)	-16.6	12.4	11	-8.0	4.1	4	-21.6	13.1	7
Pre-Grade 2									
Gp 1(75)	-28.1	10.1	6	-28.0	11.3	2	-28.2	11.3	4
B LEVEL									
Post-Grade 2									
Gp 2(77)	-15.9	6.5	8	-15.2	4.9	4	-16.5	8.7	4
Post-Grade 3									
Gp 1(77)	-18.2	7.8	4	-13.0	7.1	2	-23.5	4.9	2
<u>SUBTESTS 6-8</u>									
A LEVEL									
Post-Grade 1									
Gp 3(77)	-10.9	7.7	12	-5.2	2.1	4	-13.7	8.0	8
Pre-Grade 2									
Gp 1(75)	-21.2	8.8	6	-19.0	7.1	2	-22.2	10.4	4
B LEVEL									
Post-Grade 2									
Gp 2(77)	-10.4	4.1	8	-10.0	3.4	4	-10.7	5.3	4
Post-Grade 3									
Gp 1(77)	-11.7	6.1	4	-6.5	0.7	2	-17.0	0	2

The Slingerland screening error total for Group 1 preceding grade 2 far exceeded the breakoff point for language disability (Figure 6). Following grade 3, the reduction in auditorily-based errors brought Group 1 within a few points of the breakoff point. Group 1 still had a higher screening score, however, due to auditory rather than visual errors, than the groups 1 and 2 years younger (Groups 3 and 2) which scored within a point of the breakoff score. These younger groups showed a constancy in auditorily-based errors similar to that in visually-based errors across the primary school age-grade levels (Slingerland A-B). To date, only Group 1 has substantially changed error scores during the grade-school years in either auditorily- or visually-based Slingerland subtests.

Boys consistently scored more poorly on auditorily-based subtests than did girls whereas, excepting Group 3, they equalled or exceeded girls in visually-based skills. Group 3 girls were superior to boys in both auditory and visual perceptual-motor skills as measured by the Slingerland tests.

The rank ordering of scores across all eight Slingerland A-B level subtests agrees significantly over the four administrations of these subtests to date ($W(7) = .83$, $p < .01$; see Table 6). Subtest 7 which focuses on discrimination and visual representation of language sounds (e.g., writing initial and ending sounds) was the lowest. Just slightly higher in rank total was subtest 6 involving auditory memory and visual representation (e.g., recall and writing of letters, digits, words). Subtest 8 involving visual recognition of auditory language stimuli (e.g., matching oral to written numbers, letters, words), and thus having more visual cues

TABLE 6

SLINGERLAND SCREENING (A B LEVELS): RANKS OF ALL SUBTESTS (1-8) BY
AGE-GRADE LEVEL, GROUP AND SEX

Entire Group								
A-B subtests	1	2	3	4	5	6	7	8
A LEVEL								
Post-Grade 1 Gp 3(77)	6	1	2	2	5	8	7	4
Pre-Grade 2 Gp 1(75)	4	1.5	3	1.5	5.5	5.5	7	7
B LEVEL								
Post-Grade 2 Gp 2(77)	4	1	2	3	6	7	8	5
Post-Grade 3 Gp 1(77)	4	1	2	7	5	6	8	3
Overall rank	5	1	2	2	6	7	8	4
(Σ ranks)	(18)	(4.5)	(9)	(14.5)	(21.5)	(28)	(31)	(17.5)

Girls								Boys							
1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
7	1	2.5	2.5	6	8	4	5	6	1	2	4	5	7	8	3
2.5	1	4.5	2.5	4.5	7	8	6	4	3	2	1	6	7	8	5
5.5	1	3	2	4	7	8	5.5	2.5	2.5	1	4.5	4.5	6.5	8	6.5
4	1.5	3	8	5	6	7	1.5	3	1.5	1.5	7	4	5.5	8	5.5
5	1	2	3	6	7	8	4	3	2	1	4	5	7	8	6
(19)	(4.5)	(13)	(15)	(19.5)	(28)	(27)	(18)	(15.5)	(8)	(6.5)	(16.5)	(19.5)	(26)	(32)	(20)

TABLE 7

CIRCUS 7 (HOW WORDS SOUND): MEAN SCORES AND STANDARD DEVIATIONS OF FULL (1975) AND REDUCED (1975, 1976) ITEM POOL BY AGE-GRADE LEVEL, GROUP AND SEX

	Entire Group			Girls			Boys		
	M	SD	N	M	SD	N	M	SD	N
Pre-SPS									
Gp 3 (75)	36.6	5.7	10	36.5	6.2	4	36.7	5.8	6
	(19.7)	(3.1)	(9)	(21.3)	(1.1)	(3)	(18.8)	(3.5)	(6)
Gp 4 (76)	(20.6)	(1.1)	(5)	(20.0)	(1.4)	(2)	(21.0)	(1.0)	(3)
Pre-SKG									
Gp 2 (75)	38.7	2.7	8	37.7	3.8	4	39.7	0.5	4
	(20.7)	(1.7)	(8)	(20.2)	(2.2)	(4)	(21.2)	(1.0)	(4)
Post-SKG/Pre-Grade 1									
Gp 2 (75)	40.6	2.9	8	40.7	2.9	4	40.5	3.4	4
	(20.7)	(1.6)	(8)	(21.0)	(1.4)	(4)	(20.5)	(1.9)	(4)

NOTE: Scores in parentheses pertain to reduced item pool (22 rather than 44 items).

than the other two auditory subtests, was highest ranked among the three auditorily-based subtests. For girls subtest 8 actually ranked higher than the two lowest visually-based subtests, but for boys all three auditory subtests ranked lower than the five visual subtests.

Circus 7, How Words Sound, was used in the second and third program summers to assess the language sound discrimination of pre-reading children (see Table 7 and Figure 7). Pre-SPS differences between Group 3 and Group 4 (based on shortened form of Circus 7) were in favour of Group 4 but were small and not statistically significant. Also small and nonsignificant was the difference between pre-SPS and pre-SKG (Groups 3 and 2) children assessed in the second program summer. It is noteworthy that the project language specialist found success with the shortened form of Circus 7 on preschool entrance to be closely related to age.

Circus 7 was the only measure where 3-year-olds (the 5th group in the overall project, beginning in the third summer) gave a good percentage of responses, and the rank correlation between age and Circus 7 scores among 3- and 4-year-olds entering the last program summer was highly significant ($\gamma = .49, p < .004$).

No change occurred in the scores of Group 2 on shortened Circus 7 over the SKG period. Changes over this time in full Circus 7 score were small but attained statistical significance ($T = 4, p < .05$). The proportion of full score variance over this period attributable to occasion, however, fell short of significance. Preceding SKG Group 2 was slightly below the kindergarten norm for children with preschool experience, but attained this norm by pre-

grade 1, i.e., over the SKG period. Preceding SPS, and so preceding any school experience, Group 3 children which entered the program in its second summer had already attained the norm for preschool children with no previous school experience. Sex differences in Circus 7 scores were small and inconsistent.

As described earlier, different measures of verbal auditory memory were used in each of the three project years (see Table 8). Preceding the first summer's program, SPS and SKG children (Groups 2 and 1) scored very closely in sentence repetition errors, with the slight difference in favour of the older group. The mean score was at the school readiness level established in this and other rural Quebec school regions. The same two groups differed substantially, however, in WISC digit span scores ($U = 5.5$, $p < .01$). Both SPS and SKG groups were below age norm, the pre-SKG group (1) by 9 months (group mean age = 6 years-3 months; group mean score age equivalent = 5 years-6 months). Substantial differences occurred again 1 year later between SPS and SKG groups (3 and 2) when the Engelmann rather than WISC digit repetition test was used ($U = 20$, $p < .025$). Over the SKG period (Group 2) no gain was made in Engelmann digit repetition score and the variance attributable to occasions was not significant. In the third project summer, this time using the Illinois Test of Psycholinguistic Abilities digit test for auditory sequential memory, the SPS group (4) scored below age norm as did SPS and SKG groups (2 and 1) in the first project summer. Group 4 showed the same 9-month lag in ITPA digit repetition pre-SPS which Group 1 showed in WISC digit span pre-SKG 2 years earlier (group mean age = 5 years-

TABLE 8

SENTENCE MEMORY, WISC DIGIT SPAN, ENGELMANN DIGIT REPETITION, AND ITPA
AUDITORY SEQUENTIAL MEMORY (VERBAL AUDITORY MEMORY MEASURES): . MEAN
SCORES AND STANDARD DEVIATIONS BY AGE-GRADE LEVEL, GROUP AND SEX

[illegible]

Engelmann Digit Rep (score)									ITPA Aud. Seq. Mem. (score/age equivalent*)								
Entire Gp			Girls			Boys			Entire Gp			Girls			Boys		
M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N
34.4	5.4	12	35.2	4.6	4	34	6.1	8									
									5/	6.4	5	14	8.5	2	15.7	6.7	3
									4-5*								
42.4	4.4	8	40.5	5.7	4	44.2	2.1	4									
41.0	5.8	8	39.5	7.7	4	42.5	3.7	4									

TABLE 9

TAPPING (NONVERBAL AUDITORY MEMORY): MEAN SCORES AND STANDARD DEVIATIONS BY AGE-GRADE LEVEL, GROUP AND SEX

	Entire Group			Girls			Boys		
	M	SD	N	M	SD	N	M	SD	N
Pre-SPS									
Gp 3(75)	5.1	3.2	12	5.0	2.9	4	5.1	3.6	8
Gp 4(76)	6.0	4.5	5	6.0	2.8	2	6.0	6.0	3
Pre-SKG									
Gp 2(75)	7.5	3.0	8	6.7	2.1	4	8.2	3.8	4
Post-SKG/Pre-Grade 1									
Gp 2(75)	10.1	2.3	8	10.5	3.0	4	9.8	1.7	4

2 months; group mean score age equivalent = 4 years-5 months). The only considerable sex difference in verbal auditory memory measures occurred in Group 1 pre-SKG during the first program summer and favoured girls in both sentence and digit memory. Sex differences were small in the other three groups, all of whom entered the program at preschool rather than kindergarten age.

Nonverbal auditory memory as assessed with the Tapping measure (see Table 9 and Figure 8) was higher pre-SPS for Group 4 than it had been 1 year earlier for Group 3, but this difference fell short of statistical significance. As with the Engelmann measure of verbal auditory memory, the Tapping scores of SPS and SKG children preceding the second summer program differed significantly (Groups 3 and 2; $U = 25$, $p < .05$). Unlike the Engelmann test, however, a significant gain occurred in Group 2 Tapping scores from pre- to post-SKG ($T = 3.5$, $p < .05$). Over this period a significant proportion of score variance can be attributed to assessment occasion ($R^2_{\text{Tapping.Individuals,Occasions}} - R^2_{\text{Tapping.Individuals}} = .152$, $F(1,10) = 5.94$, $p < .05$). The only substantial sex difference seen was the pre-SKG superiority of Group 2 boys which disappeared by the end of SKG.

Motor abilities

As described above, fine visual-motor skill was assessed with Slingerland subtests, particularly subtest 4 at Prereading level, subtests 1 and 2 at A-B levels and subtest 5 at all levels. It can be seen in the ranking of Slingerland visual subtests at all levels (see Table 4) that the integration of extensive fine-motor responses into visual discrimination and memory tasks

consistently and ~~substantially~~ increased their difficulty over similar tasks without extensive motor requirements (subtest 3 vs. subtest 5). Increased difficulty related to increased motor response was not apparent in visual-motor tasks not involving memory (subtests 1 and 2 vs. subtest 4), and in fact the relationship was often reversed as seen particularly in the oldest group assessed to date (Group 1 post-grade 3).

In the area of gross-motor development (see Table 10), children entering SKG in the first summer (Group 1) were all over a year below age norm on the Missouri measure. In comparison, only one of eight SPS children in the same year (Group 2) was a year or more below age norm. By the second summer, three-quarters of the Group 2 children, now beginning SKG, had moved to within a year of gross-motor age norms, and the group entering SPS (3) showed a similar proportion of norm attainment (four-fifths). Group 2 maintained the three-quarters proportion through SKG to pre-grade 1 assessment. Group 1 did not attain this level even by pre-grade 2 when more than half the children were still below gross-motor norms. Except for Group 2 pre-SPS, the number of boys equalled or exceeded the number of girls over a year below age norms.

Review of perceptual-motor abilities results according to project purposes

1. Characterization. By grade 1 entrance, project children were characterized by visual-motor and language sound discrimination abilities at the readiness level. A comparable level was attained

Figure 7. Circus 7 (How Words Sound): Mean scores of full (1975) and reduced (1975, 1976) item pool by age-grade level and group.

^aScores indicated pertain to reduced item pool (22 rather than 44).

Figure 8. Tapping (Nonverbal auditory memory): Mean scores by age-grade level and group.

Figure 9. Circus 8 (How Words Work): Mean scores by age-grade level and group.

^aProrated for 2 items omitted.

Figure 10. Circus 10-II (Say and Tell): Mean scores of full (1975) and reduced (1975, 1976) item pool, by age-grade level and group.

Note. Lower scale pertains to reduced item pool (26 items at one point each). Upper scale pertains to full C10-II item pool (38 items at two points each).

Age-grade level

TABLE 10

GROSS MOTOR OBSERVATIONS (MISSOURI): NUMBER OF CHILDREN OVER 1 YEAR
BELOW AGE NORM, BY AGE-GRADE LEVEL, GROUP AND SEX

	Entire Gp		Girls		Boys	
	>1 Yr.BN	N	>1 Yr.BN	N	>1 Yr.BN	N
Pre-SPS						
Gp 2(74)	1	8	1	4	0	4
Gp 3(75)	2	10	1	3	1	7
Pre-SKG						
Gp 1(74)	4	4	2	2	2	2
Gp 2(75)	2	8	0	4	2	4
Post-SKG/Pre-Grade 1						
Gp 2(75)	2	8	0	4	2	4
Pre-Grade 2						
Gp 1(75)	3	5	1	2	2	3

TABLE 11

CIRCUS 8 (HOW WORDS WORK): MEAN SCORES AND STANDARD DEVIATIONS BY AGE-GRADE
LEVEL, GROUP AND SEX

	Entire Group			Girls			Boys		
	M	SD	N	M	SD	N	M	SD	N
Pre-SPS									
Gp 3(75)	16.1	3.0	12	16.5	3.4	4	15.9	2.9	8
Gp 4(76)	15.6 ^a		5	15.2 ^a		2	15.9 ^a		3
Pre-SKG									
Gp 2(75)	16.9	2.4	8	16.0	2.6	4	17.7	2.2	4
Post-SKG/Pre-Grade 1									
Gp 2(75)	19.5	2.6	8	19.0	2.9	4	20.0	2.6	4

^aProrated for 2 items omitted.

earlier by later starting cohorts. Auditory memory for meaningful language (sentences) was also at the regional readiness norm preceding grade school entrance (assessed in the first summer only), with little difference between preschool and kindergarten age levels. Rote sequential memory for low-meaning verbal material (digits) in the early years, however, was considerably below age norms with substantial differences in score between preschool and kindergarten levels.

Verbal (sentences) and nonverbal (tapping) auditory memory measures were the most outstanding early identifiers of highest and lowest achievers in the two largest groups -- the verbal measure most differentiating achievement extremes in Group 2; the nonverbal measure in Group 3, and in Group 2 one year later for the high achiever only (see Appendix D). Poor overall visual-motor ability identified the lowest achievers at initial assessment. Highest achievers were characterized by early strength in sequential visual processing and then in complex visual-motor integration.

The gross-motor ability of three-fourths (and more in some groups assessed) of pre-grade 1 project children was within a year of age norms, with the notable exception of Group 1 where the proportion of below norm gross-motor scores was much greater.

In the primary grades, combined auditorily-based and visually-based perceptual-motor error scores fell close to the breakoff point for language disability, again excepting Group 1 which had error totals well beyond this point until grade 3 assessment. Primary children showed distinctly greater difficulty with auditorily

based than with visually-based skills, regardless of group, program year or age. Language sound discrimination (with visual-motor response) was the poorest auditory skill area assessed; auditory memory (with visual-motor response) was just somewhat less difficult; visual recognition (matching) of auditory language stimuli was the strongest auditory skill area assessed. The largest perceptual-motor advantages of the highest achievers in the two large groups were seen in auditorily-based rather than in visually-based skills.

Visual-motor skills, all stronger than auditory skills, became more difficult with increasing demands for fine detail discrimination, sequential processing and especially for extensive integration of discrimination, memory and motor abilities (Slingerland 5). Increasing motor involvement appeared to increase the difficulty of visual-motor tasks requiring memory but to decrease the difficulty of such tasks with low memory requirements. This pattern of relative difficulty was apparent at all ages, with complex integrative visual-motor skill being beyond the ability of even the highest-ranked (developmental measures) preschool child (all receive basal Slingerland 5 score). While relatively simple visual-motor and visual memory skills did not sharply differentiate the highest and lowest achievers, the more complex sequential and integrative skills (Slingerland 2 and 5) did so on several occasions.

Perceptual-motor advantages of the two high achievers, whether auditorily- or visually-based, were less dramatic vis-à-vis their respective groups than were advantages seen in intelligence and achievement measures. Sex differences were inconsistent in

auditory skill areas. Boys exceeded girls in visual-motor ability while girls exceeded boys in gross motor ability. In Group 3 the girls exceeded the boys in both areas. Girls exceeded boys overall in perceptual-motor abilities during the primary grades. Boys continued to equal or exceed girls in visual-motor ability alone, however, as they did before entering grade school. Girls excelled in auditorily-based skills considered alone. Group 3 was again exceptional in that girls were superior in both auditory and visual perceptual-motor skills. Among specific visual-motor subskills, girls excelled in visual discrimination involving memory while boys excelled in fine visual discrimination not requiring memory.

2. Patterns of change. For the one group which entered the program at kindergarten rather than preschool level, decline previous to school entrance was apparent in two perceptual-motor areas: gross motor and auditorily-based skills. Change in visual-motor skills during preprimary years (Group 2) was positive and significant, but no further improvement was apparent during the primary grades. Preprimary change in auditorily-based perceptual-motor skills was less marked than in visual skills. Nonverbal auditory memory showed significant short-term improvement in the early years, but verbal auditory memory did not (Group 2). Auditory discrimination improved slightly at the kindergarten level if the full length test, including items subsequently excluded on the basis of local dialect, was considered (Group 2). The overall level of auditory skills at the primary level showed change only in Group 1, from an exceptionally high error total

down to the level of later groups.

3. Program effects. Significant evidence for immediate program effects (SKG) was very marked in the area of visual-motor skills, more moderate in nonverbal auditory memory, and slight for auditory discrimination. Small advantages of third summer over second summer preschool entrants was seen in these three areas, but none of these community diffusion effects was statistically significant. At kindergarten level, however, large significant differences in visual-motor skills were seen between the most recent cohort and the two groups entering the program in the first year. Verbal sequential auditory memory (digits) was not affected either directly by the program nor by indirect community diffusion.

Entering the program at preschool rather than at kindergarten level had the apparent effect of preventing early decline in gross-motor and in auditorily-based abilities. The observed lag of Group 1 in auditory abilities relative to other groups was largely recovered over the long term (by post-grade 3). The lag in gross-motor development was lessened but not recovered to the level of the other groups by grade 2 assessment.

Excepting the recovery of Group 1 from a very low auditory skill level, no long-term program effect was seen on either visual or auditory abilities. At least for the three cohorts now in primary school, early acceleration of visual-motor development has not appeared significant over the long term into the primary grades. All three groups were approximately equal in visual-motor skills at the latest assessment.

Language Acquisition

Receptive language (structural-functional) as measured by Circus 8, How Words Work (see Table 11 and Figure 9), showed a small nonsignificant difference between Groups 3 and 4 (second and third summer) at pre-SPS level. The difference between pre-SPS and pre-SKG groups assessed concurrently in the second summer (Groups 3 and 2) was also small and nonsignificant. Group 2 gained considerably in Circus 8 score over the SKG period ($T = 1$, $p < .01$), and the proportion of variance in pre- and post-SKG scores attributable to assessment occasion was significant if variance due to age was first removed ($R^2_{C8.Age, Occasions} - R^2_{C8.Age} = .184$, $F(1, 18) = 4.57$, $p < .05$; this was the only measure among those subjected to multiple regression analysis where removal of age-related variance affected the significance of the proportion of variance due to occasions; see Appendix C). Even following SKG gains, Group 2 was still below the kindergarten norm for children with preschool experience ($z = -.41$). The lag was less than a half SD at this age-grade level, as opposed to over 1 SD preceding SKG ($z = -1.22$). Groups beginning SPS in the second and third project summers (3 and 4) were also over 1 SD below the appropriate norm, i.e., for nursery school children without preschool experience ($z = -1.14$ and $z = -1.28$ respectively). Excepting Group 3 pre-SPS, boys were ahead of girls in Circus 8 score, but even the largest sex difference (Group 2 pre-SKG) was not statistically significant.

Unlike Circus 8, the Metropolitan Readiness Tests Listening subtest (see Table 12) of receptive language (comprehension of

TABLE 12

METROPOLITAN READINESS LISTENING SUBTEST: MEAN SCORES AND STANDARD DEVIATIONS
BY AGE-GRADE LEVEL, GROUP AND SEX

	Entire Group			Girls			Boys		
	M	SD	N	M	SD	N	M	SD	N
Pre-SPS									
Gp 3(75)	1.3	.9	10	1.7	1.1	3	1.1	.9	7
Pre-SKG									
Gp 2(75)	4.6	2.5	8	3.5	2.4	4	5.7	1.0	4
Pre-Grade 1									
Gp 2(75)	5.6	1.2	8	5.7	1.3	4	5.5	1.3	4

TABLE 13

CIRCUS 10-II (SAY AND TELL): MEAN SCORES AND STANDARD DEVIATIONS FOR
FULL (1975) AND REDUCED (1975,1976) ITEM POOL, BY AGE-GRADE LEVEL,
GROUP AND SEX

	Entire Group			Girls			Boys		
	M	SD	N	M	SD	N	M	SD	N
Pre-SPS									
Gp 3(75)	45.2 (16.2)	12.2 (3.8)	12 (12)	42.7 (16.0)	11.7 (3.7)	4 (4)	46.4 (16.2)	13.1 (4.1)	8 (8)
Gp 4(76)	(16.1)	(0.7)	(5)	(16.5)	(0.7)	(2)	(15.7)	(0.6)	(3)
Pre-SKG									
Gp 2(75)	50.5 (17.2)	7.9 (3.4)	8 (8)	49.5 (17.2)	8.2 (4.5)	4 (4)	51.5 (17.2)	8.7 (2.5)	4 (4)
Post-SKG/Pre-Grade 1									
Gp 2(75)	55.7 (20.1)	6.0 (2.4)	8 (8)	55.0 (19.5)	2.6 (1.7)	4 (4)	56.5 (20.7)	8.7 (3.1)	4 (4)

NOTE: Scores in parentheses pertain to reduced item pool (from 38 items scored at 2 points each to 25 items scored at 1 point each).

content messages) showed a significant difference between pre-SPS and pre-SKG level children (Groups 3 and 2) in the second summer ($U = 7.5$, $p < .01$). Group 2 did not increase in Listening score during SKG as it did in Circus 8 score, and the proportion of variance in pre- and post-SKG scores attributable to occasion was nonsignificant. The only substantial sex difference, favouring boys, occurred in Group 2 pre-SKG but disappeared by the end of SKG.

Expressive language as measured by Circus 10-II, (Say and Tell; see Table 13 and Figure 10) was similar to the Circus 8 receptive language measure in showing no initial difference (on shortened Circus 10 form) between first and second summer SPS groups (3 and 4). Again as with Circus 8, the difference between SPS and SKG scores on the full Circus 10 item pool preceding the second summer's program was not significant. The increase in Group 2 Circus 10 score during SKG was significant, whether short or long form was considered ($T = 0$, $p < .025$), although multiple correlation analysis did not show a significant proportion of variance over this period to be related to assessment occasion. Even more than with Circus 7, the language specialist noted a strong relationship between age and Circus 10 scores of SPS children preceding the third project summer ($r = .84$, $p < .001$ for all 3- and 4-year-old children who responded to Circus 10; the seven youngest children in the fifth project group were unable to respond). Sex differences in Circus 10 scores were small, with even the largest, favouring Group 3 boys pre-SPS on the full Circus 10 test, being nonsignificant.

Review of language acquisition results according to project purposes

1. Characterization. In the preprimary years, structural aspects of receptive language were very substantially below age norms (Groups 2, 3 and 4) and remained so at pre-grade 1 assessment (Group 2). Expressive language was also below norm in the early years but by a much smaller margin, and surpassed age norm by the pre-grade 1 assessment (Group 2). In the two large project groups, expressive language was an outstanding early discriminator of highest and lowest achieving children (kindergarten level in Group 2; preschool level in Group 3; see Appendix D). Receptive language also clearly predicted highest and lowest achievers, although by a somewhat smaller margin. In Group 3 the highest achiever was identified by a receptive language advantage on initial assessment. In Group 2, the lowest achiever was identified earlier than the highest achiever by an extreme receptive language score. Sex differences in language acquisition were not consistent or significant.

2. Patterns of change. The structure-oriented receptive language measure (Circus 8) showed significant positive change during SKG (Group 2). The expressive language measure showed positive but much smaller immediate change during the program period. In expressive language, age level differences were apparently greater in the years preceding than during preschool and kindergarten. Some early decline in expressive language abilities among children entering the program at kindergarten rather than preschool level was suggested by teacher observation in the first summer.

3. Program effects. No community diffusion effect from second to third summers was evident in either receptive or expressive structural language measure. Teachers informally noted much improvement in the expressive language of second summer preschool entrants over first summer entrants into both preschool and kindergarten.

The receptive language measure involving content messages showed a significant advantage of kindergarten age children with preschool experience over preschool age children entering school for the first time. Neither structure-oriented language measure, receptive or expressive, showed such an age and experience effect. The SKG program did show a significant immediate effect on structure-oriented receptive language, but only after removal of age-related variance (Group 2). Immediate program effect on expressive language was smaller and significant in ordinal but not proportion-of-variance terms. An increase in the quantity but perhaps not the quality of verbal expression was emphasized in teachers' anecdotal reports during the second and third summer programs. Content-oriented language was not significantly improved during the SKG program period (Group 2). Scores remained well below maximum on this readiness level measure.

Conceptual-Factual Learning and Academic Achievement

Conceptual-factual learning (preschool and kindergarten levels)

At pre-SKG assessment with the Circus 2 measure of quantitative and relational concepts (see Table 14 and Figure 11), Group 4 exceeded the score made by Group 2 two years earlier in the history of the program. This difference fell short of

TABLE 14

CIRCUS 2 (HOW MUCH AND HOW MANY): MEAN SCORES AND STANDARD DEVIATIONS FOR FULL AND REDUCED (1975, pre-SPS, Group 3 only) ITEM POOL BY AGE GRADE-LEVEL, GROUP AND SEX

	Entire Group			Girls			Boys		
	M	SD	N	M	SD	N	M	SD	N
Pre-SPS									
Gp 3(75)	(20.0)	(5.9)	(11)	(23.7)	(4.6)	(4)	(17.8)	(5.7)	(7)
Pre-SKG/Post-KG									
Gp 2(75)	30.4	4.4	8	29.0	4.2	4	31.7	4.8	4
Gp 4(77)	33.8	3.5	5	37.5	.7	2	31.3	1.2	3
Post-SKG/Pre-Grade 1									
Gp 2(75)	34.4	2.0	8	34.7	2.6	4	34.0	1.4	4

NOTE: Scores in parentheses pertain to selectively reduced item pool (digit items omitted, 30 rather than 40 item total).

TABLE 15

OPPOSITES (VERBAL CONCEPTS): MEAN SCORES AND STANDARD DEVIATIONS BY AGE-GRADE LEVEL, GROUP AND SEX

	Entire Group			Girls			Boys		
	M	SD	N	M	SD	N	M	SD	N
Pre-SPS									
Gp 3(75)	10.6	3.1	12	11.4	2.8	4	10.2	3.3	8
Gp 4(76)	12.4	2.7	5	11.7	4.6	2	12.8	1.8	3
Pre-SKG									
Gp 2(75)	13.5	1.5	8	14.0	1.8	4	13.0	1.3	4
Post-SKG/Pre-Grade 1									
Gp 2(75)	14.6	1.6	8	14.5	1.9	4	14.7	1.5	4

Figure 11. Circus 2 (How Much and How Many): Mean scores by age-grade level and group.

Note. Selectively reduced pre-SPS C2 assessment not included.

Figure 12. Opposites (Verbal concepts): Mean scores by age-grade level and group.

Figure 13. Circus 11 (Do You Know): Mean scores by age-grade level and group.

Figure 14. Circus 5 (Finding Letters and Numbers -- capital letter and number sections): Mean scores by age-grade level and group.

statistical significance. The gain made by Group 2 over the SKG period was statistically significant ($T = 3$, $p < .025$). Approximately one third of the variance in Circus 2 scores over this period, a highly significant proportion, was attributable to assessment occasion ($R^2_{C2.Individuals, Occasions} - R^2_{C2.Individuals} = .340$, $F(1,9) = 15.30$, $p < .01$). Group 2 attained the norm for kindergarten children with preschool experience only after this SKG gain, whereas Group 4 children scored at this norm before the summer of their kindergarten year. Group 4 girls exceeded boys at this post-kindergarten assessment, as did Group 3 girls by an even larger margin at the pre-SPS assessment in the second year when a selectively shortened form of Circus 2 was used (items containing digits removed; $U = 5$, $p < .025$). Kindergarten level children in the second summer (Group 2) did not show a significant sex difference.

The Opposites measure of verbal concepts (see Table 15 and Figure 12) showed a difference between the two groups (3 and 4) assessed at the pre-SPS level in favour of the group beginning later in the history of the program (Group 4). This difference fell short of statistical significance. Preceding the second program summer, SKG (Group 2) children were significantly ahead of the younger SPS (Group 3) children ($U = 21$, $p < .025$). During that SKG session, Group 2 children increased significantly in Opposites score ($T = 5.5$, $p < .05$). As with quantitative and relational concepts (Circus 2), a significant, but much smaller, proportion of variance over this period was attributable to assessment occasion ($R^2_{Opposites.Individuals, Occasions} - R^2_{Opposites.Individuals}$).

duals = .172, $F(1,10) = 5.85$, $p < .05$). Sex differences in Opposites scores were small and inconsistent.

General factual knowledge as measured by Circus 11, Do You Know (see Table 16 and Figure 13), was significantly higher for SKG children than for SPS children assessed concurrently during the second project summer ($U = 22.5$, $p < .05$). Change in Group 2 scores over the SKG period was highly significant ($T = 0$, $p < .025$) as was the proportion of variance, approximately one third, over this period due to assessment occasion ($R^2_{C11.Individuals, Occasions} = R^2_{C11.Individuals} = .339$, $F(1,8) = 64.53$, $p < .01$). By the end of SKG, Group 2 children were above the Circus 11 norm for kindergarten children with preschool experience ($z = +.24$), a change from almost 1 SD below the norm ($z = -.90$) pre-SKG. The year younger SPS children were also well below the appropriate norm preceding their entrance into the summer program ($z = -.69$ in terms of the norm for nursery school children with no preschool experience). Sex differences in Circus 11 scores were negligible (within 1 point).

Letter-number knowledge as measured by Circus 5 (see Table 17 and Figure 14) was dramatically higher at the SKG (Group 3) than at the SPS (Group 2) level preceding the second project summer ($U = 1$, $p < .001$). The Circus 5 scores for Group 2 increased significantly during the SKG program ($T = 0$, $p < .025$) and showed the largest proportion of variance due to assessment occasion of all the 14 measures used with Group 2 pre- and post-SKG and subjected to multiple correlation analysis (See Appendix C;

TABLE 16.

CIRCUS 11 (DO YOU KNOW): MEAN SCORES AND STANDARD DEVIATIONS BY AGE-GRADE LEVEL, GROUP AND SEX

	Entire Group			Girls			Boys		
	M	SD	N	M	SD	N	M	SD	N
Pre-SPS									
Gp 3(75)	23.0	3.6	11	22.3	4.6	4	23.4	3.3	7
Pre-SKG									
Gp 2(75)	25.4	4.3	8	25.2	5.6	4	25.5	3.4	4
Post-SKG/Pre-Grade 1									
Gp 2(75)	29.4	2.0	7	30.0	2.0	4	29.3	1.5	3

NOTE: All scores prorated for 2 specifically U.S. items omitted.

TABLE 17

CIRCUS 5 (FINDING LETTERS AND NUMBERS - CAPITAL LETTER AND NUMBER SECTIONS): MEAN SCORES AND STANDARD DEVIATIONS BY AGE-GRADE LEVEL, GROUP AND SEX

	Entire Group			Girls			Boys		
	M	SD	N	M	SD	N	M	SD	N
Pre-SPS									
Gp 3(75)	4.6	2.6	9	6.0	2.8	2	4.1	2.7	2
Pre-SKG									
Gp 2(75)	10.0	2.6	8	9.0	2.4	4	11.0	2.6	4
Post-SKG/Pre-Grade 1									
Gp 2(75)	12.9	1.7	8	12.0	1.4	4	13.7	0.5	4

TABLE 18

EARLY CHILDHOOD INVENTORY IDENTIFICATIONS OF BODY PARTS, SHAPES, COLOURS: MEAN SCORES AND STANDARD DEVIATIONS BY AGE-GRADE LEVEL, GROUP AND SEX

	Entire Group			Girls			Boys		
	M	SD	N	M	SD	N	M	SD	N
Pre-SPS									
Gp 3(75)	22.6	5.9	10	24.0	6.2	3	22.0	6.2	7
Pre-SKG									
Gp 2(75)	24.4	4.4	8	24.2	1.7	4	24.5	6.6	4
Post-SKG/Pre-Grade 1									
Gp 2(75)	26.4	2.1	8	25.7	1.7	4	27.0	2.6	4

$R^2_{C5.Individuals, Occasions} - R^2_{C5.Individuals} = .426, F(1,9) = 13.65, p < .01$). Preceding SKG, Group 2 children were at the norm for kindergarten children with preschool experience and they surpassed that norm by early grade 1. SPS children were far below the appropriate norm on program entrance, obtaining approximately half the mean score for nursery school children with no preschool background. At both Group 2 assessments boys were ahead of girls, while Group 3 girls exceeded boys at pre-SPS assessment.

Knowledge of colour, shape and body parts as measured by the Early Childhood Inventory (see Table 18) showed a pattern of results similar to the measures of general (Circus 11) and letter-number (Circus 5) knowledge. SKG children (Group 2) were ahead of the younger SPS children (Group 3) preceding the second summer session, and scores for the SKG children increased during the session. In the case of Early Childhood Inventory scores, however, the difference between age groups and the SKG gains in the second program summer were not statistically significant. The proportion of variance in SKG scores due to assessment occasions was also nonsignificant, while variance over this period due to individual children was highly significant ($R^2_{ECI.Individuals} = .912, F(10,10) = 10.33, p < .01$). Sex differences were very small but similar to Circus 5 in favouring boys at both Group 2 assessments (pre- and post-SKG) and girls at the Group 3 pre-SPS assessment.

Academic achievement (grade school levels)

Four of the Stanford Achievement Tests (see Table 19 and summary in Table 21) have been administered to all three groups (1, 2, and 3) who had completed grade 1 by last assessment date: Word Meaning, Paragraph Meaning, Vocabulary and Word Study Skills. Following grade 1, Group 3 showed consistently higher grade equivalent (GE) scores in these tests than the two groups preceding it in the program and having less preschool and kindergarten experience. In all cases, differences between Group 3 (with 6-week SPS plus full-year kindergarten) and Group 2 (with 3-week SPS and 6-week SKG) were greater than differences between Group 2 and Group 1 (with just 6-week SKG). The post-grade 1 GE increment between Group 1 and Group 3 two years later was largest for Word Study Skills (+.7 GE), followed by Word Meaning (+.5 GE), Vocabulary (+.3 GE) and Paragraph Meaning (+.2 GE, but +.4 GE between Groups 2 and 3 after a .2 GE decrement between Group 1 and Group 2; this was the only achievement measure where an earlier group exceeded a group starting later in the history of the project). Spelling (Stanford), not used with Group 1, showed a larger increment in post-grade 1 score between Groups 2 and 3 than any of these other four verbal tests (+.8 GE). Word reading as assessed by the Slosson test (see Table 20 and summary in Table 21), requiring an oral response rather than the written (choice marking) Stanford responses, showed the largest increment of any achievement area between Groups 1 and 3, +1.3 GE. Unlike these language-related areas of achievement, Arithmetic (Stanford) scores post-grade 1 for Group 1 and Group 3 were identical.

TABLE 19

STANFORD ACHIEVEMENT TESTS (PRIMARY I AND II BATTERIES): MEAN SCORES AND
STANDARD DEVIATIONS IN GRADE EQUIVALENT UNITS, BY AGE-GRADE LEVEL, GROUP
AND SEX

[illegible]

Word Study									Arithmetic (computation/concepts)								
Entire			Girls			Boys			Entire			Girls			Boys		
M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N
1.4	.3	6	1.6	.5	2	1.3	.1	4	2.0	.4	6	2.0	.5	2	2.0	.5	4
1.6	.3	8	1.7	.3	4	1.6	.4	4									
2.1	1.2	12	2.0	.5	4	2.1	1.4	8	2.0	.6	12	2.0	.4	4	1.9	.7	8
1.7	.3	6	1.6	.3	2	1.8	.2	4									
									2.9	.6	8	2.5	.5	4	3.2	.5	4
									2.5	.7	8	2.2	.4	4	2.7	.9	4
									4.1	1.1	4	4.3	1.0	2	4.0	1.7	2
									4.1	1.1	4	4.2	.1	2	3.9	1.8	2

Language									Science								
Entire Cp			Girls			Boys			Entire Cp			Girls			Boys		
M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N
2.3	.4	8	2.4	.2	4	2.1	.5	4	2.5	.5	8	2.7	.6	4	2.3	.5	4
2.5	.4	4	2.4	.3	2	2.5	.6	2	3.0	.7	4	3.5	.6	2	2.5	.1	2

TABLE 20

SLOSSON ORAL READING TEST: MEANS AND STANDARD DEVIATIONS IN GRADE EQUIVALENT UNITS, BY AGE-GRADE LEVEL, GROUP AND SEX

	Entire Group			Girls			Boys		
	M	SD	N	M	SD	N	M	SD	N
Post-Grade 1/Pre-Grade 2									
Gp 1(75)	.6	.5	5	.9	.5	2	.4	.4	3
Gp 3(77)	1.9	1.6	11	2.0	.7	4	1.9	.9	7
Post-Grade 2									
Gp 2(77)	3.0	1.1	8	2.7	.7	4	3.3	1.5	4
Post-Grade 3									
Gp 1(77)	4.5	1.7	4	5.4	.3	2	3.6	2.3	2

TABLE 21

SUMMARY OF ACADEMIC ACHIEVEMENT RESULTS (SLOSSON AND STANFORD TESTS): MEAN GRADE EQUIVALENTS BY AGE-GRADE LEVEL AND GROUP, WITH COMPARISON DATA FROM 1975 (POST-GRADE 2)

	Slosson Oral Reading	Stanford Achievement Tests							
		Word Mng.	Paragr, Mng.	Vocab- ulary	Spell- ing	Word Study	Lang- uage	Sci- ence	Arith- metic
Post-Gr 1/Pre-Gr 2									
Gp 1(75)	0.6	1.3	1.7	1.5		1.4			2.0
Gp 2(76)		1.4	1.5	1.5	1.1	1.6			
Gp 3(77)	1.9	1.8	1.9	1.8	1.9	2.1			2.0
Post-Gr 2 (75)		(1.7)				(1.9)	(1.8)		(2.0)
Gp 1(76)		2.0	1.9			1.7			
Gp 2(77)	3.0						2.3	2.5	2.9 ^a
Post-Gr 3									
Gp 1(77)	4.5						3.0		4.1

NOTE: Scores in parentheses pertain to comparison data (children 1 year preceding SPS-SKG groups).

^aArithmetic computation score, arithmetic concepts score + 2.5, combined arithmetic grade equivalent score = 2.7.

At higher grade levels (post-grade 2 and post-grade 3), no groups were assessed concurrently with the same achievement measures. In all cases where two grade levels were administered an achievement measure on the same date, older children were ahead of younger. The differential was particularly large between post-grade 2 and post-grade 3 (Groups 2 and 1) Slosson Oral Reading scores (1.5 GE), and particularly small between post-grade 1 and post-grade 2 (Groups 2 and 1) Word Study scores (.1 GE).

Group 1 grade equivalent scores in Paragraph Meaning and Word Study increased during grade 2 but lost ground relative to grade norms—from .7 GE and .6 GE respectively below post-grade 1 norm, to 1.1 and 1.3 GE respectively below post-grade 2 norm. The post-grade 2 Word Study score was also .2 grade equivalents lower than the Word Study score of post-grade 2 children one year previously who had not been part of the SPS-SKG project. The Word Meaning score of Group 1 children during their grade 2 year rose considerably more, +.7 GE, but was still a year below grade norm. This score did, however, surpass by .3 GE the Word Meaning score of grade 2 children in the previous year (preceding SPS-SKG program). Over the 2 school years from post-grade 1 to post-grade 3, Group 1 increased Slosson Oral Reading score dramatically by 3.9 GE, from .4 GE below grade norm to .5 GE above norm. Arithmetic scores during this 2 year period increased by just over 2 grade equivalents, falling within .1 GE of norms at both post-grade 1 and post-grade 3 assessments.

Arithmetic was ahead of all language-related achievement areas and Word Meaning (comprehension and reading of words) was

the lowest academic skill in Group 1 following grade 1. At the most recent assessment occasion 2 years later, word reading as measured by the Slosson test surpassed the arithmetic grade equivalent score. The third achievement area assessed post-grade 3, Science (Stanford), is content rather than skill oriented.

Here Group 1 children were more than 1 GE below their concurrent levels in Word Reading and Arithmetic, and 1 GE below grade norm.

Group 2 averaged just under half a grade equivalent below norm on language-related achievement measures post-grade 1. One year later following grade 2 the Slosson test of word reading was at norm, while Language and Science (both Stanford tests not used previously with Group 2) were .7 GE and .5 GE respectively below grade norm. The Science score, however, was .7 GE higher than the last group of grade 2 children who had not been part of the SPS-SKG project. The Group 2 Slosson score represented a lesser deficit in grade equivalent terms than seen in the year older (post-grade 3) Group 1 children who had entered the program at SKG rather than SPS level. As with Group 1, Group 2 Arithmetic scores at most recent assessment were intermediate between Slosson word reading and Science scores. For Group 2 post-grade 2, the mean arithmetic score (average of computation and concepts tests) was .7 GE above that of children 2 years previously without pre-school or kindergarten, but .3 GE below grade norm. Group 3, completing grade 1 at most recent assessment, approached grade norms in all achievement areas with little spread among areas (range = .3 GE).

Slosson reading and Stanford arithmetic (computation) scores for all project children in grade school at latest assessment, Groups 1 and 2 as well as Group 3, approached grade level at the last assessment date (see Table 21). Grade 1 (Group 3) was above grade norm (+.5 and +.1 GE respectively for reading and arithmetic) and grades 2 and 3 (Groups 2 and 1) were within .1 GE of grade norm. In comparison, the grade 2 class just preceding children from SPS-SKG programs (1975) averaged 1.2 GE below grade norm in two Stanford verbal subtests (Word Meaning, Word Study) and in Science, and 1 GE below grade norm in Stanford Arithmetic. The grade 3 class just preceding project children (1976) averaged 1.5 GE below grade level in four Stanford verbal subtests (Word Meaning, Paragraph Meaning, Spelling, and Word Study), while the first grade 3 composed of project children was .5 above norm in Slosson Oral Reading and at norm in Stanford Arithmetic tests.

Past comparison data available for higher grade levels in the Grösse Ile School (Stanford Achievement Tests for grades 4 and 5, 1975; Canadian Tests of Basic Skills for grades 4 to 7, 1971) show a gradual decline in verbal achievement areas (reading, language, vocabulary and word study) from (a) 1.5 GE below grade norm post-grade 4, to (b) 1.7 GE below norm post-grade 5, (c) 1.9 GE below norm post-grade 6, and (d) to 2.3 GE below grade norm post-grade 7. Arithmetic scores averaged .5 GE higher than language-related achievement at all levels.

Following grade 1, Group 1 girls were ahead of boys by .5 GE or less in Paragraph Meaning, Vocabulary and Word Study while boys were equal to or slightly ahead of girls in Word Meaning and

Arithmetic. By post-grade 3, girls' advantage in reading as measured by the Slosson test increased dramatically to +1.8 GE, in addition to a 1 GE advantage in Science. Boys and girls have remained much closer in Arithmetic scores, with girls only .3 GE ahead at last assessment. In Group 2, post-grade 1 Stanford Achievement scores for the two sexes were more similar, with the largest difference being a .3 GE advantage for girls in Spelling. By post-grade 2, Group 2 boys surpassed girls in Slosson reading and in Arithmetic by over half a grade equivalent (+.6 GE in both areas). Girls had a smaller advantage (+.3 GE) in Science and Language. In the Slosson Oral Reading Test and all Stanford Achievement Tests, Group 3 at the most recent assessment occasion (post-grade 1) showed negligible sex differences (maximum of .1 GE), even less than seen in group 2 at the same age-grade level.

Review of learning and achievement results according to project purposes

1. Characterization. On preschool entrance, Group 3 children were well below norms in general information and letter-number knowledge. On grade 1 entrance, Group 2 children were at norm in these two factual learning areas and in quantitative concepts. In Group 3, but not in the preceding Group 2, the highest achiever was identified by preprimary superiority in conceptual learning measures (Circus 2, Opposites). Factual learning was not as clear an early discriminator of high and low achievement in either large group. (see Appendix D).

At the most recent assessment, all three groups in the primary grades were near or above norm in both word reading and arithmetic.

achievement. Earlier assessments of project children had shown verbal achievement areas (including word reading) below grade level but arithmetic near norm. Science and language achievement areas (two oldest cohorts) were still below grade norm at last assessment.

Arithmetic and reading skill (Stanford Word Study, Slosson) most distinguished highest and lowest achievers during primary years. The greatest grade equivalent differential between high and low achievers was seen in word reading (Slosson) scores (see Appendix D). Vocabulary and language were the least discriminating measures in both large groups, with high and low achievers in Group 2 actually reversed in ranking on these two measures.

The spread among achievement areas was less in younger groups with more ECE experience than in older groups with less pre-grade educational experience. Sex differences were also less clear with each successive cohort.

2. Patterns of change. Positive change occurred in all assessed areas of conceptual-factual learning during the preprimary years, with letter-number knowledge showing the most immediate change after program entrance and the largest short-term (SKG) gain (Group 2). Achievement gains continued over the long term into the primary grades, with word reading measures showing the greatest improvement up to grade 3. Reading skills eventually surpassed arithmetic skill. At least in older cohorts with less early educational experience, positive changes in academic achievement continued and increased for some years following program participation. The largest yearly change so far in grade equivalent

terms occurred during grade 2 (Group 1). The full extent of gains in academic achievement relative to age-grade norms did not become apparent for some time after program participation.

3. Program effects. Short-term positive program effects were evident in all early measures of conceptual-factual learning used in the project. Quantitative concepts, general information and letter-number knowledge all showed large and significant gains over the SKG period (Group 2); the Opposites measure of conceptual learning showed more moderate gain; gain in knowledge of colours, shapes and body parts (Early Childhood Inventory) was nonsignificant apparently due to a ceiling effect. Both conceptual learning measures showed an advantage for the latest cohort over an earlier cohort (Circus 2 at the kindergarten level; Opposites and the pre-school level), but these advantages fell short of statistical significance.

Evidence for increasing program effectiveness with each program year was seen in achievement measures during the primary grades. In language-related measures, later cohorts surpassed earlier cohorts at post-grade 1 assessment (Groups 1, 2 and 3). The difference was greater between the groups entering the pre-school program in first and second summers (Group 2 versus Group 3) than between the groups starting preschool and kindergarten in the first project year (Group 2 versus Group 1). Arithmetic achievement at the grade 1 level did not appear to be differentially affected by program change from year to year, but was at norm.

Overall program effects on school achievement were evidenced by the recent grade equivalent superiority of primary level project

children over previous classes in Grosse Ile school. To date this superiority has become most apparent in the oldest cohort (Group 1, which is 2 grade equivalents above the previous grade 3 class in reading) even though this group experienced less pre-grade 1 educational intervention than the other groups. This pattern of increasing advantage with age contrasts the pattern of increasing deficit with age seen in Grosse Ile in previous years, and in older non-project children (grades 4 and 5) in the middle project year. At least with the cohorts now in primary school, program effect has continued to grow for some time following preschool-kindergarten participation.

To date, project children's greatest advantage over previous years has been in verbal academic skill areas (Group 1). This has had the effect of erasing the consistent inferiority of verbal relative to numerical achievement seen in the past. In the two oldest cohorts, in fact, reading skill (Slosson) surpassed arithmetic assessment. This change in achievement pattern occurred even though arithmetic was at grade level from the first primary school assessment, well ahead of the available past result (grade 2). A relatively smaller advantage was seen in the content area of science.

Interaction Process Ratings

Item Usability-Answerability

Of the six raters involved over the course of the project, only one (B) omitted a substantial proportion of items when rating

a group where she was a full-time teacher. From this, her first use of the rating items, to her second rating of the same group (pre- to post-SKG, Group 2) the proportion of omissions dropped very significantly from .25 to .08 ($\chi^2_A(1) = 53.69, p < .001$). Of the remaining five raters, three (G, E, F) made no omissions when they were teachers of the group being rated. Rater A made a .003 proportion and rater D a .002 proportion of item omissions as teachers of groups being rated. When rating groups where they were part-time teachers or assistants, raters A and E made a somewhat higher proportion of omissions (.05 and .09 respectively). The one occasion when a rater (F) was only a temporary observer - not a teacher or assistant - of the group being rated showed a much higher proportion of omissions than for any other rater on any other occasion (.38). All omissions involved preschool or kindergarten children, not older grade-school children. The total number of omissions over all administrations of the 30 interaction rating items did not vary significantly from child to child within any of the four groups.

Considering all administrations of the rating items, there was a significant variation in answerability among items ($\chi^2_F(29) = 85.39, p < .001$). The items were ranked into five usability-answerability categories based approximately on the distribution of omissions, as follows: category 1: $z = \text{under } -1$ (3 items); category 2: $z = 0 \text{ to } -1$ (11 items); category 3: $z = 0 \text{ to } +1$ (9 items); category 4: $z = +1 \text{ to } +2$ (4 items); category 5: $z = \text{over } +2$ (2 items) (see Table 22). The mean proportion of omissions per item was .03; the maximum proportion of omissions, occurring for

TABLE 22

FIVE CATEGORY RANKING OF INTERACTION RATING ITEMS FOR USABILITY-ANSWERABILITY, INTERRATER AGREEMENT, DISCRIMINATION AND PREDICTION OF ACHIEVEMENT LEVELS

Item	Usability - Answerability	Interrater Agreement	Discrimination		Prediction	
			Low ^a	High ^a	Low ^b	High ^b
1	2	4	1		1	
2	2	1	4	x	3	x
3	3	4	2		4	
4	2	1	3		2	
5	5	4	1		3	
6	2	2	4	x	2	x
7	1	1	3		3	
8	2	2	4	x	4	
9	3	4	3		4	
10	5	4	5	x	4	x
11	4	1	3		3	
12	1	3	2		3	
13	2	3	4		2	
14	2	1	4	x	2	
15	3	3	5		4	
16	2	2	4		4	
17	3	5	4	x	4	x
18	4	5	4		4	
19	2	2	3		4	
20	3	2	5		5	
21	3	3	5		3	
22	2	4	3		3	
23	2	2	4		4	
24	4	2	4	x	5	
25	4	3	4	x	4	
26	2	3	3		3	
27	3	3	5		4	
28	3	2	3	x	3	
29	3	5	3		4	
30	1	3	2		1	

Category

1: $\bar{z}$ under -12: $\bar{z}$ = 0 to -13: $\bar{z}$ = 0 to +14: $\bar{z}$ = +1 to +25: $\bar{z}$ over +2

(Omissions

range = 0 to

.08, $\bar{M}$ = .03)

Category

1: $\bar{z}$ over +12: $\bar{z}$ = 0 to +13: $\bar{z}$ = 0 to -14: $\bar{z}$ = -1 to -25: $\bar{z}$ under -2

(C range =

.62 to .89,

 $\bar{M}$ = .79)

Category

1: $\bar{z}$ over +1.52: $\bar{z}$ = +1 to +1.53: $\bar{z}$ = 0 to +14: $\bar{z}$ = 0 to -15: $\bar{z}$ = under -1

(see text for details of discrimination

-prediction point system)

Category

1: $\bar{z}$ over +1.52: $\bar{z}$ = +1 to +1.53: $\bar{z}$ = 0 to +14: $\bar{z}$ = 0 to -15: $\bar{z}$ under -1

^a Items indicated (x) discriminate predominantly high or low achievers (minimum 2:1 point ratio).

^b Items indicated (x) predict only high or low achievers.

3

TABLE 23

INTERRATER AGREEMENT ($\bar{C}$) AND RELIABILITY (γ) OF INTERACTION RATINGS BY
AGE-GRADE LEVEL

Age-grade level	Group	<u>N</u>	Raters	<u>C</u> (average)	Total negative γ
post-SPS(76)	4	5	D,E	.72	3
			D,C	.72	4.26
			C,E	.65	2.89
pre-SKG(75)	2	11	A,B	.74	2.67
post-SKG(75)	2	11	B,D	.79	2.2
pre-Gr 1(76)	3	15	D,E	.77	1.9
pre-Gr 2(76)	2	8	E,F	.89	} 0
pre-Gr 3(76)	1	6	E,F	.94	

the two items in category 5 (items 5 and 10), was .08. Each teacher-rater identified different items as "most difficult to rate". (together they identify items 4, 6, 15, 19, and 23), but all reported that the items generally become less difficult as the children become more familiar.

Inter-rater Agreement

The five age-grade levels which were rated by two or more raters (see Figure 1) showed a steady increase in proportion of agreement ($\bar{C}$ averaged over all 30 items) with increase in age-grade level, from .65 to .72 at preschool to .94 at grade 2 level (see Table 23). This increase occurred with level regardless of project year (1975-1977) or group size (from $N = 5$ to $N = 11$). Furthermore, the degree of systematic difference between a pair of raters in their ordering of children on the rating items (negative γ values) decreased as proportion of agreement rose. The proportion of agreement ($\bar{C}$) for individual items, averaged over 7 administrations with different pairs of raters and with groups from $N = 5$ to $N = 15$, varied from .62 to .89 with a mean of .79. The items were ranked into five proportion-of-agreement categories based approximately on the average $\bar{C}$ distribution, as follows: category 1: $\bar{z} = \text{over } +1$ (4 items); category 2: $\bar{z} = 0$ to $+1$ (8 items); category 3: $\bar{z} = 0$ to -1 (9 items); category 4: $\bar{z} = -1$ to -2 (6 items); category 5: $\bar{z} = \text{under } -2$ (3 items).

Patterns of Rating Distribution (see Appendix E for distributions of three selected items)

Approximately half the rating items showed a discernable monotonic trend as children grew older and/or experienced more

formal schooling. Three of these items (6, 20, 21) largely maintained the same modal rating at the extreme low end of the scale (1) throughout the project. For item 6 (fails to follow adults) the proportion of 1 ratings increased with age except in Group 1. In rating both Group 3 and Group 4 one year later, D assigned a higher proportion of 2's at preschool than at kindergarten level. At pre-grade 1 level, she assigned some 2's and 3's to Group 2, but none to Group 3 in the following year. Rating 1 also predominated throughout for item 20 (manipulates peers) and for item 21 (manipulates adults) excepting the youngest group rated (Group 3 pre-SPS). At preschool and kindergarten levels, D's ratings on these items showed a lower proportion of 1's for the group later in the history of the program (4) than for the group preceding it (3).

Two items (13, 14) shifted modal rating one step to the extreme end of the scale and relatively quickly obtained a high proportion of the extreme rating. Item 13 (refuses to follow peers) moved from a mode of 2 to a mode of 1; item 14 (enjoys verbal activities) moved from a mode of 3 to a mode of 4. The oldest children (Group 1 post-grade 3), with no SPS background, were exceptions to this pattern of mode concentration at one end of the scale, especially for item 13. At preschool, kindergarten and grade 1 levels, D assigned more extreme ratings (low for 13; high for 14) in the later-starting and more experienced group than in the group preceding it in program history.

Three items (7, 15, 25) shifted modal rating two steps from midrange to extreme 4 ratings. Item 7 (uses language in communication) reached 100 percent 4 rating by post-grade 1 (Group 3).

At all three age-grade levels where she rated two groups, D assigned more 4 ratings to the later than to the earlier and less experienced group. Items 15 and 25 (obtains materials and advantages, expresses affection to peers) shifted from mode 2 to mode 4, only reaching 100 percent 4 rating following grade 2 (Group 2). At post-SPS and post-SKG levels, D assigned item 15 (obtains materials and advantages) a mode of 3 in Group 3 but a mode of 4 in the more experienced Group 4. Preceding grade 1 she assigned a 3 mode to both Groups 2 and 3. For item 25 (expresses affection to peers), D's ratings did not show the later more experienced groups ahead in attaining a 4 mode.

Two items (16, 27) showed a similar overall pattern of mode shift from midrange ratings to a relatively high proportion of extreme 4 ratings, but involved an earlier and/or intermediate step of flatter rating distribution. For item 16 (uses materials effectively), only the youngest children rated (Group 3 pre-SPS) showed a 2 mode. The proportions of 3 and 4 rating generally increased with age and preschool experience. Group 4 obtained 100 percent 4 ratings following kindergarten; Group 3 never reached 100 percent 4 ratings even at 1 year older (.83 maximum); and Groups 2 and 1, at 2 and 3 years older, showed even lower proportions of ratings on the 4 mode (.67 and .75 respectively). At kindergarten and grade 1 levels, D assigned a 4 mode to the more experienced group and a 3 mode to the earlier starting and less experienced group. Item 27 (praise by adults rewarding) moved towards a 4 mode in all groups, with a high proportion of 4 ratings (up to 100 percent) at younger ages in more experienced groups.

Following preschool and kindergarten, the more experienced group (4) was given a higher proportion of 4 ratings by D.

Items 4, 8 and 30 moved from rather flat midrange distributions of ratings to a 3 or 4 mode. Items 4 and 8 (cooperates in group activities, and shows sense of social rules) moved to a 4 mode in the two youngest groups with most preschool experience (3 and 4) but fell back to a 3 mode in older but less experienced groups (1 and 2). At preschool, kindergarten and grade 1 levels, D's rating for item 4 (cooperates in group activities) clearly showed a 3 mode for the less experienced but a 4 mode for the more experienced group of the same age. Item 30 (shows flexibility) reached 100 percent 4 mode for Group 4 following kindergarten. This is a substantially higher proportion of 4 ratings than ever obtained by the other three groups even at higher age-grade levels (maximum .75). Groups 3 and 4 differed sharply in D's ratings at preschool and kindergarten levels, in the direction of the 4 extreme for Group 4.

The remaining items showed more complex patterns of change in rating distribution, among which three items (3, 18, 19) followed an apparently quadratic rather than monotonic pattern. Item 3 (exploration rewarding) began with a 2 mode in the two youngest groups rated (Group 3 pre-SPS and Group 4 post-SPS), moved towards the 4 extreme, and then moved back towards lower ratings, as children grow older. Cross-sectional comparison of D's ratings indicate that groups with more pre-grade 1 school experience followed the quadratic pattern at younger age-grade levels than groups preceding them. By the latest group in the project (4), the high proportion

of 4 ratings was not seen, but only the shift towards higher proportions of lower ratings. The quadratic patterns of ratings for items 18 and 19 (persists at tasks and leads group) showed less age difference among groups. For item 18 there was an intermediate high mode of 4 ratings, with a particularly high proportion in Group 3. For item 19, there was an intermediate high mode of 2 or 3 rather than 4 ratings.

Each of the remaining items had a relatively unique distribution. Items 10, 24 and 28 (uses peers as resources, uses materials with originality and follows verbal directions) all showed indications of quadratic trend. For item 24 (uses materials with originality), the younger more experienced groups moved to and maintained a 3 mode (with a particularly high proportion for Group 4) vs. a 2 mode for the earlier groups. This upward direction with the history of the program was seen in D's ratings at the kindergarten (Groups 3 and 4) and grade 1 (Groups 2 and 3) levels. For item 28 (follows verbal directions) a similar upward trend over time was seen in D's ratings at preschool and kindergarten levels (Groups 3 and 4), with only the youngest and most experienced group (4) attaining 100 percent 4 ratings.

Items 9, 11 and 12 (experimental in exploration, uses adults as resources, and selects activities independently) showed fluctuations in mode, but overall a discernable upward direction of change. For item 9 (experimental in exploration), extreme low ratings decreased while extreme high ratings increased. D's item 9 ratings at preschool, kindergarten and grade 1 levels clearly placed later

more experienced groups further towards the 4 end of the rating scale. Low ratings also decreased for item 11, and D assigned later-starting groups a higher proportion of 3 ratings compared to the low-to-midrange ratings of earlier groups.

Items 1, 17 and 22 (stands ground in novel situations, seeks adult approval, and reproduces dialogue-rhymes) showed only slight changes in mode except for the latest starting and most experienced group (4). For these items, only Group 4 moved from high proportion of midrange ratings to a clear 4 mode. Particularly for item 17 (seeks adult approval), D assigned the later groups ratings which are further towards the 4 extreme at all three levels where she rated two groups (post-SPS, pre-SKG or post-kindergarten and pre-grade 1). Item 2 (solitary play) showed a clear direction for the two later-starting and more experienced groups (3 and 4), in this case towards the low extreme (1) of the rating scale. Group 4 attained a 1 mode at an earlier age-grade level than did Group 3 which preceded it in the history of the program.

Relationships Between Interaction Process Ratings and Development-Achievement Measures

Rating Item Discrimination and Prediction of Achievement Levels

These results focus on the two highest and the two lowest achieving children in Group 2 and 3 at last assessment, and two of the children most changed in each direction for initial developmental to most recent achievement assessment (see Appendix E for for these individualized results on three selected items). The two negatively changed children studied in detail were those from

Group 2 (negative change in Group 3 is slight with multiply tied ranks); the two positively changed children studied were those from Group 3. A point system (see Method) quantified the degree to which each item distinguished these high, low and changed children from the modal rating of their groups. Based approximately on the distribution of points-per-item, the items were divided into (a) five categories of overall discrimination and (b) five categories of prediction (derived from the first rating occasion only; see Figure 22). Discriminative categories fell as follows: category 1: $z = \text{over } +1.5$ (2 items); category 2: $z = +1 \text{ to } +1.5$ (3 items); category 3: $z = 0 \text{ to } +1$ (9 items); category 4: $z = 0 \text{ to } -1$ (11 items); category 5: $z = \text{under } -1$ (5 items). Predictive categories were defined by the same z scores as discriminative categories and contain 2, 4, 9, 13, and 2 items respectively from category 1 to category 5.

Over all ratings to date, item 1 (holds ground in novel situations) was the best identifier of high-achieving individuals. Together with item 5, item 1 was also the best overall discriminator of extreme achievement levels regardless of direction (high or low), and was one of the two best predictors (with item 30) from early rating (first occasion) to later achievement level. High achievers were frequently rated above the modal rating (towards the 4 extreme) while low achievers were rated below the mode.

Second in discriminating power for high-achieving individuals were items 9 (experimental in exploration) and 19 (leads group). High achievers rated above the mode in both these items. Item 19 was also one of the two best identifiers of the two children

showing most positive change in rank from preschool (developmental) to grade 1 (achievement) rank. The second was item 20 (manipulates peers) which distinguished the two positively changing children by high early ratings, and was the best predictor of positive change from just the first rating occasion. Item 21 (manipulates adults) was just slightly lower as an early indicator of positive change, along with items 3 (freedom to explore rewarding), 9 (experimental in exploration), 27 (praise by adults rewarding), and 29 (expresses pride).

Third in discriminating power for high-achieving children were item 3 (freedom to explore rewarding) and item 26 (task completion rewarding). Except for the two most positively changing children in Group 3, the discriminating power of item 3 increased with age-grade level. High achievers tended to remain at a high rating in these two items rather than falling back to lower ratings as did the majority of children in Groups 2 and 3.

The clearest identifier of low achieving children, and of those showing most negative change from kindergarten to grade 2 was item 5 (attends to two things simultaneously). Low achievers and negatively changing individuals stayed quite consistently below the modal rating. Second in discriminating power for low achievers (and one of the two best early predictors in either direction) was item 30 (shows flexibility) where the low children rated below the mode. Item 14 (enjoys verbal activities), while only moderately discriminating overall, was the best early predictor of low achievement (below mode ratings). Item 28 (follows verbal directions) was third in discriminating low achievers, with

the two lowest scoring consistently below the group mode. Together with item 11 (uses adults as resources), item 28 also ranked second in identifying the two Group 2 children showing most negative change during the project. These negatively changing children were best identified early (first rating occasion) by item 2 (solitary play) and then by item 6 (fails to follow adults). In both items these children did not move to 1 ratings as early as the rest of their groups. Item 6 was unique in distinguishing only low achievers (fourth rank in both discrimination and prediction) while high achievers always received modal rating.

Among the most discriminating items overall (fourth), but identifying both high and low achievers rather than predominantly one extreme, was item 12 (selects activities independently). High (4) ratings were seen for high achievers while below mode ratings were seen for low achievers. Item 4 (cooperates in group activities) was only moderately discriminating over all rating occasions but relatively high (fourth) in the early prediction of high and low achievement. At kindergarten rating, Group 2 low achievers were below the mode; at preschool rating, Group 3 high achievers were two steps above the mode at the 4 end of the scale. Item 13 (refuses to follow peers) was well below item 4 in overall discrimination but equally powerful as a predictor of both the highest and lowest achievers from just the first rating occasion.

Correlation of Selected Interaction Rating Items with Development-Achievement Measures

Panel correlation of the three selected items (7, 19 and 30) with the development and achievement ranks of Group 2 children at

most recent assessment is presented in Figure 15. Item 7 (uses language in communication; see Appendix E) was in the highest category of both usability-answerability and interrater agreement ($C_{\text{average}} = .88$), but close to the median rank in overall discrimination (14.5 out of 30 items). This item discriminated high-achieving and low-achieving children about equally. Longitudinally, item 7 ratings moved in a clear monotonic pattern from midrange to the extreme 4 end of the scale, followed by increasing proportions of 4 ratings. Pre-SPS, item 7 ratings correlated with achievement measures at a moderately high level ($\gamma = .62$) similar to the concurrent correlation between development and achievement measures ($\gamma = .60$). Pre-SPS correlation between item 7 and developmental measures is considerably lower.

As a pre-SPS predictor of post-grade 1 achievement ($\gamma = .53$), item 7 fell close to both development ($\gamma = .51$) and achievement ($\gamma = .51$) measures. Predictive correlation of post-SPS item 7 ratings and post-grade 1 achievement was even higher ($\gamma = .78$). Predictive correlation with developmental measures was also highest at this point but still far lower than with achievement measures ($\gamma = .42$). Prediction from item 7 ratings to post-grade 1 development and achievement levels dropped sharply 1 year later. No subsequent correlations could be calculated due to the high proportion of 4 ratings for item 7 at the grade 1 level.

Item 19 (leads group; see Appendix E) was in the second highest category of both usability-answerability and interrater agreement ($C_{\text{average}} = .83$). It was one of the two best discriminators of high-achieving and of positively changing individuals.

Figure 15. Predictive (to post-grade 1) and concurrent correlations (Y) of selected rating items (7, 19 & 30) with development and achievement measures (Group 2).

The distribution of item 19 ratings suggested an overall quadratic pattern with high achievers maintaining high ratings rather than falling back to lower ratings as they grow older. The pre-SPS correlation of item 19 with achievement was lower than the concurrent correlation of development with achievement measures. There was negligible pre-SPS relationship between item 19 and developmental measures ($\gamma = .09$).

Prediction from pre-SPS item 19 ratings to post-grade 1 achievement ($\gamma = .41$) was lower than prediction from either development or achievement measures. The predictive correlation of item 19 and post-grade 1 achievement rose during preschool and kindergarten years ($\gamma = .41$, then $.51$, then $.69$) but fell preceding grade 1 ($\gamma = .18$). The high concurrent post-grade 1 correlation between item 19 and achievement ($\gamma = 1.00$) was due to just 4 children in the group who are not assigned the modal rating of 3. The highest child was rated 4 while the three lowest children were rated 1. The prediction of developmental rank from item 19 ratings gradually increased to a moderate concurrent correlation post-grade 1 ($\gamma = .65$).

Item 30 (shows flexibility; see Appendix E) was in the highest category of answerability and the middle category of interrater agreement ($C_{\text{average}} = .79$). It was the second best discriminator of low-achieving children and one of the two best early predictors of achievement at either extreme. Item 30 ratings generally moved from midrange to the 4 end of the scale with increasing age and preschool experience (only the last group to enter the program attained 100 percent 4 ratings). High

achievers were distinguished by higher item 30 ratings at younger ages, while low achievers were rated quite consistently below the mode. Of the three items included in the panel correlation, item 30 had the highest concurrent correlation with pre-SPS achievement measures ($\gamma = .66$), and an even higher correlation with pre-SPS developmental measures ($\gamma = .85$).

The predictive correlation from pre-SPS item 30 ratings to post-grade 1 achievement ($\gamma = .46$) was below that from pre-SPS achievement and development measures. Item 30 improved as a predictor until at post-SKG rating it had the highest predictive correlation with post-grade 1 achievement shown by any of the three items at any point ($\gamma = .85$). As with item 19, correlation fell preceding grade 1 but rose to a very high concurrent correlation with post-grade 1 achievement ($\gamma = .90$). Preceding SPS, item 30 showed a high concurrent correlation with developmental measures ($\gamma = .85$) and also a high predictive correlation with post-grade 1 developmental measures ($\gamma = .85$). The other two items being correlated did not show either concurrent or predictive relationships of this magnitude with developmental measures. The relationship between item 30 ratings and post-grade 1 developmental measure gradually decreased from preschool to concurrent grade 1 assessments.

Concurrent correlations between the pre-SPS ratings of all three items and achievement level were only slightly higher than the predictive correlations between pre-SPS ratings and post-grade 1 achievement (see Figure 16). Item 7 rose to its highest predictive value early, by post-SPS rating ($\gamma = .78$). Items 19 and 30 predicted best 1 year later following SKG, item 19 with a lower

Figure 16. Predictive correlations (γ) from selected interaction rating items (7, 19 & 30) to post-grade 1 achievement (Group 2).

Note. First and last sets of points represent concurrent correlations (pre-SPS and post-GR1) presented for comparison.

correlation ($\gamma = .69$), and item 30 with a higher correlation ($\gamma = .85$) than the highest predictive correlation of item 7. All three items fell rapidly following their peaks of predictive correlation. A sharp rise followed to high levels of concurrent post-grade 1 correlation between item ratings (items 19 and 30; item 7 showed no variation post-grade 1) and achievement ranks. These post-grade 1 correlations were very much higher than pre-SPS correlations, and higher than any predictive correlations seen.

Review of Interaction Ratings and their Relationships with Development-Achievement Measures According to Project Purposes

1. Characterization. Standing ground in novel situations (item 1) was the most salient characteristic of high achievers from preschool through the primary grades. They appeared distinct from the majority in this characteristic right from their earliest participation in the program. These children were also more flexible (30) than most during their early educational experience. High achievers were more experimental than their peers in exploring their environments (9) and they found greater reward in the freedom to explore for its own sake (3). These exploratory interaction characteristics were apparent at a very early stage in children who subsequently achieved beyond the level suggested by pre-SPS developmental assessment. As well as exploration, high achievers found task completion rewarding (26) regardless of others' attention or approval. Unlike many children who found self-directed exploration and completion less rewarding as they grew older, high achievers continued through the primary years to enjoy such independent interactions with the physical environment. They were

also able to select activities independently (12) with greater ease than other children.

Socially, high achieving children showed somewhat greater group cooperation at young ages than did their peers (4, 13) and they emerged as group leaders (19). Leadership was also shown by those who increased their achievement rank over time, and these positively changing children were in fact manipulative of their peers (20) during the early stages of their educational career. During this early period they were also somewhat more manipulative of adults (21) than their classmates, and were more rewarded by adult praise (27). These children who achieved increasingly beyond initial predictions were further distinguished by early expressions of pride in their work and their personal characteristics (29).

Poor ability to attend to two things simultaneously (5) was the most salient interaction characteristic of low achievers and of children whose school achievement declined in rank from pre-school developmental assessment. In contrast to high achievers, low achievers were markedly inflexible regarding changes in routine and activity (30). Also in contrast to top achievers, low achievers were somewhat less able than the majority to select activities independently (12) and less likely to stay in novel situations (1). Very early in their educational experience, these low achieving children showed less than average enjoyment of talking, listening and verbal interaction generally (14). Through the primary grades, low achievers and those falling increasingly below early predictions followed verbal directions poorly (28).

Socially, these low achieving children again contrasted high achievers in their poor group cooperation (4) and comparatively frequent refusals to follow peers (13). An above average level of solitary play (2) and of failure to follow adults (6) were early characteristics of children whose achievement rank subsequently fell over time. In both preschool and primary years these children also made relatively little use of adults for information or other help (11).

In recent primary level rating, highest and lowest achievers in both large groups were most widely divergent in their enjoyment of exploration for its own sake (3). In Group 2, the two children at achievement extremes also contrasted sharply in their independent selection of activities (12) and their use of adults as resources (11). In Group 3, where the highest and lowest achieving children were earlier and more widely separated, it was the dual focussing of attention (5) along with enjoyment of exploration (3) which most discriminated achievement extremes at the primary level.

2. Patterns of change. Among those areas of interaction showing an essentially linear pattern of change as children progress from preschool through the primary grades, the following characteristics showed the clearest and greatest change: use of language in interaction (7), defense of materials and advantages (15) and expression of affection to peers (25). Preschool and primary children showed more variability (range) in the following characteristics, but tended generally to increase steadily with age: effective use of materials (16), pleasure in adult praise

(27), acceptance into and cooperation in group activities (23 and 4), and flexibility (30). During the preschool and primary years, children generally reduced the following interactive characteristics to very low levels: manipulation of peers (20 and adults (21), failure to follow adults (6) and active resistance of peers (13). Enjoyment of language and verbal interactions (14) generally rose to a high level in the early educational years.

The following interactive characteristics showed more fluctuations as children grow older, but overall appeared to increase with age-grade level: experimental exploration (9) and independent selection of activities (12), and use of adults as resources (11). Among those interaction characteristics which showed nonlinear patterns of change, i.e., about half the items studied, several showed indications of quadratic trend. Children from preschool through primary grades tended to increase and then decrease in their persistence at tasks (18), pleasure in exploration (3), and leadership in group activities (19). Children's use of peers as resources (10), original use of materials (24) and following of verbal instructions (28) also appeared to increase and then decrease with age-grade level, although less definitively than the three above characteristics. Children who entered the project in the first two summers have shown little change in their level of seeking adult approval (17), standing ground in novel situations (1), or of using language to reproduce stories, rhymes, etc. (22). Children in the latest cohorts have shown marked early increases in these three interactive characteristics.

Highest and lowest achievers showed definite longitudinal changes in interaction ratings relative to the modal rating of their groups. The highest achieving and most positively changing Group 2 child showed a very dramatic change (increase of two rating steps) towards greater independence in selecting activities (12) soon after his developmental-achievement profile diverged to distinguish him clearly from the group. The lowest achiever and most negatively changed child in Group 2 showed a change from below to above average cooperation in group activities (4) at latest assessment (post-grade 2) when she also showed signs of gain relative to the group in certain development-achievement areas (e.g., verbal IQ, visual discrimination). The highest achieving Group 3 child continued to increase in group leadership following kindergarten entrance while other children actually decreased in leadership. He emerged as the principle group leader by the end of grade 1. The lowest Group 3 achiever became unique at last rating occasion (post-grade 1) in his level of expressing affection towards peers (25) and of repeating dialogue, songs, rhymes and other verbal material (22).

Sequential relationships between many interaction characteristics rated and development-achievement ranks differed across cohorts. Within a single cohort (Group 2), interaction characteristics differed markedly in their pattern of longitudinal change relative to development-achievement ranks. Use of language in interaction (7), for instance, predicted grade 1 achievement best following SPS, while leadership (19) and flexibility (30) predicted best following SKG. Flexibility is the only one of these three

predictively analysed items which predicted developmental as well as achievement rank with a high correlation ($\gamma = .85$).

3. Program effects. Among those interaction characteristics for which a linear developmental trend can be detected, several showed distinct advances of later cohorts over earlier cohorts. When ratings of a given age-grade level were examined, particularly those at preschool, kindergarten and grade 1 levels by teacher-rater D, the following items evidenced such direct and/or indirect program effects in the accelerated or extended development of later cohorts. Group cooperation (4) showed a clear upward shift at grade 1 level between children who entered the program in first and second summers, and again between children who entered in second and third summers. Effective use of materials (16) showed a similar positive change across these three cohorts. Use of language in interactions (7) showed an increased proportion of children on the extreme high mode across the three cohorts (Group 3 over Group 2 at grade 1 level; Group 4 over Group 3 at preschool and kindergarten levels). Defense of materials (15) and particularly flexibility (30) showed upward mode shifts at preschool and kindergarten levels between the latest (Group 4) and the preceding cohorts. Group 4 also showed more enjoyment of verbal activities (14) and of adult praise (27), and manipulation of both adults (21) and peers (22) during early education years than did preceding cohorts. Resistance to peers (13) decreased from both first to second summer preschool entrants (mode difference at grade 1 level), and from second to third summer preschool entrants (higher proportion on extreme low mode). Failure to follow adults decreased between first and second summer preschool entrants (not evident until grade 1

level) but not to date between the latest two cohorts.

Several marked differences between cohorts in rating distribution were also evident in interaction areas which did not show such a clear monotonic developmental direction. Experimental exploration (9) showed a definite increase across the three project summers with the difference between second and third summer entrants being particularly marked (two step mode difference at preschool and kindergarten levels). Use of adults as resources (11) and, even more distinctly, concern with adult approval (17) increased across the three cohorts. Original use of materials (24) increased as well from one cohort to the next, with the latest preschool group (4) showing less variation in rating than their predecessors. Each cohort also made more use of language to reproduce dialogue, etc., (22), with a particularly marked difference between the two latest cohorts at preschool and kindergarten levels (mode shifted upwards). A clear positive difference between the latest cohort and all earlier groups was seen in following verbal instructions (28), standing ground in novel situations (1), and in solitary play (2) during preprimary years where Group 4 moved to the lowest rating mode earlier than other groups.

DISCUSSION

Development and Achievement

Characterization of the Project Population

The most unexpected development-achievement characteristic of the project population as a whole was the early (preschool-kindergarten levels) superiority and the subsequent similarity of verbally-measured (PPVT) compared to nonverbally-measured (GH) intelligence. This finding challenges the verbal deficit assumption with which the project began, as well as the concurring conclusion from recent regional study (Cram, 1974) that children in this and comparable isolated communities lack specifically verbal skills rather than general intellectual abilities.

Some speculation is in order concerning the one occasion, out of 11 assessments using both IQ indicies, which did not conform to this pattern of verbal superiority over or similarity to nonverbal IQ score. Verbal IQ scores of the youngest group preceding the first project summer (Group 2 per-SPS) were lower than those of the older group in that same summer, than those of any subsequent preschool group, and than the nonverbal scores of this or any other group on any occasion. Project staff observed that before the initial summer program parents were much more apprehensive about school attendance for their preschool-aged children than for their kindergarten children who would be entering

school (grade 1) in any case in three months time. Very probably the younger children were less prepared and confident regarding the exposure to strangers and the new environments and activities of school, and otherwise shared their parents uneasiness. Perhaps the verbal task was more unfamiliar and intimidating than the nonverbal draw-a-man task in its formal structure, question-answer format, and its close interaction with a strange adult. One might expect the score-depressing effects of apprehension and a strange situation to dissipate in a short time following exposure to the situation. This appears to have occurred here due largely to attitudinal changes on the part of both children and community parents.

A striking contrast was seen between the verbal-nonverbal pattern of intellectual development in the Grosse Ile population and findings from a similar preschool-kindergarten project in a U.S. urban setting (Beller, 1973, p. 550). Evaluation of the Philadelphia project, which made longitudinal use of the same two intelligence measures (PPVT, GH), revealed early verbal scores down in the 4th quartile rather than in the 3rd quartile as here. The Philadelphia verbal scores only reached initial (preschool) Grosse Ile levels by grade 4, and never reached the highest group scores seen here to date. Philadelphia nonverbal scores were higher initially than either verbal or nonverbal scores in Grosse Ile, but subsequently fell rather than rising as in this project. Such comparison of relative verbal-nonverbal IQ profiles provides initial confirmation of the assumption that a distinct local pattern of strengths and weaknesses among development-achievement areas

would be discovered in this relatively isolated community.

The margin here between verbal and nonverbal intelligence seemed less, and even opposite in direction during the early years, than in the bulk of ECE literature which is predominantly from the urban U.S.A. It is significant that the director of an experimental preschool in a small Canadian city (Wright, 1975a) also made the observation, similar to Grosse Ile results, of comparable verbal and nonverbal ability levels in her longitudinal evaluation studies. She suggested that perhaps, due to this more even pattern of ability development, a specific language emphasis in ECE is less relevant in this country than in the U.S.A. Reanalysis of existing ECE evaluation data and design of future studies in terms of cultural setting variables would be necessary to determine whether the verbal-nonverbal similarity seen in Grosse Ile is related to rural-urban factors, to Canadian-U.S. factors, and/or to other as yet undocumented community factors. Such ECE evaluation tied to cultural setting variables is largely lacking to date but, as demonstrated here, is urgently needed if ECE and its study is to be relevant beyond a single model of early development-achievement or of cultural deprivation.

Further impetus for seeking an alternative to or refinement of the verbal deficit assumption regarding Grosse Ile children came from case study development-achievement profiles. Verbal versus nonverbal content did not seem to affect the power of intelligence measures at the preschool level to predict high and low primary school achievers -- neither IQ measure was remarkable in this respect. Likewise the verbal-nonverbal distinction seemed

irrelevant to the power of auditory memory measures as early discriminators of later achievement extremes -- both sentence repetition (verbal) and tapping (nonverbal) memory tasks appeared outstanding predictors of high and low achievement. At higher age-grade levels following program participation, certain nonverbal achievements (e.g., arithmetic computation) were close to certain verbal skills (e.g., word reading) in discriminating the highest and lowest achievers in the two large project groups.

Together with group-averaged results, these case study observations suggest future culture-specific early childhood studies which use a variety of alternative development-achievement models or categories. It is suggested from the present results, for instance, that the important factor(s) of cultural disadvantage in this population must be (a) an underlying cognitive ability which cross-cuts verbal and nonverbal development, and/or (b) a specific component or subskill of those verbal abilities particularly relevant to school success. Theory-based evaluation (Fitzgibbon & Morris, 1975) is required, focussing on both more specific and more comprehensive development-achievement characteristics than the verbal-nonverbal distinction being discussed here, and on alternative concurrent and sequential interrelationships among these characteristics. Such theory-based evaluation can aid generally in determining an effective level of analysis for development-achievement variables as well as specifically in characterizing a given group or population.

Some direction towards an alternative analytical focus on the verbal deficit assumption is suggested by language assessments

of project children. Receptive language (Circus 8) remained substantially below age norms up to grade 1 level whereas expressive language (Circus 10) had well surpassed age norm by pre-grade 1 assessment. This suggests that while Grosse Ile children in pre-primary years are limited in the language structures they comprehend, they make fuller use than the average of those structures they have assimilated. Since receptive precedes expressive language acquisition, it also suggests that language development may become noticeably retarded in later years in consequence of the early, but largely masked, lag in receptive language acquisition. In recent primary level assessment, in fact, those achievement areas which deal most with the use and content messages of language, i.e., Stanford Language and Science subtests, were markedly lower than was achievement in less contextualized, applied academic skills, i.e., reading and arithmetic mechanics.

Case study development-achievement profiles suggest some cultural limitation or levelling effect specific to this area of language use as distinct from functional linguistic and reading skills per se. Highest and lowest achievers in the two large groups were close (and sometimes reversed in ranking) in language and vocabulary scores in contrast to wide margins in several basic academic skills, both verbal and nonverbal. Even at earliest assessment when both receptive and expressive measures of structural language acquisition were outstanding predictors of high and low primary school achievers, the third language measure concerned with recall, comprehension and interpretation of content messages (Metropolitan Readiness Listening subtest) was not. In light of

the similarity in verbal and nonverbal intellectual development results and relatively poor achievement in content subjects, it is suggested that this area of apparent developmental weakness may not be language-specific. It may rather involve higher order cognitive and/or affective processes general to the contextualized application of acquired skills in many areas of development. Further work might include assessment of more contextualized application of nonverbal skills, e.g., spatial reasoning, physical-mechanical problem solving.

The comprehensive Rand (1974) proposal for future ECE study supports a focus on such pervasive higher order cognitive-affective processes, i.e., strategies for using cognitive abilities and for communicating and play, which the Rand researchers label meta-cognition and metalanguage (p. 398). Attention to these processes of applying abilities in context may be a key to understanding patterns of development and learning particular to at least certain types of disadvantaged situations, and eventually a key to designing appropriate ECE intervention and its evaluation. In the last few years, several very prominent developmentalists and early childhood educators have recognized that these processes of contextualized application may be quite specific to the various subcultures of North America (e.g., Cole & Bruner, 1972; Shapiro, 1973; Sigel, 1972a, 1974). Based on this realization, a refined goal for ECE is nurturance of the capacity to apply abilities within both the dominant, mainstream culture and one's own subculture, and to shift appropriately according to context. This capacity for using abilities in context is largely comprised of the child-environment

interactive processes which, with experience in the Grosse Ile program, came to replace a confined language development goal.

Unlike the assumption of general weakness in the verbal relative to the nonverbal abilities of these children, the assumption of weakness in auditorily-based skills was at least initially confirmed. At preprimary levels, sequential auditory memory (digits) was below age norms in both first and last project years. Primary level assessments throughout the project consistently showed a preponderance of auditorily-based errors in component language and reading skills. A direct relationship between the auditorily-based skills measured at primary levels and school achievement was suggested by Group 1 results. Only by the end of grade 3, when Group 1 had reduced auditorily-based difficulties to the level of other project groups, did these children progress sufficiently to attain expected grade equivalents in basic achievement areas (reading, arithmetic). Case study profiles further emphasize the importance of auditory skills to the school success of Grosse Ile children. Out of the development-achievement batteries used at the preschool level, two auditory memory measures (sentences, tapping) were the clearest predictors of later high and low achievers. Highest achievers subsequently showed greater advantages over their groups in auditory than in visual skills.

It is noteworthy that auditory discrimination as measured by Slingerland screening subtest 7 was the weakest perceptual-motor ability at all primary assessments, while auditory discrimination as measured by Circus 7 was close to age norms at preprimary levels. Also noteworthy was the keen nonverbal auditory

discrimination among preprimary children remarked by the staff as these results were discussed, e.g., Grosse Ile children could identify individual fishing boats by only the sounds of their motors. The group of Circus authors note that phoneme discrimination difficulty in their standardization sample is much less than in numerous other ECE studies which find auditory discrimination to be the root factor in cultural disadvantage (e.g., C.P. Deutsch, 1964; M. Deutsch in White & Watts, 1973; Zigmond & Cicci, 1968). The Circus 7 authors suggest that other measures incorporate more tasks irrelevant to auditory discrimination per se. This is certainly true of the Slingerland measure which demands auditory memory involving sequential processing (selection of beginning or end sound), visual and kinaesthetic recall of letter symbol related to phoneme, and motor response (letter formation). In contrast, the Circus measure involves whole meaningful words with picture association, rather than the abstraction of single language sounds, and demands only minimal memory, sequential processing or fine-motor response (picture marking). It was perhaps (a) the relatively abstract nature of the Slingerland task (low context or meaning; symbolic rather than concrete) and/or (b) its demands for integrating several perceptual-motor skills into memory and sequencing tasks which caused difficulty for Grosse Ile children, not auditory discrimination itself.

Support for this reinterpretation of the observed auditory skill deficits in the Grosse Ile population came from several other perceptual measures. Slingerland subtest 6 was only slightly less difficult than subtest 7. Subtest 6 again involves integrating

auditory input with visual-motor output, and even more than subtest 7 emphasizes sequential memory of abstract sound-sight units. In contrast, the third auditorily-based Slingerland subtest (8) was considerably less difficult than the first two (6 and 7); for girls even less difficult than the two most complex visually-based subtests. Relative to the other auditory subtests, Slingerland 8 involves considerably less abstracting of auditory stimuli, less integration of the auditory mode with other perceptual-motor skills, and less sequential processing of auditory or visual input (whole visual units are presented for matching to auditory stimuli, with minimal memory or motor demands). Among auditory skills assessed at preprimary levels, it was the abstract sequential memory test (digit span), not memory for a meaningful sentence unit nor auditory discrimination, which was below norm.

Although overall the visual mode appeared relatively strong in this population, increased demands for abstraction, integration and sequential memory increased the difficulty of visually-based as well as auditorily-based tasks. At all age-grade levels, it was tasks demanding fine discrimination and sequencing and the integration of several perceptual-motor abilities which were most difficult, and which were seen to discriminate high and low achievers where more simple and concrete visual-motor tasks did not. The abstraction and integration-sequencing factors incorporated in these low performance perceptual measures, both auditorily-based and visually-based, are central to phonic reading instruction which Grosse Ile teachers continually reported as the greatest learning problem for their children, and which showed the greatest achievement

deficit early in the program (Slosson, Group 1).

Despite these clear weaknesses in certain academically relevant perceptual-motor skills, it was intelligence and achievement measures, rather than perceptual-motor measures with either auditory or visual basis, which most widely separated high and low achievers during primary years. Several recent large scale ECE studies also found perceptual-motor characteristics relatively poor in differentiating and/or predicting particularly competent, successful children (Shipman, 1972, 1976; White & Watts, 1973). These findings in combination with the intelligence, perceptual-motor and language results here suggests refinement of a commonly accepted hypothesis concerning the nature and origin of culturally disadvantaged learning patterns, i.e., the model of verbal deficiency deriving from an early auditory perception deficiency (C.P. Deutsch, 1964; Parker & Day, 1972). Blank (1973a, 1973b) has suggested that the assumed perceptual deficits of disadvantaged populations, particularly auditory discrimination and cross-modal deficits, are in fact more comprehensive cognitive deficits.

The concept which Blank proposes for the capacity involved -- symbolizing, coding and organizing the environment -- corresponds closely to the abstracting, integrating and sequencing processes suggested here as a basic learning difficulty for Grosse Ile children. Although this cognitive difficulty was apparently generalized to some degree across various ability areas, perhaps the capacity to incorporate the auditory mode into such processes was a very particular weakness in this community. Recent work in the field of learning disabilities is revealing the importance of such cross-

modal integration and sequential processing to academic achievement and particularly to successful reading (e.g., Doehring, 1975; Gibson & Levin, 1975). Particularly implicated is temporal ordering (Bakker, 1970) which would be most directly reflected in auditory memory. Poor auditory memory abilities and the power of auditory memory, both verbal and nonverbal, as an identifier of the highest achievers were repeated observations in this population. These findings suggest that particular instances of culturally disadvantaged learning difficulties may derive from cognitive processing deficits similar to those being defined in learning disability children. Research here and elsewhere aimed at further characterization of such particular difficulties, and at designing pertinent intervention, may find valuable direction in this growing body of learning disability research regarding analysis models for development-achievement variables.

Patterns of Change and Program Effects

Differences seen in the change patterns of verbal and non-verbal intelligence measures seem consistent with rejecting the hypothesis of a specifically verbal deficiency as the limiting factor in the development and achievement of this population. Both longitudinally and cross-sectionally, the verbal measure rose to near norm level with less direct and indirect intervention. In light of accumulated past ECE evaluation results, interpretation of these quick early gains largely in terms of specific verbal and test response skills, in addition to the attitudinal changes as discussed above, seems more reasonable than an intelligence gain interpretation. Such skills and attitudes may have been particularly

emphasized in the program, e.g., question-answer patterns. Nonverbal intelligence scores required more time and program exposure to rise to near norm level. As compared with Wright's urban Canadian preschool evaluation (1975a), the PPVT gain pattern here was characteristic of high-SES ECE participants, while the more gradual GH gain was characteristic of low-SES children who eventually reached levels similar to their high-SES peers in cognitive measures. Grosse Ile children thus appeared more like urban disadvantaged children in nonverbal than in verbal development. For this population, whatever intellectual capacities were measured by the GH may have been more indicative of global intelligence as it is commonly understood, while PPVT scores may have reflected more specific learnings during the preschool and primary years.

The limited data available regarding short-term change in development-achievement measures (Group 2 SKG session changes primarily) provide some support for the initial assumption of more immediate change in specific, concrete learnings than in comprehensive higher order capacities. As with the verbal deficiency assumption, however, observed patterns of change suggest some refinement of such an assumption for this particular population. It is proposed that the development areas showing most short-term change were those which least involved the abilities apparently basic to the learning deficiency of these children, i.e., certain higher level applied cognitive processes and particularly abstract integrating and sequencing processes as they involve the auditory mode.

The four development-achievement measures which showed both significant SKG change and a very high proportion of variance (over one third) attributable to occasion over the short summer program period largely support this interpretation. These measures include general information, quantitative concepts, and especially letter-number knowledge and visual-motor skills. All these measures quickly rose to age norm level, and visual-motor skill, with continued to be assessed longitudinally, remained stable thereafter. It is suggested that children were ready to learn in these areas and did so rapidly once exposed to the skills and material involved. Their initial below norm scores evidently reflected lack of cultural experience in these areas but no fundamental ability deficit. Letter-number knowledge showed the most dramatic immediate gain, perhaps because this area involves visual-motor skills -- which are normally developing very rapidly at this age -- in a very specific concrete learning.

A second group of development-achievement measures also showed significant short-term change but a more moderate proportion of variance (over one tenth) due to assessment occasion. These measures -- verbal conceptual learning, receptive language and non-verbal auditory memory -- were all important predictors of later school success. Their early improvement during a short program session is thus an encouraging result. These achievement related abilities can apparently be positively affected by short-term ECE as provided here. Future work might be directed towards increasing the program effect in these areas since both the need and possibility for further improvement are indicated by present results. In

particular, a wide gap remained between the receptive language level of the most and the least successful individuals, and the mean receptive language level remained below age norms even after gains were made.

Among the remaining areas which showed lesser and/or nonsignificant degrees of short-term change were those apparently age appropriate and not needing improvement, and those needing improvement but remaining unaffected by the program over the short-term. Among the latter, sequential auditory memory for verbal material, which seemed highly discriminating of subsequent high and low achievers, requires future attention regarding its relation to school achievement and its resistance to intervention. In relation to nonverbal memory, this verbal sequential memory ability appeared more closely tied to age preceding any educational intervention and less susceptible to positive change related to intervention. Also showing little change over short-term ECE participation was auditory discrimination of phonemes comprising local dialect. It is of interest, however, that perception of some standard English phonemes did improve during the SKG period. Even at this early age, these children appeared to be learning to cope with a standard dialect in addition to their own common usage. This learning, considered by many language development specialists as crucial to the success of cultural minorities in North American society (Ervin-Tripp, 1972; Labov, 1970) might well be overlooked in culturally adapted curricula and, as partially occurred here, in assessment procedures. Beyond this early discrimination learning, the auditorily based language skill deficiencies of this population appeared largely

resistant to improvement throughout primary grades to date, thus underlining a need for further study.

Recall, comprehension and interpretation of the content messages of language (Metropolitan Readiness Listening) also merits further study concerning short-term resistance to intervention, as well as long term follow up which was not done here. It would be of interest, for instance, to examine the relative importance of maturation and experience in the preschool-kindergarten discrepancy seen at the beginning of the second summer. Neither expressive nor receptive structural language measure showed the wide age difference seen in the content-oriented language measure. This area is of particular importance since it may involve the same higher order cognitive abilities as those content-focussed achievement areas which did not subsequently gain at the pace of basic skills, and/or did not as clearly distinguish high achievers from the group average. Expressive language, while averaging in age norm range, might be also studied regarding elimination of wide individual variability since it was one of the most discriminating early development-achievement measures.

Both intellectual development measures fell into the category of minimal short-term change, both with and without previous preschool experience. Over the longer term, either preceding or including SKG, both did rise significantly in the two large groups. The earliest stage of this positive change is presumed due to familiarization and specific test-taking skill learning. These effects can account for at least a 3 to 5 point improvement for naive test takers as documented by Campbell (1967, p. 218), and

possibly much more in the first preschool group here. Any further improvement, and particularly the long-term maintenance of any gains, would provide much more significant evidence of positive program effects on conventionally measured intellectual development. To date the first two cohorts have shown continued gains in both verbal and nonverbal IQ measures beyond their initial gains. This result contrasts sharply with the major evaluated ECE programs (1966-1972; see Introduction) which commonly showed a substantial decline in IQ soon after the initial dramatic gain. It also contrasts with IQ studies of disadvantaged populations not subject to special educational or other cultural intervention in which IQ declined steadily with increasing age (Anastasi, 1968, p. 208). The only decline to date for Grosse Ile children was small and statistically nonsignificant, and occurred only three years after program participation in the group which entered the program at kindergarten rather than at the younger preschool level. The IQ scores remained close to normative level even with this minimal decline, and further follow up is needed to determine if this was the beginning of a ~~more~~ substantial decline, a levelling off in IQ gain, or merely a temporary plateau in a pattern of continuing gain.

Two factors of documented significance in other situations may account for the relative success of this program in maintaining and continuing intellectual development gains following ECE participation. The first is the initial and increasingly more direct involvement of parents in the preschool-kindergarten program and in associated educational experience for themselves (Bronfenbrenner, 1975; Gotts, 1973; Horowitz & Paden, 1973). The second factor is

the involvement of teachers from the community and the regular school program, with the result of great continuity between summer ECE and later full year programs for these children (Banta, 1971; Miller & Dyer, 1975; Weikart, 1970), and perhaps also the further result of considerable community diffusion effect outside the school program per se.

Even more clearly than IQ, academic achievement showed continued gains in all follow-up to date subsequent to program participation. Both slow-acting effects of preschool-kindergarten and/or continued input via school and community may be responsible for continuing improvement. Results indicated a longer lag from ECE experience to academic achievement gains than to IQ gains. This reverses the sequence seen in several studies of inner city children (Dyer & Miller, 1974) where achievement gains occurred more immediately following Head Start experience than did IQ gains. As in the area of verbal and nonverbal intellectual development, this difference may reflect either Canadian-U.S., urban-rural or other factors of cultural difference among the populations studied. Further work is advocated regarding the degree of generality to rural Canada or specificity to this local situation in the observed patterns of change in intelligence and achievement measures.

The greatest program effect on academic achievement in relation to past levels and standardized norms has been in the improvement of reading mechanics. These gains have in fact eliminated, and in the two oldest cohorts actually reversed, the large deficit of verbal relative to numerical academic skill traditionally seen in Grosse Ile and other disadvantaged settings. While this

particularly large specific gain may have been the result of the initial language development emphasis by program personnel, further analysis is necessary in light of the limited gains to date in higher level language use and content areas. Recent theorizing and research concerning the developmental lateralization of perceptual-motor and language capacities (e.g., Satz & Van Nostrand, 1973; Semmes, 1968) suggests the possibility that an unintended program effect may be reflected in this rather specific differential pattern of academic achievement gains. As suggested in reference to learning disability children, these Grosse Ile children may be developing a slow perceptually-based learning style, and especially reading strategy, at the expense of faster language-based strategies (Bakker & Satz, 1970). For these children, this may be either a necessary stage to be followed at a later date (normally between 8 and 11 years of age) by a switch to the language-based approach, or it may be an unnecessarily limiting pattern of cognitive development. It will be most important to follow these children past the grade 5 level which in the past has been the observed levelling-off point for academic achievement in this community. If such levelling-off appears in preschool-kindergarten participants, the relative dominance of perceptually-based and language-based abilities, especially reading skills, should be closely examined for a possible role in limiting academic progress. Program elements related to perceptual and language emphases would also need examination for possible alteration and a planned variation type of evaluative study.

In many development-achievement areas this study cannot clearly attribute longitudinal change within cohorts to program-specific factors as distinct from maturational or cultural change factors. To do so would require the ideal proposed by developmental theorists such as Wohlwill (1973): an extensive set of prototypic functions for defined developmental dimensions, describing both the nature and norms of long term change. There is no realistic expectation of ever establishing such functions which would be local and contemporary enough to unequivocally isolate longitudinal program effect. This study, like other accumulating longitudinal developmental research, can merely work towards such prototypic functions in order that subsequent ECE work, especially in the same or a similar community, will have a somewhat more extensive comparative base for detecting and interpreting alterations in the pattern and rate of development.

Cross-sectional comparisons in the present study provide sounder evidence for program effectiveness than do longitudinal data. In IQ and achievement areas, comparisons with national norms and with the two preceding cohorts in the community at two points in time (pre-program and at the most recent assessment occasion) indicate meaningful gains related to preschool-kindergarten participation. Furthermore, convincing evidence is seen that each yearly improvement in the program (increase in child, parent and teacher participation), in combination with additional time for any community diffusion effects, increased program effectiveness. To date, two cohorts have surpassed preceding cohorts in verbal intelligence scores. Academic achievement gain over past comparison data has been quicker and greater for each successive cohort entering

primary school after preschool-kindergarten participation. Continued monitoring of cohort differences will be revealing concerning the long term stability of these yearly increments in effect, e.g., will Group 3 surpass Group 2 in verbal IQ, and Group 4 surpass Group 3 in nonverbal IQ; will the three later cohorts with preschool as well as kindergarten increase, hold or decrease their achievement margin over Group 1 by the time all complete primary school?

Each successive cohort has also shown reduced differences between sexes in academic achievement level, and reduced spread among the areas of achievement considering both sexes together. Sex differences in cohorts entering the project early in its history conformed to patterns observed in many developmental and school achievement studies. Boys excelled in nonverbal while girls excelled in verbal intellectual ability; boys excelled in visual-motor skills and especially in fine visual discrimination while girls excelled in auditorily-based skills, in early gross-motor and in later visual memory abilities. However, only in the first cohort with no preschool experience do girls show a clear school achievement advantage which increases with age through the primary grades. The second cohort showed inconsistent sex advantages across achievement areas, and the third cohort showed negligible sex differences. This third cohort also reversed the relative levels of verbal and nonverbal ability considered typical of each sex and seen in the earlier project groups. Group 3 girls excelled in all perceptual-motor areas, visual as well as auditory, and in nonverbal intelligence, while boys excelled in the verbal intelligence measure. This youngest primary school group also eliminated the verbal-numerical

gap in academic achievement one and two years earlier, respectively, than the two preceding cohorts.

In contradiction of the early Head Start results (Cicirelli, 1969), some clear positive effects seem related specifically to the summer preschool component of the present project. Both verbal and nonverbal IQ measures (both are age-corrected measures) showed significant gains in the first year preschool group before any further educational experience. In these intellectual development measures, and to a lesser degree in academic achievement, a superiority of preschool over kindergarten entrants (Group 2 versus Group 1) which was not at all apparent initially (older group actually exceeded younger group) emerged as these first year entrants proceeded through primary school. This lag in the effect of participation in the preschool program component, in conjunction with lags in the full effects of SPS-SKG experience on academic achievement generally (see above), strongly supports the practice of long term follow up evaluation of even very short-term programs.

Preschool entrance also had the apparent effect, noted by several researchers in a variety of settings (Beller, 1973; Bronfenbrenner, 1975; Spaulding, 1971), of preventing decline in several development-achievement areas. Most immediately apparent was gross-motor ability in which the younger preschool entrants far exceeded older kindergarten entrants in the first summer. Teachers in the first summer also noted very marked superiority of preschool children over kindergarten children in auditorily-based skills relating to language and reading. This observation was supported by the early primary level Slingerland screening scores of the two groups in

which kindergarten entrants (Group 1) made twice the auditory error score of the other project groups. While Group 1 auditory skill deficiencies relative to the groups with preschool experience were slowly eliminated over several years (by post-grade 3), gross motor deficits were not, at least by latest grade 2 level assessment.

As expected, considerable evidence for the power of indirect community effects, apart from actual program participation, is seen in the cross-sectional comparison of pre-SPS scores in a range of development-achievement areas: verbal IQ between first and second program years; nonverbal IQ and memory, and quantitative and verbal conceptual learning between second and third years; and visual-motor skills between both first and second and second and third program years. The two largest statistically significant diffusion effects, involving verbal IQ and visual-motor skills, were both still apparent after one year's program participation by the groups involved. The large and early community diffusion effect in verbal intelligence scores may well have been at least partly due to familiarization and changed parental attitudes, as was postulated regarding the large early longitudinal gains in verbal IQ by the first preschool cohort. Several factors may have been involved in the particular susceptibility of visual-motor skills to indirect diffusion as well as direct longitudinal program effects: the generally rapid development of visual-motor skills during the preschool to primary years (Piaget & Inhelder, 1969); the early and continuing attention given to this area by project teachers (e.g., crayons and scissors sent home after the first summer); and

the apparent general strength of this population in visually-based compared with auditorily-based skills. The possibility of cultural changes other than the ECE program being responsible for these large community diffusion effects must be considered, although no probably alternative cause has been noted.

Language acquisition (Circus 8 and Circus 10) and language sound discrimination (Circus 7) showed little or no community diffusion effect between second and third summers in which they were used. The high age correlation seen among very young children for two of these measures (Circus 7, Circus 10) suggests they may be more closely tied to maturational variables, and thus more resistant to intervention, than some other development-achievement area assessed. On the basis of teacher anecdotal reports, however, one suspects that a community diffusion effect occurred at least in expressive language between the first and second summers before formal language measures were introduced. One wonders whether the higher expressive in comparison to receptive language level of second summer program entrants reflects a stable cultural characteristic and/or any such indirect program effect. It would be valuable to further explore relative expressive and receptive language levels at younger ages both before and after educational intervention, and in relation to urban-rural and other cultural variables. Culture-specific patterns of early expressive and receptive language development, and of their susceptibility to the indirect and/or direct effects of ECE intervention, may have particular relevance in light of their predictive power and the later deficits in higher order language use seen here.

Although the development-achievement profiles of highest and lowest achievers in the two large program groups may have been largely idiosyncratic, certain patterns in these case studies support other evidence for a significant indirect community diffusion effect of this ECE program. In the group of first year preschool entrants (Group 2), initial development-achievement profiles of the highest and lowest primary school achievers were not distinctly separated, nor were the developmental ranks of these two children at the extreme top and bottom of their groups. In the group of preschool entrants one year later in the history of the program (Group 3), highest and lowest achievers were clearly identified by extreme developmental ranks and widely separated developmental-achievement profiles right from initial assessment preceding any educational intervention. At this first assessment, the highest achievement profile in Group 3 already showed many characteristics only seen one year later in the Group 2 highest achievement profile, i.e., particular superiority over the rest of the group in expressive language, nonverbal auditory memory and sequential visual-motor skills. (It should be noted that only the latter area was assessed in Group 2 pre-SPS.) In addition the Group 3 high achiever showed an initial superiority in conceptual learning (Circus 2, Opposites) not seen in the Group 2 high achievement profile even one year later after preschool participation. He also had a wider margin over his group in verbal intelligence score, initially and even more markedly post-grade 1. All these developmental-achievement areas in which the Group 3 high achiever showed earlier or greater superiority than the Group 2 high achiever

seemed susceptible to indirect community diffusion effects as seen in group-averaged data.

Continued follow-up of these high achievement individuals will determine if this earlier advantage, and additional advantages in areas possibly related to the postulated higher order cognitive deficit of these children, is reflected in greater and/or longer-continuing development-achievement gains. To date this appears so in the exceptionally wide margin which the Group 3 high achiever has gained over his group in all areas but vocabulary and certain visual-motor skills. Together with all other results discussed regarding the characterization of, patterns of change in and program effects on the development and achievement of Grosse Ile children, these individual profiles confirm teacher observations of vast individual differences among this superficially homogeneous group of children. As discovered in the Educational Testing Service (Shipman, 1976) and Harvard (White & Watts, 1973) early childhood studies, much individual variability even at very young ages cannot be accounted for by cultural (e.g., SES, urban-rural), family (e.g., size, stability) or child (e.g., early developmental capacities) variables. Here as in these two large studies, there is a need for much more descriptive case study involving those children showing outstanding resilience and competency in their lives in and out of school.

Interaction Processes

Characterization of the Project Population

In discussing the interactive characteristics of the project population, the three principle elements of child-environment interaction which are most prominent in the rating items used here, and in the literature from which they were derived, provide a useful orientation: (a) interactions between the child and the overall physical-situational aspects of his or her environment, (b) interactions between the child and the other persons, both peers and adults, in his or her environment, and (c) the child's use of language as a means of interacting with both physical and social environments. It was the first element, quantity and quality of interaction with the physical environment, which most sharply distinguished both high and low achievers from their classmates and from each other. Among these essentially nonsocial, nonverbal interactive characteristics which clearly related to school achievement were enjoyment of exploration, independent selection of activities, standing ground in novel situations and attending to two things simultaneously. This discriminating power of physical world interaction characteristics was seen both predictively from the preschool years and concurrently through the primary grades. High achievers' self-directed, persistent, outgoing and creative approach contrasted with low achievers' more dependent, uncertain, rigid and simple approach to their physical surroundings.

This finding is supported by the Head Start-Educational Testing Service longitudinal study (Bridgeman & Shipman, 1975) which

showed persistence and concentration (achievement motivation components) at the preschool level, to contribute substantially to prediction of grade 3 success in reading and mathematics. That study found a substantial correlation between teacher ratings on Schaefer's (1976; and see Introduction) task orientation dimension at both preschool and grade 3 levels, and achievement at the grade 3 level. Similarly Kohn and Rosman (1972b) found Schaefer task orientation ratings at kindergarten level to be useful predictors of grade 2 achievement. In Miller & Dyer's (1975) longitudinal followup of four preschool programs, it was the Montessori program with its emphasis on self-directed interaction with the physical environment which showed most positive and lasting effects on reading and mathematics scores at grade 2 level. It was the physical-situational interactive characteristics of inventiveness and curiosity, measured using Banta's autonomy battery (1970; and see Introduction), which showed gains in the preschool year preceding any cognitive-achievement gains, and were still strong relative to other programs at grade 2 assessment. The present findings and their support in these larger studies suggest that emphasis on positive interaction with the general physical-situational aspects of the environment is perhaps more important in preschool situations, at least in the Grosse Ile situation, than more specifically social or verbal interactions.

Among the competence-based constructs of interactive development reviewed in the introduction, effectiveness motivation (Sharp & Stott, 1976; and see Introduction) seems best to describe the interactive element which emerged here as most related to school

success. The summary definition offered by Stott and colleagues for this construct of child-physical world interaction is "willingness to face novel and complex tasks" (Stott & Albin, 1975, p. 159). It is positive interaction with novelty (item 1) which most clearly characterized high achievers, and a low level of interactive complexity in task situations (item 5) which most clearly characterized low achievers in this population. It is of note that the item used here as indicative of interactive complexity, dual focussing of attention, was also identified by the Harvard study (White & Watts, 1973) as one of the two most salient characteristics of competent preschool children. As emphasized in the introduction regarding all child-environment interaction characteristics, this significant element of interaction with the physical-situational environment involves both the affective (e.g., motivation as emphasized by Stott and colleagues) and the cognitive (e.g., competence as emphasized by White and colleagues) development of the child. Further structured observation, both ethological and experimental in approach, is advocated in this area, as well as additional teacher rating studies.

Social interactions appeared second in importance to physical world interactions as predecessors and correlates of primary school success for Grosse Ile children. With their peers, high achievers' interactions were frequent and positive, and increasingly dominant through the primary grades. Project teachers in the first two summers noted a general lack of peer interaction skills in Grosse Ile children relative to other populations in their experience. Perhaps it was this lack in the majority rather than an exceptional

quality and quantity of peer interactions in the highest achievers that was reflected in these results. In relations with both peers and adults, poor achievers showed an early low level and a later relatively negative style of social interaction. Among the social interaction characteristics rated, the best students contrasted poor achievers most dramatically in their effective use of the adults around them. White and Watts (1973) found social interaction characteristics to be more distinctive than cognitive or linguistic factors in identifying competent preschoolers. In particular, the effective use of adults as resources was one of the two (with dual focussing of attention, as discussed above) most discriminating characteristics in their study of the "competent" or "educable" preschool child. The present study confirms this often-questioned link between social interaction and achievement, and supports one of Piaget's few directly stated ECE principles: the importance of social collaboration in cognitive development (Piaget, 1969; Kamii, 1972).

The third element in child-environment interactions, use of language in dealing with physical and social worlds, distinguished low achievers but not high achievers from their classmates in this study. It is perhaps the lowest achievers in this population which would benefit from the intensive early language stimulation advocated by many urban U.S. early childhood educators and researchers. This finding has implications for overall ECE programming, but also for early identification and individualization for a minority of particularly low language children. The basic program should probably have an exploratory and social interaction emphasis, to be

combined with -- not replaced by -- a specific language emphasis for those individual cases of exceptionally low language development. For the majority of Grosse Ile children the original language deficit premise of the project was again challenged by these interaction results as it was by the intelligence, language and perceptual-motor results. Development related to early nonverbal interactions seemed more basic than early verbal interaction to academic success in this community as a whole. It must be remembered that the apparently greater importance of physical world compared to social world interactions, and of both compared to verbal interactions, may well be specific to the Grosse Ile setting. To the extent that this hierarchy can be generalized to other situations, however, it has relevance to the ongoing enrichment-instruction debate in ECE. The enrichment argument is supported, with qualifications for exceptional low language individual cases.

Patterns of Change and Program Effects

This study demonstrates the clear observability of change from preschool through primary years in many aspects of interactive behaviour, and the wide diversity of patterns of developmental change among the various interaction characteristics observed. The present findings concerning the observability and diversity of change in interactive characteristics during early childhood are supported in a number of recent larger scale longitudinal ECE studies. A similar degree and complex variety of changes, i.e., differences in form, speed and individual variability (range) across interactive variables, was observed by Wright (1975b, 1976). Based partly on data derived from White and Watts' (1973) social

competency scale, both Wright (1975a) and the Educational Testing Service researchers (Shipman, 1976) reject the contention of White and colleagues that competency is established by 3 years of age. Both these studies also confirm the need for future work which is sensitive to complex change patterns in interactive characteristics, particularly to curvilinear or nonisomorphic (Emmerich, 1971) as well as linear change functions.

In the present study only about half the interaction items studied showed a discernable linear pattern of developmental change. Even for these items, followed through primary school to date in only the oldest project cohort, more complex change patterns may be revealed over the longer term. Several characteristics which showed curvilinear change here seem related to those so described in the Wright (1975b, 1976), White and Watts (1973) and/or Educational Testing Service (Bridgeman & Shipman, 1975) studies, e.g., pleasure in exploration without adult attention or other reward, leadership, use of peers as resources. Other items which showed tentative quadratic trend here are not so described in the literature reviewed, e.g., original use of materials, persistence. Among the nonlinear items, several appeared to discriminate high achievers by not declining with age following early increases as occurred in the majority of the group, e.g., exploration, leadership, task completion. These items which showed quadratic patterns for all but the top students suggest, as have many popular critiques of education, that there may be something in the conventional school situation which begins to discourage interactive characteristics such as curiosity and initiative after a certain age-grade level.

Perhaps further studies similar to the present one, limited to those relevant items showing quadratic trend, would offer a valuable objective focus on such allegations.

For the present study, speculative interpretation on longitudinal interactive change must be limited to the Grosse Ile program alone, and even to particular individuals with the program. For instance, an individual pattern of change which differs from the group pattern might provide clues to a low achiever's attempt to cope and compensate in the school situation. The lowest child in Group 3 provided an example of such an idiosyncratic change in his unique late rise in affection to peers and reproduction of dialogue. Continued follow-up may determine whether these changes are stable and are positively or negatively related to any achievement gains in the next few years. Such individual-focussed as well as group-focussed study in the area of interaction characteristics is advocated for educators and specialists working with exceptional children in the schools. Even among a non-exceptional preschool group, Wright (1976) documents wide individual differences in the ways in which gains are made in all areas of development.

Among those interactive characteristics which to date have shown an overall linear pattern of developmental change in project children, several were among those most distinctive of highest and/or lowest achievers. These are of interest to future research, early childhood programming and educational individualization since they suggest an accelerated or lagging development in these characteristics on the part of high or low achievers respectively. In the area of physical world interactions, variation in the rate

of developing an exploratory and independent approach to things and situations, also seen by Miller and Dyer (1975) to improve in a linear fashion (inventiveness and curiosity), seemed to distinguish both achievement extremes. The use of materials in more structured and directed ways, on the other hand, appeared to develop linearly, according to both teacher ratings and their anecdotal comments, but did not show such distinctive variation between high and low achievers.

In the area of social interactions, group cooperation showed definite increases with age and experience as in Wright's (1976) study. This positive linear change was particularly noted by parents and teachers of the third summer preschoolers as well as in overall rating patterns, and was discriminative of both achievement extremes. Effective use of adults as resources also increased over time (with some fluctuation), at an apparently above average rate for high achievers and below average rate for lowest achievers. While the majority of project children quickly decreased certain negative social interactions with peers and adults early in their school experience (e.g., failure to follow adults, resistance of peers), low achievers seemed to maintain more of these behaviours, again as seen by Miller and Dyer (1975; e.g., aggression). It is noteworthy that initially low-ranked children who were gaining in achievement level also maintained certain social behaviours which might often be considered negative, e.g., manipulation of peers and adults and an orientation towards adult attention. The need for very specific descriptions of interaction characteristics in future research, and particularly in any application of findings to ECE

programming and teacher training, is apparent in the distinctions seen here.

The use of language in interactions increased rapidly in a linear fashion during the early stages of school experience according to both teachers' ratings and their anecdotal comments in all three years. Low achievers apparently failed to develop in verbal interaction at a pace characteristic of the group as a whole, but high achievers were not accelerated in this respect.

It was initially assumed that changes in interaction characteristics would be discernible earlier than in development-achievement characteristics. While it is true that certain high discrimination interaction characteristics showed change very early after program entrance, e.g., language use and social resistance, and that to date certain development-achievement characteristics have shown maximum change only several years after preschool-kindergarten experience, e.g., nonverbal IQ and academic achievement, the situation is much too complex for any generalization across characteristics. As seen in just the three interaction characteristics subjected to predictive correlational analysis, one (item 7, language use) predicted post-grade 1 achievement best from post-SPS rating while the other two items (19, leadership; 30, flexibility) predicted best from post-SKG rating. Case study profiles of the highest achievers in the two largest groups suggest that certain other interaction characteristics change coincidentally with or even following changes in achievement level, e.g., marked increase (two rating steps) in independent selection of activities by the Group 2 high achiever following grade 1, more than one year after his dramatic development-achievement gains relative to the group (see Appendix D).

If interaction characteristics are carefully defined and considered individually, the assumption that interactive changes are among the earliest detectable effects of ECE intervention does appear valid for selected characteristics. All three items analysed ~~predictively~~, for example, did at some preprimary stage predict post-grade 1 achievement with higher correlations than did developmental or achievement measures. A number of characteristics were found to predict highest achievers (e.g., standing ground in novel situations, flexibility), most positively changing (e.g., experimental exploration and finding exploration rewarding) and lowest achievers (e.g., little enjoyment of language interactions, high level of solitary play) from the first rating occasion alone. Supportive of these findings indicating that certain interactive change does precede achievement change is Bridgeman and Shipman's (1975) finding that preschool task orientation ratings were more useful as predictors of grade 3 than of grade 1 achievement. In the future more predictive correlational studies are suggested, perhaps with larger samples which would make possible analyses using strength of association indices for single characteristics over various time spans (lagged and cross-lagged analyses). Such designs would allow for study of curvilinear association between interaction characteristics and development-achievement variables, not merely the linear correlations imposed here by small group sizes.

One observation which should be followed up in any further study is that, for all three items analysed predictively, correlation with post-grade 1 achievement was higher at the end than at the

beginning of preschool. One questions whether children quickly developed in the interaction characteristics involved (flexibility, leadership, language use) during their first exposure to an educational program, and/or whether teachers required familiarization time to distinguish variations among children in these interactive characteristics. The latter factor is suspected on the basis of numerous psychometric methodologists' (Brandt, 1972, 1973; Nunnally, 1967) emphasis on knowing subjects well in order to obtain reliable rating results. Omissions declined and interrater agreement rose in the present study as teachers gained more exposure to the children -- both more daily time (e.g., observers versus full-time teachers) and more years with the children. The alternative explanations for these changes in omission and agreement levels, increasing age of children and/or increasing rating experience of some teachers (E.D. Evans, 1974), also need further examination. The possible importance of a teacher experience variable is suggested by the one instance of high omissions by a full time teacher (B). Teacher B herself and other project personnel describe her as less oriented to the observation and rating of children than other teacher-raters in the project. While response and interrater agreement levels were generally adequate throughout, factors relating to the observed improvement of these levels during the course of the project merit attention.

Even more than with development-achievement characteristics, it is impossible to attribute longitudinal changes in the interactive characteristics of project groups to specifically maturational cultural-environmental or program factors. Little developmental

data concerning the nature and norms of child-environment interaction characteristics are available, nor are any regional or local comparative data. What can be considered evidence of program effect, either direct effects of program participation and/or indirect effects via changes in the community, are the clear differences across cohorts at preschool, kindergarten and grade 1 levels in the two project years when interaction characteristics were rated. For many characteristics these differences from cohort to cohort were substantial enough to preclude general observations concerning the timing of developmental change in the population as a whole.

Each yearly increment in the preschool-kindergarten program (increased participation and community exposure time) did appear to further accelerate the early development of several interaction characteristics which (a) showed essentially linear change with age and experience, and (b) were highly discriminating of high and low achieving individuals. The program as experienced by the latest cohort to enter preschool (Group 4) showed such positive impact on a greater range of interaction characteristics than did the program as experienced by the cohort entering preschool one year earlier (Group 3). This latest cohort exceeded children at the same age in preceding cohorts in characteristics incorporating all three interaction elements being considered here, e.g., flexibility, social manipulation and attention to adult praise, enjoyment of verbal interactions. The latest cohort was also ahead of its predecessors on several high discrimination interaction characteristics less clearly linear in overall pattern, e.g., standing ground in novel situations, experimental exploration, solitary play. It is

of note that standing ground in novel situations, the most discriminative and predictive interactive characteristic studied, showed little longitudinal change within groups until its sharp early increase in the latest cohort.

A lesser number of achievement-related characteristics showed positive differences between both first and second and second and third year entrants into the program, e.g., group cooperation, use of adults, use of language in interactions. Specifically social and verbal interaction characteristics, rather than interactions with the overall physical-situational environment, showed this pattern of earlier and continuing program-related change. These characteristics were apparently affected sooner by lesser degrees of direct intervention and/or community diffusion. Among these discriminative characteristics which showed linear patterns of change and which increased across all three successive groups of preschool entrants are several items also prominent in teacher and parent comments concerning positive changes observed in their children, e.g., group cooperation and use of language. Increased group cooperation was also particularly noted by Wright (1976), and by Miller and Dyer (1975) along with verbal participation, as an early and lasting ECE program effect. In several other interactive areas showing high correlation with achievement, both in teacher ratings and according to parent and teacher anecdotal observations, no improvement across successive cohorts was noted. These apparently unaffected but developmentally significant interaction characteristics include certain aspects of independence (activity selection, leadership) and interactive complexity (dual focussing, following

sequential directions). These characteristics are predominantly physical-situational rather than social or verbal interactions, and should receive further careful study regarding their importance to school success and their resistance to change.

An important program effect of the preschool component alone is suggested by teacher observation in the first project year before interaction ratings were introduced. Preschool entrants had apparently not developed certain negative interaction characteristics which the year older kindergarten children had acquired. These negative characteristics involved all three elements of child-environment interaction: low initiative and task orientation, low social cooperation and high aggression, and limited use of language in interaction. Numerous other investigators of disadvantaged populations have described a progressive alienation of children from school culture preceding their entrance into educational programs (e.g., Beller, 1973; Gray & Klaus, 1969). Based on the similarity of her high- and low-SES preschool children in many interactive characteristics, Wright (1975a) suggests that this often-reported alienation, especially regarding interaction with teachers, must occur after preschool or after 4 years of age. This observation was supported during the first summer in the Grosse Ile situation where, despite parental apprehension regarding preschool particularly, the 4-year-old preschoolers were more positive in their school setting interactions than were the 5-year-old kindergarten children who anticipated entering grade 1 in several weeks. Whether fully attributable to an alienation factor or to other direct program effects of preschool, there are several achievement

related interaction characteristics in which the first cohort, with kindergarten but no preschool experience, have continued to show more negative interactive behaviour than subsequent cohorts, e.g., refusing to follow peers.

Conclusions

The purpose of this study has been to explore some more adequate concepts and more appropriate methods for ECE and its study, and so particularly to generate hypotheses for future local action research in ECE. The conclusions of this study are presented as recommended starting points for future context specific ECE programming and evaluation. Eight major conclusions with significant research implications are proposed.

1. Progress was made in the identification of child-environment interaction variables particularly related to school success. The relationship of early interaction characteristics to subsequent development and school learning is perhaps the most significant conclusion of the study with respect to future directions in ECE. Interaction with the physical environment appeared of prime importance, especially interactions with novel and complex situations. Interactions with the social environment, both peers and adults, followed in discriminative and predictive importance vis-à-vis achievement. Use of language in interactions seemed significant at the low but not at the high achievement extreme. Widely differing rating results from superficially similar items demonstrated the importance of precise definitions for interaction characteristics. In future

research, carefully described interaction characteristics should be studied individually for their relationship to high, low and changing achievement levels.

2. The expected sequential relationship of interactive changes preceding development-achievement changes was confirmed for certain specific variables only, e.g., language use in interactions, experimental exploration. Interaction characteristics varied widely in the form and speed of changes observed, including certain quadratic patterns which should be pursued for possible developmental and/or educational significance (e.g., decreasing self-direction during primary grades). Development-achievement characteristics also varied widely in the age and time span during which changes occurred, from immediate preprimary gains (e.g., visual-motor and factual learning) to later and continuing grade school gains (e.g., reading achievement and specific preschool effects on IQ). Future studies should be designed to detect such nonlinear and sometimes very slowly manifest (lagged and cross-lagged) patterns of longitudinal change and of longitudinal relationships among interactive and developmental characteristics.

3. The initial assumption that a distinct pattern of development-achievement strengths and weaknesses would be found in this relatively isolated population was supported. A need is implied for much further work concerned with developmental assessment in situations of cultural diversity. Such work towards culturally-specific developmental functions is of immediate relevance in Canada (and especially in Quebec, Cram, 1977) as ECE and its evaluation are attempted in distinct and relatively isolated cultural

contexts. The work must be multifactorial and interdisciplinary to avoid lasting biases concerning categories and levels of analysis for development-achievement variables, and must include more contextual and procedural data than was possible in the present study.

4. The initial assumption that this population would be characterized by specifically verbal relative to nonverbal learning difficulty was largely rejected, and the assumption of auditory relative to visual weakness was refined and modified during the course of the project. Three alternative models for the development-achievement characterization of this population, and for further theory-based ECE evaluation, are suggested by the pattern of development-achievement results. These models all focus on higher-order development and learning difficulties than were initially considered in this project, as follows:

(a) contextualized application or use of both verbal and nonverbal abilities, as distinct from skill acquisition per se, involving those cognitive-affective processes of child-environment interaction which gradually displaced language development as the Grosse Ile program goal;

(b) developmental lateralization of perceptual-motor and language abilities in relation to academic learning, and particularly in relation to the perceptual and higher-order cognitive processes of reading; and

(c) abstracting, integrating and sequencing of environmental input, especially as these processes involve the auditory mode and temporal ordering.

5. In contrast to initial staff assumption, but corresponding to recent observations in larger scale studies, a wide range of individual differences was observed within this group of children

in both (a) level and strength-weakness patterns of development-achievement characteristics and (b) interactive characteristics and their patterns of change over time relative to development-achievement changes. The range and complexity of these variations demonstrates the value of continuing to include descriptive case studies in future ECE evaluations, particularly case profiles of outstandingly successful children in apparently disadvantaged settings. Some degree of individualized assessment and programming, as well as evaluation, appeared necessary for effective ECE even in a small and superficially homogeneous community (e.g., exceptional low language cases here).

6. A considerable weight of evidence for positive program effects on both interactive and development-achievement characteristics accumulated during the course of the project, including evidence for the increasing effectiveness of each yearly increment in the program (i.e., increases in community involvement) and for the effect of the preschool program component alone. Direct longitudinal change appeared particularly prompt and substantial in those areas minimally implicated in the postulated basic learning difficulties of the Grosse Ile children. The combining of cross-sectional with longitudinal analysis was valuable in detecting program effects here, and is advocated in the common longitudinal research situation where adequate prototypic developmental functions (control or comparison data) are lacking. Of particular importance for future longitudinal study regarding ECE program effects are those developmental areas apparently significant in limiting the achievement of many Grosse Ile children but resistant to intervention as provided here, e.g., verbal sequential memory, content-oriented language use, and certain physical world interactive characteristics

such as independent selection of activities and dual focussing of attention.

7. The anticipated indirect community diffusion effect of the SPS-SKG program was clearly seen in several development-achievement (e.g., verbal intelligence, visual motor skill) and interaction (e.g., flexibility, experimental exploration) areas. An extension and perhaps more formalization of community involvement in such small-scale ECE projects is suggested by these results. The collection of data concerning community and specifically parent and teacher change in future longitudinal evaluations would be valuable in understanding and maximizing this indirect ECE effect.

8. The above conclusions and recommendations support a final conclusion that local action research, adopting many elements of a naturalistic paradigm, can produce valuable information and germane hypotheses for further study. Many naturalistic research elements, particularly as developed in descriptive or exploratory evaluation research, characterized this study:

(a) minimal manipulation of, intentional responsiveness to and progressive involvement of the research context; i.e., balancing of service and research functions, inclusion of individual-diagnostic and personnel training functions;

(b) broad and multifactored outcome measurement in terms of both content, i.e., developmental-learning domains; and technique, i.e., development and use of an observational rating measure (including demonstration of reliability) as well as formal testing;

(c) multiple analytical approaches, i.e., case study anecdotal description, summary descriptive statistics, nonparametric significance testing, multiple regression;

(d) small, local program and research setting oriented to a large, ongoing stream of research, involving result interpretation based on accumulated comparison data rather than on experimental control.

These naturalistic elements provided an appropriate methodology for this ECE study in making possible research which was relatively contextualized, differentiated, individualized and sequential. It is suggested that such naturalistic action research, as opposed to experimentation or criteria-outcome judgemental evaluations, remains a much needed step in the development of adequate conceptualizations for ECE and its study..

REFERENCES

- Abelson, W.D., & Zigler, E. Effects of a four-year follow through program on economically disadvantaged children. Journal of Educational Psychology, 1974, 66, 756-777.
- Alkin, M.C. Evaluation theory development. Evaluation Comment, 1969, 2, 1.
- Alkin, M.C. & Kosecoff, J.B. A new eclectic model for the redirection of evaluation efforts. Los Angeles, California: Center for the Study of Evaluation, University of California, 1973. (ERIC Document Reproduction Service No. ED 075 086).
- Almy, M. Commentary on J. Stallings "Implementation and child effects of teaching practices in Follow Through classrooms". Monographs of the Society for Research in Child Development, 1975, 40, 120-123.
- Anastasi, A. Psychological Testing. London: McMillan, 1968.
- Anderson, S.B. Circus manual and technical report (preliminary version). Princeton, N.J. : Educational Testing Service, 1974-1975.
- Anderson, S.B. & Messick, S. Social competency in young children. Developmental Psychology, 1974, 10, 282-293.
- Angrist, S.S. & Borke, H. An observational system for the evaluation of preschool programs : research and development. Urban Education, 1974, 9, 284-298.
- Appleford, B. & Ryan, T.J. An examination of social processes in a preschool. Professional Journal of the Association for Early Childhood Education of Ontario, in press.
- Armington, D.E. A plan for continuing growth. In U.S. Dept. of Health, Education & Welfare, Office of Education, Descriptions of follow-through programs. Washington, D.C.: Author, 1968.
- Averch, H.A., Carroll, S.J., Donaldson, T.S., Kiesling, H.J., & Pincus, J. How effective is schooling? A critical review and synthesis of research findings. Santa Monica, California : Rand Corporation, 1971.
- Bakker, D.J. Temporal order perception and reading retardation. In D.J. Bakker & P. Satz (Eds), Specific reading disability: Advances in theory and method. Rotterdam: Rotterdam University Press, 1970.

- Bakker, D.J. & Satz, P (Eds.) Specific reading disability: Advances in theory and method. Rotterdam: Rotterdam University Press, 1970.
- Banta, T.J. Autonomy in an absurd society. University of Denver Magazine, 1964, 2, 10-15.
- Banta, T.J. Is there really a Montessori method? Cincinnati: University of Cincinnati, 1966a.
- Banta, T.J. Educating children for adulthood. Young Children, 1966b, 21, 272-280.
- Banta, T.J. Two voices of psychology and the new psychologist. Ohio Psychologist, 1967, 13, 6-11.
- Banta, T.J. Tests for the evaluation of early childhood education: The Cincinnati Autonomy Test Battery (CTAB). In J. Hellmuth (Ed.), Cognitive Studies (Vol. 1). New York: Bruner/Magel, 1970.
- Banta, T.J. Montessori: Myth or reality. In R.K. Parker (Ed.), The preschool in action. Boston: Allyn and Bacon, 1972.
- Barker, R.G. Ecological psychology: Methods and concepts for an eco-behavioral science. Stanford, California: Stanford University Press, 1968.
- Barth, R.S. Open education. In Theory and philosophy of education. Kensington, Australia, 1969.
- Baumrind, D., & Black, A.E. Socialization practices associated with dimensions of competence in preschool boys and girls. Child Development, 1967, 38, 291-327.
- Baumrind, D. The contributions of the family to the development of competence in children. Schizophrenia Bulletin (U.S. Dept. of Health, Education & Welfare), Fall, 1975, pp. 12-37.
- Becker, W.C. & Krug, R.S. A circumplex model for social behavior in children. Child Development, 1964, 35, 371-396.
- Behar, L. & Stringfield, S. A behavior rating scale for the preschool child. Developmental Psychology, 1974, 10, 601-610.

- Beller, E.K. Dependency and independence in young children. Journal of Genetic Psychology, 1955, 87, 25-35.
- Beller, E.K. Teaching styles and their effects on problem-solving behavior in Headstart programs. In E. Grotberg (Ed.), Critical issues in research related to disadvantaged children. Princeton, N.J. : Educational Testing Service, 1969.
- Beller, E.K. Impact of early education on disadvantaged children. In S. Ryan (Ed.), A report on longitudinal evaluation of preschool programs. Washington, D.C. : Office of Child Development, 1972.
- Beller, E.K. Research on organized programs of early education. In R.M. Travers (Ed.), Second handbook of research on teaching. Chicago : Rand McNally, 1973.
- Bennet, E.L., Diamond, M., Krech, D., & Rosenzweig, M.R. Chemical and anatomical plasticity of the brain. Science, 1964, 146, 610-619.
- Bentley, R.J., Washington, E.D., & Young, J.C. Judging the educational progress of young children: Some cautions. Young Children, 1973, 29, 5-18.
- Bereiter, C. An academic preschool for disadvantaged children: conclusions from evaluation studies. In J.C. Stanley (Ed.), Preschool programs for the disadvantaged. Baltimore: Johns Hopkins University Press, 1972.
- Bereiter, C., & Englemann, S. Teaching disadvantaged children in the preschool. Englewood Cliffs, N.J. : Prentice Hall, 1966.
- Bernstein, D.A. A sociolinguistic approach to socialization with some reference to educability. In F. Williams (Ed.), Language and poverty. Chicago : Markham, 1970.
- Bernstein, I & Freeman, H.E. Academic and entrepreneurial research : Consequences of diversity in federal evaluation studies. New York : Russell Sage Foundation, 1974.
- Bertalanffy, L.C. General systems theory: A critical review. General Systems, 1962, 7, 1-20.

Biber, B. A learning-teaching paradigm integrating intellectual and affective processes. In E.B. Bower & W.C. Hollister (Eds.), Behavioral science frontiers in education. New York: Wiley, 1967.

Biber, B., & Franklin, M.B. The relevance of developmental and psychodynamic concepts to the education of the preschool child. Journal of the American Academy of Child Psychiatry, 1967, 6, 5-24.

Biber, H., Miller, L.B. & Dyer, J.L. Feminization in preschool. Developmental Psychology, 1972, 7, 86.

Bijou, S.W. & Baer, D.M. Child development I: A systematic and empirical theory. New York: Appelton-Century-Crofts, 1961.

Bissell, J.S. Implementation of planned variations in Head Start. Washington, D.C.: U.S. Department of Health, Education & Welfare, 1971.

Bissell, J.S. Planned variation in Head Start and Follow Through. In J.C. Stanley (Ed.), Compensatory education for children ages two to eight. Baltimore: John Hopkins, 1973.

Black, M.S. The development of personality in children and adolescents. Educational and Psychological Measurement, 1965, 24, 767-785.

Blank, M. Implicit assumptions underlying preschool intervention programs. In B. Spodek, Early childhood education. Englewood Cliffs, N.J.: Prentice-Hall, 1973a.

Blank, M. Teaching learning. Columbus, Ohio: Merrill, 1973b.

Blatz, W.E. Human security: Some reflections. Toronto: University of Toronto Press, 1966.

Bloom, B.S. (Ed.). Taxonomy of educational objectives: The classification of educational goals. Handbook I. Cognitive domain. New York: Macmillan, 1929.

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

Bloom, B.S., Hastings, J.T. & Madaus, G.T. Handbook on formative and summative evaluation of student learning. New York: McGraw Hill, 1971.

Borowitz, G.H., Costello, J., & Hirsch, J.G. Clinical observations of ghetto four-year-olds: Organization, involvement, interpersonal responsiveness, and psychosexual content of play. Journal of Youth and Adolescence, 1972, 1, 59-79.

- Bortner, M. & Birch, H.G. Cognitive capacity and cognitive competence. American Journal of Mental Deficiency, 1970, 74, 735-744.
- Brandt, R.M. Studying behavior in natural settings. New York: Holt, Rinehart & Winston, 1972.
- Brandt, R.M. Observational methodology for evaluation of early childhood programs. Journal of Research and Development in Education, 1973, 6, 94-109.
- Bridgeman, B & Shipman, V.C. Disadvantaged children and their first school experiences: Predictive value of measures of self-esteem and achievement motivation in four- to nine-year-old low income children. Princeton, N.J.: Educational Testing Service, 1975.
- Bronfenbrenner, U. Early deprivation in mammals and man. In G. Newton (Ed.), Early experience and behavior. Springfield, Ill.: Thomas, 1968.
- Bronfenbrenner, U. Is early intervention effective? In E.L. Struening & M. Guttentag (Eds.), Handbook of evaluation research. Beverly Hills, California: Jones, 1975.
- Bronfenbrenner, U. The experimental ecology of education. Teachers College Record, 1976, 78 (2), 152-204.
- Bruner, J.S. Towards a theory of instruction. Cambridge, Mass.: Harvard University Press, 1966.
- Bruner, J.S. The nature and uses of immaturity. American Psychologist, 1972a, 27, 687-708.
- Bruner, J.S. Poverty and childhood. In R.K. Parker (Ed.), The preschool in action. Boston: Allyn & Bacon, 1972b.
- Brunswik, E. Representative design and probabilistic theory in a functional psychology. Psychological Review, 1955, 62, 193-217.
- Bureau of Head Start and Early Education. Review of research: 1965 to 1969. Washington, D.C.: Department of Health, Education & Welfare, Office of Child Development, 1969.
- Burgess, E. Values in early childhood education (2nd ed.). Washington, D.C.: National Education Association, 1965.
- Bushell, D. Behavior analysis: a research approach to follow-through. Washington, D.C.: Dept. of Health, Education & Welfare, 1969.

- Busse, T.V., Ree, M. & Gertride, M. Environmentally enriched classrooms and the play behaviors of Negro preschool children. Urban Education, 1970, 5, 128-140.
- Bussis, A. & Chittenden, E. The horizontal dimension of learning. In A. Tobier (Ed.), Evaluation reconsidered. New York: Workshop Center for Open Education, 1973.
- Caldwell; B.M. A new approach to behavioural ecology. In J.P. Hill (Ed.), Minnesota symposia on child psychology (Vol. 2). Minneapolis: University of Minnesota Press, 1969.
- Caldwell, B.M. & Smith, L.E. Day care for the very young - prime opportunity for primary prevention. American Journal of Public Health, 1970, 60, 690-697.
- Camp, J.C. A skill development curriculum for 3-, 4-, & 5-year-old disadvantaged children. In B. Spodek, Early childhood education. Englewood Cliffs, N.J.: Prentice-Hall, 1973.
- Campbell, D.T. From description to experimentation: Interpreting trends as quasi-experiments. In C.W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1967.
- Campbell, D.T. Reforms as experiments. American Psychologist, 1969, 24, 4.
- Campbell, D.T. Critical problems in the evaluation of social programs. Paper prepared for the Annual Meeting of the Division of Behavioral Sciences, National Academy of Sciences, 1972.
- Campbell, D.T. & Erlebacher, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), The disadvantaged child (Vol. 3). New York: Bruner/Magel, 1970.
- Campbell, D.T. & Frey, P.W. The implications of learning theory for the fade-out of gains from compensatory education. In J. Hellmuth (Ed.), Disadvantaged child (Vol. 3). New York: Bruner/Magel, 1970.
- Campbell, D.T. & Stanley, J.C. Experimental and quasi-experimental designs for research. Chicago: Rand-McNally, 1963.
- Carini, P.F. The Prospect School: Taking account of progress. Childhood Education, 1973, 49, 350.

- Carini, P.F. Observation and description: An alternative methodology for the investigation of human phenomena. Grand Forks, N.D.: North Dakota Study Group in Evaluation, University of North Dakota Press, 1975.
- Carter, N. & Wharf, B. Evaluating social development programs. Ottawa: Canadian Council on Social Development, 1973.
- Carver, R.P. Special problems in measuring change with psychometric devices. In American Institutes for Research, Evaluative research: Strategies and models. Pittsburgh: Author, 1970.
- Carver, R.P. The case against statistical significance testing. Harvard Educational Review, 1978, 48 (3), 378-399.
- Case, R. Social class differences in intellectual development: A Neo-Piagetian investigation. Canadian Journal of Behavioural Science, 1975, 1, 244-260.
- Cass, J.E. & Gardner, D.E. The role of the teacher in the infant school. Oxford: Pergamon, 1965.
- Cazden, C. Functions of language in the classroom. New York: Teachers College Press, 1972.
- Central Advisory Council for Education (England). Children and their primary schools (The Plowden report). London: Her Majesty's Stationary Office, 1967.
- Chazan, M. The use of tests in the evaluation of preschool educational programs. In Council of Europe (Documentation Center for Education in Europe), Problems in the evaluation of pre-school education. Strasbourg: Author, 1975. (ERIC Document Reproduction Service No. 117 158).
- Cherns, A. Social research and its diffusion. Human Relations, 1969, 22, 3.
- Cicchetti, D.V. A new measure of agreement between rank-ordered variables. Proceedings of the 80th Annual Convention of the American Psychological Association, 1972, 7, 17-18.
- Cicirelli, V.G. The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development. Westinghouse Learning Corporation & Ohio University, 1969.
- Cicirelli, V.G., Evans, J.W. & Schiller, J.S. The impact of Head Start: A reply to the report analysis. Harvard Educational Review, 1970, 40 (1), 105-129.
- Clark, B. The Brunswick-Cornwallis Pre-School: A program for disadvantaged white and negro children. Child Study, 1967, 29, 17-27.

- Cohen, D.K. Politics and research: Evaluation of social action programs in education. Review of Educational Research, 1970, 40, 213-238.
- Cohen, S. The problem with Piaget's child. Teachers College Record, 1966, 68, 211-218.
- Cole, M. & Bruner, J.S. Preliminaries to a theory of cultural difference. In I.J. Gordon (Ed.), Early childhood education: The 71st yearbook of the National Society for the Study of Education (Part 2). Chicago: University of Chicago Press, 1972.
- Coleman, J.S. Equality of educational opportunity. Washington D.C.: U.S. Dept. of Health, Education & Welfare, Office of Education, 1966.
- Coller, A.R. Systems for the observation of classroom behavior in early childhood education. Urbana, Ill.: ERIC Clearinghouse on Early Childhood Education, 1972.
- Colligan, R.C. Prediction of kindergarten reading success from preschool report of parents. Psychology in the Schools, 1976, 13, 304.
- Cook, T.D. & Campbell, D.T. The design and conduct of quasi-experiments and true experiments in field settings. In M.D. Dunnette (Ed.), Handbook of industrial and organizational research. In press.
- Coopersmith, S. The antecedents of self-esteem. San Francisco: Freeman, 1968.
- Council of Europe (Documentation Center for Education in Europe). Problems in the evaluation of preschool education. Strasbourg: Author, 1975. (ERIC Document Reproduction Service No. 117 158).
- Cram, J.M. Educational interpretation of the Wechsler Intelligence Scale for Children for remote and isolated schools in Quebec. Montreal: Faculty of Education, McGill University, 1974.
- Cram, J.M. Summer learning centre, Magdalen Islands, P.Q., 1974-1976: Final report. Montreal: Faculty of Education, McGill University, 1977.
- Cronbach, L.J. Course improvement through evaluation. Teachers College Record, 1963, 64, 672-683.

- Cronbach, L.J. Heredity, environment and educational policy. Harvard Educational Review, 1969, 39, 338-347.
- Cronbach, L.J. & Snow, R.E. Aptitudes and instructional methods: A handbook for research on interactions. New York: Irvington, 1977.
- Cronbach, L.J. & Snow, R.E. Final report: Individual differences in learning ability as a function of instructional variables. Stanford, California: Stanford University Press, 1969.
- Dennis, W. Goodenough scores, art experience, and modernization. Journal of Social Psychology, 1966, 68, 211-228.
- Deutsch, C.P. Auditory discrimination and learning: Social factors. Merrill-Palmer Quarterly, 1964, 10, 277-296.
- Deutsch, M. Facilitating development in the preschool child: Social and psychological perspectives. Merrill-Palmer Quarterly, 1964, 10, 249-263.
- Deutsch, M. Regional research and resource center in early childhood: Final report. Washington, D.C.: U.S. Office of Economic Opportunity, 1971.
- Dewey, J. Experience and education. New York: Collier, 1963. (Originally published, 1938).
- Dewey, J. The school and society. Chicago: University of Chicago Press, 1965a.
- Dewey, J. The child and the curriculum. Chicago: University of Chicago Press, 1965b.
- DiLorenzo, L.T. Prekindergarten programs for educationally disadvantaged children (Final report). University of the State of New York, New York State Education Department, Office of Research and Evaluation, 1969.
- Drugokinski, E., Weiss, S. & Johnston, S. Preschoolers at risk: social, emotional and cognitive considerations. Psychology in the Schools, 1976, 13, 134-138.
- Doehring, D.G. The evaluation of two models of reading disability. Montreal: School of Human Communication Disorders, McGill University, 1975.
- Douglas, J. & Douglas, H. Balancing the taxonomies. Audio-visual Instruction, 1972, 17, 7-8.
- Duckworth, E. Evaluation of the African Primary Science Program. Newton, Mass.: Education Development Center, 1976.

Dyer, J.L. & Miller, L.B. Note on Crano, Kenny, and Campbell's "Does intelligence cause achievement?". Journal of Educational Psychology, 1974, 66, 49-51.

Educational Research Service. Summary of research on prekindergarten programs. Arlington, Virginia: Author, 1976.

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

Educational Testing Service. Young children and their first school experiences: A brief description of the ETS - Head Start Longitudinal Study. Princeton, N.J.: Author, 1976.

Educational Testing Service. Instruments used in ETS - Head Start Longitudinal Study of young children and their first school experiences for 1969-1974. Princeton, N.J.: Author, no date.

Edwards, W., Guttentag, M. & Snapper, K. A decision-theoretic approach to evaluation research. In E.L. Stuenkel & M. Guttentag (Eds.), Handbook of evaluation research. Beverly Hills, California: Sage, 1975.

Eisner, E.W. Emerging models for educational evaluation. School Review, 1972, 80, 575-590.

Eisner, E.W. English primary schools: Some observations and assessments. Washington, D.C.: National Association for the Education of Young Children, 1974.

Elkind, D. Piaget and Montessori. Harvard Educational Review, 1967, 37, 4.

Elkind, D. Children and adolescents: Interpretive essays on Jean Piaget. New York: Oxford University Press, 1970.

Elkind, D. Preschool education: enrichment or instruction. In B. Spodek (Ed.), Early childhood education. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

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.

Emmerich, W. Disadvantaged children and their first school experiences: Preschool personal-social behaviors - relationships with socioeconomic status, cognitive skills, and tempo. Princeton, N.J.: Educational Testing Service, 1973.

- Engel, B.S. A handbook on documentation. Grand Forks, N.D.: North Dakota Study Group on Evaluation, University of North Dakota Press, 1975.
- Engin, A.W., Leppaluoto, J.R. & Petty, S.Z. Multifactorial unbiased assessment: A guide for school psychologists. School Psychology Digest, 1977, 6, 42-47.
- Erikson, E.L., McMillan, J., Bennett, J., Callahan, O.D. Experiments in Head Start and early education: Curriculum structures and teacher attitudes. Washington, D.C.: Office of Economic Opportunity, Project Head Start, 1969.
- Ervin-Tripp, S.M. Children's sociolinguistic competence and dialect diversity. In I.J. Gordon (Ed.), Early childhood education: The 71st yearbook of the National Society for the Study of Education (Part 2). Chicago: University of Chicago Press, 1972.
- Evans, E.D. Contemporary influences in early childhood education. New York: Holt, Rinehart & Winston, 1971.
- Evans, E.D. Measurement practices in early childhood education. In R.W. Colvin & E.M. Zaffiro, Preschool education: A handbook for the training of early childhood educators. New York: Springer, 1974.
- Evans, E.G.S. Modern educational psychology: An historical introduction. London: Routledge & Kegan Paul, 1969.
- Evans, J.W. Head Start: Comments on the criticisms. In D.G. Hays (Ed.), Britannica Review of American Education (Vol. 1). Chicago: Encyclopedia Britannica, 1969.
- Evans, R. Nursery education: **Need** for research. Child Welfare, 1975, 54, 495.
- Ezrin, S. An operational evaluation of a preschool for Canadian disadvantaged children. Unpublished Master's thesis, Carleton University, 1971.
- Fairweather, G.W. Methods for experimental social innovation. New York: Wiley, 1967.
- Featherstone, H.J. Cognitive effects of preschool programs on different types of children (Doctoral dissertation, Harvard University, 1973). Dissertation Abstracts International, 1974, 34, 7100A.
- Filstead, W.J. (Ed.), Qualitative methodology. Chicago: Markham, 1970.

Fitzgibbon, C.T. & Morris, L.L. Theory-based evaluation.
Evaluation Comment, 1975, 5, 1.

Flapan, D., Gunn, G. & Neubauer, P.B. Sources of information
for developmental assessment. Psychosocial Process, 1970,
2, 98-115.

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

Forness, S.R., Guthrie, D. & Hall, R.J. Follow-up of high risk
children identified in kindergarten through direct class-
room observation. Psychology in the Schools, 1976, 13, 45-49.

Fowler, W. Cognitive learning in infancy and childhood.
Psychological Bulletin, 1962, 59 (2), 116-152.

Gagné, R.M. Instruction and the conditions of learning.
In L. Siegel (Ed.), Instruction: Some contemporary view-
points. San Francisco: Chandler, 1967.

Gagné, R.M. Contributions of learning to human development.
Psychological Review, 1968, 75, 177-191.

Gibson, E.J. & Levin, H. The psychology of reading. Cambridge,
Mass.: MIT Press, 1975.

Gilmer, B., Miller, J.O. & Gray, S.W. Intervention with mothers
and young children: Study of intra-family effects. Nashville,
Tenn.: DARCEE Demonstration and Research Center for Early
Education, 1970.

Glaser, B.G. & Strauss, A.L. Discovery of substantive theory:
A basic strategy underlying qualitative research. In W.J.
Filstead (Ed.), Qualitative methodology. Chicago: Markham,
1970.

Goodwin, W.L. Evaluation in early childhood education. In
N.W. Golvin & E.M. Zaffiro, Preschool education: A handbook
for the training of early childhood educators. New York:
Springer, 1974.

Gordon, I.J. An instructional theory approach to the analysis
of selected early childhood programs. In I.J. Gordon (Ed.),
Early childhood education: The 71st yearbook of the National
Society for the Study of Education (Part 2). Chicago:
University of Chicago Press, 1972.

Gordon, I.J. An early intervention project: A longitudinal
look. Gainesville, Fla.: University of Florida, Institute
for Development of Human Resources, College of Education, 1973.

- Gordon, I.J. & Jester, R.E. Techniques of observing teaching in early childhood and outcomes of particular procedures. In R.M. Travers (Ed.), Second handbook of research on teaching. Chicago: Rand McNally, 1973.
- Gotts, E.E. Head Start research, development and evaluation. In J.L. Frost, Revisiting early childhood education. New York: Holt, Rinehart & Winston, 1973.
- Gray, S.W. & Klaus, R.A. Experimental preschool program for culturally-deprived children. Child Development, 1965, 36, 887-898.
- Gray, S.W. & Klaus, R. The early training project: A 7th year report. Nashville, Tenn.: JFK Center for Research on Education and Human Development, George Peabody College, 1969.
- deGroot, Adriaan. Methodology: Foundations of inference and research in the behavioural sciences. Hague: Mouton, 1969.
- Grosse, M.E. Effectance behavior in preschool children. (Doctoral dissertation, University of Connecticut, 1972). Dissertation Abstracts International, 1972, 32, 6679B.
- Guba, E.G. The failure of educational evaluation. Educational Technology, 1969, 9, 5.
- Gump, P.V. Ecological psychology and children. Chicago: University of Chicago Press, 1975.
- Guralnick, M.J. A research-service model for support of handicapped children. Exceptional Children, 1973, 39, 274-282.
- Guttentag, M. Models and methods in evaluation research. Journal of the Theory of Social Behavior, 1971, 1, 1.
- Halasa, O. Evaluation of preschool education. Educational Technology, 1972, 12, 38-40.
- Hammond, R.L. Context evaluation of instruction in local school districts. Educational Technology, 1969, 9, 1.
- Harris, C.W. (Ed.), Problems in measuring change. Madison, Wisconsin: University of Wisconsin Press, 1967.
- Harris, D.B. (Ed.), The concept of development. Minneapolis: University of Minnesota Press, 1957.
- Harris, M.B. & Wagoner, P.S. Predicting future school success through the use of a preschool developmental rating scale: A longitudinal study. Training School Bulletin, 1973, 70, 172-178..

- Hawkrige, D.H. Designs for evaluative studies. In American Institutes for Research, Evaluative research: Strategies and models. Pittsburgh: Author, 1970.
- Hays, W.L. Statistics for psychologists. New York: Holt, Rinehart & Winston, 1963.
- Hebb, D.O. The organization of behavior. New York: Wiley, 1949.
- Hechinger, F.M. (Ed.), Pre-school education today. Garden City, N.Y.: Doubleday, 1966.
- Hein, G.E. An open education perspective on evaluation. Grand Forks, N.D.: North Dakota Study Group on Evaluation, University of North Dakota Press, 1975.
- Hellmuth, J. Cognitive studies (Vol. 1). New York: Bruner/Magel, 1970.
- Hertzog, M.E., Birch, H.G., Thomas, A. & Mendez, O.A. Class and ethnic differences in the responsiveness of preschool children to cognitive demands. Monographs of the Society for Research in Child Development, 1969, 33, 1.
- Herzog, E., Newcomb, C. & Cisin, I. But some are more poor than others: SES differences in a preschool program. American Journal of Orthopsychiatry, 1972, 42, 4-22.
- Hess, R. & Baer, R. (Eds.), Early education. Chicago: Aldine Press, 1968.
- Hess, R.D., Kramer, D., Slaughter, D., Torney, J., Berry, D., & Hull, E. Techniques for assessing cognitive and social abilities of children and parents in Project Head Start. Chicago: University of Chicago, 1966. (ERIC Document Reproduction Service No. ED 015 772).
- Hess, R.D. & Shipman, V. Early experience and socialization of cognitive modes in children. Child Development, 1965, 36, 869-886.
- Hewett, F.M. Educational engineering with emotionally disturbed children. Exceptional Children, 1967, 33, 459-467.
- Hodge, W.L., McCandless, B.R. & Spicker, H.H. The development and evaluation of a diagnostically based curriculum for pre-school psychosocially deprived children: Final Report. Bloomington, Ind.: Indiana University, 1967. (ERIC Document Reproduction Service No. ED 021 948).

- Hodges, W.L. The worth of the Follow Through experience. Harvard Educational Review, 1978, 48 (2), 186-192.
- Hoge, R.D. & Luce, S. Predicting academic achievement from classroom behaviour. Ottawa: Department of Psychology, Carleton University, no date.
- Horowitz, F.O. & Padin, L.Y. The effectiveness of environmental interventional programs. In B.M. Caldwell & H.N. Ricciuti (Eds.), Review of child development research (Vol. 3), Chicago: University of Chicago Press, 1973.
- House, E.R., Glass, G.V., McLean, L.D. & Decker, F.W. No simple answer: Critique of the follow through evaluation. Harvard Educational Review, 1978, 48 (2), 128-160.
- Huberty, C.J. & Swan, W.W. Preschool experience and first grade achievement. Journal of Educational Research, 1974, 67, 311-316.
- Hunt, D.E. Person-environment interaction: A challenge found wanting before it was tried. Review of Educational Research, 1975, 45, 209-230.
- Hunt, J.McV. Intelligence and experience. New York: Ronald Press, 1961.
- Hunt, J.McV. The psychological basis for using preschool enrichment as an antidote for cultural deprivation. Merrill-Palmer Quarterly, 1964, 10, 209-248.
- Hunt, J.McV. Early childhood education and social class. Canadian Psychologist, 1972, 13, 305-328.
- Hyman, H.H.* & Wright, C.R. Evaluating social action programs. In R.F. Lazarsfeld (Ed.), The uses of sociology. New York: Basic Books, 1967.
- Isaacs, N. The growth of understanding in the young child. London: Ward Lock, 1961.
- Isaacs, N. Piaget: Some answers to teachers questions. London: Froebel Foundation, 1965.
- Isaacs, S. Intellectual growth in young children. London: Routledge & Kegan Paul, 1930.
- Isaacs, S. Social development in young children. London: Routledge & Kegan Paul, 1937..
- Isaacs, S. The children we teach. London: University of London Press, 1963.

Jackson, D.N. & Messick, S. (Eds.), Problems in human assessment. New York: McGraw Hill, 1967.

Jacobson, L.I., Berger, S.E., Bergman, R.L., Millham, J. & Greeson, L.E. Effects of age, sex, systematic conceptual learning, acquisition of learning sets, and programmed social interaction on the intellectual and conceptual development of preschool children from poverty backgrounds, Child Development, 1971, 42, 1399-1415.

Jencks, C. Inequality: A reassessment of the effect of family and schooling in America. New York: Basic Books, 1972.

Jensen, A.R. How much can we boost I.Q. and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.

Johnson, G.H. The purpose of evaluation and the role of the evaluator. In American Institutes for Research, Evaluative research: Strategies and methods. Pittsburgh: Author, 1970.

Jones, N.B. (Ed.), Ethological studies of child behaviour. Cambridge: Cambridge University Press, 1972.

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

Kamii, C. Evaluation of learning in preschool education: Socio-emotional, perceptual-motor, cognitive development. In B.S. Bloom, J.T. Hastings & G.F. Madaus, Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill, 1971.

Kamii, C. An application of Piaget's theory to the conceptualization of a preschool curriculum. In R.K. Parker (Ed.), The preschool in action. Boston: Allyn & Bacon, 1972.

Kamii, C.K. & Radin, N.L. A framework for a preschool curriculum based on Piaget's theory. In I.J. Athey & D.O. Rubadeau (Eds.), Educational implications of Piaget's theory. Waltham, Mass.: Ginn-Blaisdell, 1970.

Karnes, M.B. Research and development program on preschool disadvantaged children: Final report. Washington, D.C.: Office of Education, 1969.

Karnes, M.B., Hodgins, A.S., Teska, J.A. & Kirk, S.A. Research and development program on preschool disadvantaged children (Vol. 1). Urbana, Ill.: Institute for Research on Exceptional Children, 1969.

- Karnes, M.B., Zehrbach, R.R. & Teska, J.A. An ameliorative approach in the development of curriculum. In R.K. Parker (Ed.), The preschool in action. Boston: Allyn & Bacon, 1972.
- Kerlinger, R.N. & Pedhazur, E.J. Multiple regression in behavioral research. New York: Holt, Rinehart & Winston, 1973.
- Kirk, S.A. Early education of the mentally retarded. Urbana, Ill.: University of Illinois Press, 1958.
- Klaus, R. & Gray, S. The early training project for disadvantaged children: A report after five years. Monographs of the Society for Research in Child Development, 1968, 33.
- Klein, S.P. Ongoing evaluation of educational programs. Paper presented at the American Psychological Association convention, Honolulu, 1972. (ERIC Document Reproduction Service No. ED 069 725).
- Koep, R.G. The effect of social and cognitive interaction strategies on children's motivation to achieve in school (Doctoral dissertation, University of Oregon, 1973). Dissertation Abstracts International, 1973, 34, 3144A.
- Kohlberg, L. Montessori with the culturally disadvantaged: A cognitive-developmental interpretation and some research findings. In R.D. Hess & R.M. Baer (Eds.), Early education. Chicago: Aldine, 1966.
- Kohlberg, L. Early education: A cognitive-developmental view. Child Development, 1968, 39, 1013-62.
- Kohlberg, L. & Mayer, R. Development as the aim of education. Harvard Educational Review, 1972, 42, 4.
- Kohlberg, L. & Turiel, E. Moral development and moral education. In G. Lesser (Ed.), Psychology and educational practice. Chicago: Scott Foresman, 1971.
- Kohn, M. & Cohen, J. Emotional impairment and achievement deficit in disadvantaged children - fact or myth? Genetic Psychology Monographs, 1975, 92, 57-78.
- Kohn, M. & Rosman, B.L. A social competence scale and symptom checklist for the preschool child: factor dimensions, their cross-instrument generality and longitudinal persistence. Development Psychology, 1972a, 6, 430-444.
- Kohn, M. & Rosman, B.L. Relationship of preschool social-emotional functioning to later intellectual achievement. Developmental Psychology, 1972b, 6, 445-452.

Krathwohl, D.R., Bloom, B.S. & Masia, B.B. Taxonomy of educational objectives: The classification of educational goals. Handbook II. Affective domain. New York: McKay, 1964..

Kratochwill, T.R. N = 1: An alternative research strategy for school psychologists. Journal of School Psychology, 1977, 15, 239-249.

Kuzma, K.J. & Stern, C. The effects of three preschool intervention programs on the development of autonomy in Mexican-American and Negro children. Journal of Special Education, 1972, 6, 197-205.

Labov, W. The logic of nonstandard English. In F. Williams (Ed.), Language and poverty. Chicago: Markham, 1970.

Langer, J. Theories of development. New York: Holt, Rinehart & Winston, 1959.

Lavatelli, C.S. Early childhood curriculum: A Piaget program. Boston: American Science & Engineering, 1970.

Lay, M.Z. The Responsive Care model. Syracuse: Early Childhood Education Center, Syracuse University, 1972.

Levenstein, P. Cognitive growth in preschoolers through verbal interaction with mothers. American Journal of Orthopsychiatry, 1970, 40, 426-432.

Levine, M. Scientific method and the adversary model. American Psychologist, 1974, 29, 9.

Little, B.R. & Ryan, T.J. Children in context: The social ecology of human development. Ottawa: Department of Psychology, Carleton University (Research essay prepared for National Health and Welfare Canada, Policy Research and Long Range Planning Branch), 1978.

Lofland, J. Analyzing social settings. Belmont, California: Wadsworth, 1971.

Maw, W.H. A definition of curiosity: A factor analysis study. Newark, Del.: University of Delaware, 1967. (ERIC Document Reproduction Service No. ED 012 382).

Maxwell, A.E. Analyzing qualitative data. London: Methuen, 1961.

McCandless, B.R. Children: Behavior and development. New York: Holt, Rinehart & Winston, 1967.

- McCall, R.B. Smiling and vocalization in infants as indices of perceptual-cognitive processes. Merrill-Palmer Quarterly, 1972, 18, 342-347.
- McClelland, D.C. Testing for competence rather than for intelligence. In E. Flaxman (Ed.), Educating the disadvantaged (1972-1973). New York: AMS Press, 1976.
- McMillan, M. The nursery school (Rev. Ed.). London: J.M. Dent, 1930.
- McNemar, Q. A critical examination of the University of Iowa studies of environmental influences upon the I.Q. Psychological Bulletin, 1940, 37, 63-92.
- McPhee, S. Canada: Basic issues identified. In Centre for Educational Research and Innovation (CERI), Developments in early childhood education. Paris: Organization for Economic Cooperation and Development (OECD), 1975.
- Medley, D.M. OSCAR goes to nursery school: A new technique for recording pupil behavior. Princeton, N.J.: Educational Testing Service, 1969.
- Mendelsohn, R.L. Headstart - success or failure. Clinical Pediatrics, 1969, 8.
- Messick, S. Evaluation of educational programs as research on educational process. Paper presented to American Psychological Association, Washington, D.C., 1969.
- Messick, S & Barrows, T.S. Strategies for research and evaluation in early childhood education. In I.J. Gordon (Ed.), Early childhood education: The 71st yearbook of the National Society for the study of Education. Chicago: University of Chicago Press, 1972.
- Miller, J.O. & Camp, J.C. Towards individual competency - a curriculum in the child's ecology. In R.K. Parker (Ed.), The preschool in action. Boston: Allyn & Bacon, 1972.
- Miller, L.B. & Dyer, J.L. Experimental variation of Head Start curricula: A comparison of current approaches (Annual progress report, 1969-70). Louisville, Kentucky: Child Development Laboratory of Louisville, 1970. (ERIC Document Reproduction Service No. ED 045 196).
- Miller, L.B. & Dyer, J.L. Four preschool programs: Their dimensions and effects. Monographs of the Society for Research in Child Development, 1975, 40, 5 & 6 (No. 162).

Minuchin, P. Correlates of curiosity and exploratory behavior in preschool disadvantaged children. Child Development, 1971, 42, 291-293.

Mirante, T.J. & Ryckman, D.B. Classroom Behavior Inventory: Factor verification. Journal of Research in Personality, 1974, 8, 291-293.

Moffitt, A.R. Schools without curricula: Evaluation of some early childhood programs. Paper presented to the Canadian Psychological Association, June, 1964.

Moncada, C.R. A longitudinal observational study of the behavior of low-income preschool children at home and in school compared with the behavior of similar but non-schooled children at home (Doctoral dissertation, University of Minnesota, 1972). Dissertation Abstracts International, 1973, 33, 5558A.

Montessori, M. [The Montessori Method: Scientific pedagogy as applied to child education in "The Children's Houses" (Rev. ed.)] (A.E. George, Trans. with Introduction by J. McV. Hunt). New York: Schocken Books, 1964. (Originally published, 1909.)

Moriarity, A. Coping patterns of preschool children in response to intelligence demands. Genetic Psychology Monographs, 1961, 64, 3-127.

Murphy, L.B. The widening world of childhood. Chicago: Basic Books, 1962.

Murphy, L.B. & Moriarity, A.E. Vulnerability, coping and growth. New Haven: Yale University Press, 1976.

Nedler, S. A development process approach to curriculum design. In R.K. Parker (Ed.), The preschool in action. Boston: Allyn & Bacon, 1972.

Nesselroade, J.R. & Reese, H.W. (Eds.), Life-span developmental psychology: Methodological issues. New York: Academic Press, 1973.

Nimnicht, G.P. A model program for young children that responds to the child. In R.K. Parker (Ed.), The preschool in action. Boston: Allyn & Bacon, 1972.

Northway, M. Child study in Canada: A casual history. In L. Brockman, J. Whiteley & J. Zubrek (Eds.), Child development: Selected readings. Toronto: McClelland & Stewart, 1973.

Nunnally, J. Psychometric theory. New York: McGraw-Hill, 1967.

Nunnally, J. Research strategies and measurement methods for investigating human development. In J.R. Nesselroade & H.W. Reese (Eds.), Life-span developmental psychology. New York: Academic, 1972.

Nunnally, J. The study of change in evaluation research. In E.L. Struening & M. Guttentag, (Eds.), Handbook of evaluation research. Beverly Hills, California: Sage, 1975.

Nuthall, G.A. A review of some selected recent studies of classroom interaction and teaching behavior. In J.J. Gallagher, G.A. Nuthall & B. Rosenshine (Eds.), Classroom observation (AERA Monograph Series on Curriculum Evaluation). Chicago: Rand McNally, 1970.

O'Connor, E.F. Extending classical test theory for the measurement of change. Los Angeles, California: Center for the Study of Evaluation, University of California, 1970. (ERIC Document Reproduction Service No. ED 050 175).

Painter, G. Infant education. San Rafael, California: Dimensions, 1968.

Parker, R.K. Theory in early education curricula. In R.W. Colvin & E.M. Zaffiro (Eds.), Preschool education: A handbook for the training of early childhood educators. New York: Springer, 1974.

Parker, R.K. & Day, M.C. Comparisons of preschool curricula. In R.K. Parker (Ed.), The preschool in action. Boston: Allyn & Bacon, 1972.

Parlett, M. & Hamilton, D. Evaluation as illumination: A new approach to the study of innovatory programs (Occasional Paper No. 9). Edinburgh: Center for Research in the Educational Sciences, University of Edinburgh, 1972.

Patton, M.Q. Alternative evaluation research paradigm. Grand Forks, N.D.: North Dakota Study Group on Evaluation, University of North Dakota, 1975.

Payne, J.S., Mercer, C.D., Payne, R.A. & Davison, R.G. Head Start: A tragicomedy with epilogue. New York: Behavioral Publications, 1973.

Peterson, D.R. The age generality of personality factors derived from ratings. Educational and Psychological Measurement, 1960, 20, 461-474.

Piaget, J. Psychologie et pédagogie. Paris: Denoël, 1969.

Piaget, J. & Inhelder, B. The psychology of the child. New York: Basic, 1969.

Pick, L.L. An exploratory study of curiosity in young children. (Doctoral dissertation, Columbia University, 1974). Dissertation Abstracts International, 1974, 34, 7053A.

Pines, M. Revolution in learning. New York: Harper & Row, 1967.

Plutchick, R., Platman, S.R. & Fieve, R.R. Three alternatives to the double blind. Archives of General Psychiatry, 1969, 20, 429-432.

Polanyi, M. Logic and psychology. American Psychologist, 1968, 23, 1.

Provus, M. Evaluation of ongoing programs in the public school system. In R.W. Tyler (Ed.), Educational evaluation - new roles, new means: The 68th yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press, 1969.

Radin, N. The impact of a kindergarten home counselling program. Exceptional Children, 1969, 36, 251-256.

Rambusch, N.M. Learning how to learn: An American approach to Montessori (A Montessori bibliography of materials in the English language, 1909-1961, compiled by G.E. Donahue). Baltimore: Helicon Press, 1962.

Rand Corporation. Design for a national evaluation of social competence in Head Start children. Santa Monica, California: Author, 1974. (ERIC Document Reproduction Service No. ED 104 569).

Rardin, D.R. A study of peer interaction and its effect upon cognitive development. (Doctoral dissertation, University of Florida, 1971.) Dissertation Abstracts International, 1971, 32, 569-B.

Raths, J.D. Teaching without specific objectives. Educational Leadership, 1971, 28, 714-720.

Reese, J.W. & Overton, W.F. Models of development and theories of development. In Goulet, L.R. & Baltes, P.B. (Eds.), Life-span developmental psychology: Research and theory. New York: Academic Press, 1970.

Reitan, R.M. & Dawson, L.A. (Eds.), Clinical neuropsychology: Current status and applications. Washington, D.C.: Winston, 1974.

Rentfrow, R., Durning, D., Conrad, E. & Goldupp, O. Use of **three** new instruments in a Head Start program evaluation. Psychology in the Schools, 1975, 12, 34-39.

Resnick, L.B. Programmed instruction and the teaching of social studies skills. The 33rd Yearbook of the National Council for Social Studies, 1963, 33, 252-273.

Resnick, L.B. Design of an early learning curriculum. Pittsburgh: University of Pittsburgh Learning Research and Development Center, 1967.

Riecken, H.W., Boruch, R.F., Campbell, D.T., Caplan, N., Glennan, T.K., Pratt, J., Rees, A. & Williams, W. Experimentation as a method for planning and evaluating social innovations. New York: Seminar Press, 1975.

Rotter, J.B. Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs: General and Applied, 1966, 80, 609.

Rousseau, J.J. Emile, or education (B. Foxley, trans.). London: Dent, 1966. (Originally published, 1762.).

Rubenstein, G. & Fisher, L. The development of a measure of teachers' observations of student behavior. Rochester, N.Y.: University of Rochester Medical School, 1974.

Rusk, B. An evaluation of a six week Head Start program using an academically oriented curriculum. Toronto: Ontario Institute for Studies in Education, 1968.

Ryan, T.J. Poverty and the child: A Canadian study. Toronto: McGraw-Hill Ryerson, 1972.

Ryan, T.J. Promoting child development through a program of home visiting. Canadian Journal of Behavioural Sciences, 1976, 8 (1), 102-105.

Ryan, T.J. & Moffitt, A.R. Evaluation of preschool programs. The Canadian Psychologist, 1974, 15, 3.

Sameroff, A.J. Early influences on development: fact or fancy? Merrill-Palmer Quarterly, 1975, 21, 267-294.

Samsky, J.J. Social and non-social exploration in preschool children (Doctoral dissertation, New York University, 1973). Dissertation Abstracts International, 1973, 1262B.

- Sansone, G.F. The development of measures of autonomy among elementary school children (Doctoral dissertation, Boston University, 1974). Dissertation Abstracts International, 1974, 35, 1513A.
- Satz, P. & Van Nostrand, G.K. Developmental dyslexia: An evaluation of a theory. In P. Satz & J. Ross (Eds.), The disabled learner: An evaluation of a theory. Rotterdam: Rotterdam University Press, 1973.
- Schaefer, D., Baker, E. & Zawel, D. A factor analytic and reliability study of the Devereux Elementary School Behavior Rating Scale. Psychology in the Schools, 1975, 12, 295-300.
- Schaefer, E.S. Development of hierarchical and configurational models for parent behavior and child behavior. In J.E. Hill, (Ed.), Minnesota symposium on child psychology (Vol. 5). Minneapolis: University of Minnesota Press, 1971.
- Schaefer, E.S. Major replicated dimensions of adjustment and achievement: Cross-cultural, cross-sectional and longitudinal research. Paper presented at the American Educational Research Association annual meeting (Symposium on Dimensions of Competence in Classrooms), Washington, D.C., 1975.
- Schaefer, E.S. & Aaronson, M. Classroom behavior inventory (Pre-standardization short-form; Revised 1/76). Unpublished form and technical information, 1976.
- Schaie, K.W. A general model for the study of developmental problems. Psychological Bulletin, 1965, 64, 92-107.
- Schoggen, M. & Schoggen, P. Environmental forces in home lives of three-year-old children in three population subgroups (DARCEE Papers and Reports, 5, 2). Nashville, Tenn.: George Peabody College for Teachers, 1971.
- Schulberg, H.C. & Baker, T. Program evaluation models and the implementation of research findings. American Journal of Public Health, 1968, 58.
- Scriven, M. The methodology of evaluation. American Educational Research Association Monograph Series on Evaluation (No. 1). Chicago: Rand McNally, 1967.
- Scriven, M. Evaluating educational programs. The Urban Review, 1969, 3, 4.

- Scriven, M. Objectivity and subjectivity in educational research. In L.G. Thomas (Ed.), The 71st yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press, 1972.
- Sears, P.S. & Dowley, E.M. Research on teaching in the nursery school. In N.L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.
- Semmes, J. Hemispheric specialization: a possible clue to mechanisms. Neuropsychologia, 1968, 6, 11-26.
- Shapiro, E. Educational evaluation: Rethinking the criteria of competence. School Review, November, 1973, 523-549.
- Shapiro, E. & Biber, B. The education of young children: A developmental-interaction approach. Teachers College Record, 1972, 74, 55-79.
- Sharp, J.D. & Stott, D.H. Stott-Sharp effectiveness motivation scale. Windsor, Berks. England: National Foundation for Educational Research, 1976.
- Shipman, V.C. Disadvantaged children and their first school experiences: Structural stability and change in the test performance of urban preschool children. Princeton, N.J.: Educational Testing Service, 1972.
- Shipman, V.C. Disadvantaged children and their first school experiences: Notable early characteristics of high and low achieving black low SES children. Princeton, N.J.: Educational Testing Service, 1976.
- Siegel, S. Non-parametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.
- Sigel, I.E. The Piagetian system and the work of education. In D. Elkind & J.H. Flavell (Eds.), Studies in cognitive development: Essays in honor of Jean Piaget. New York: University Press, 1969.
- Sigel, I.E. Where is preschool education going: Or are we en route without a roadmap? Proceedings of the Invitational Conference on Testing Problems (Educational Testing Service), 1972a, 99-116.
- Sigel, I.E. Developmental theory and preschool education: Issues, problems and implications. In I.J. Gordon (Ed.), Early childhood education: The 71st yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press, 1972b.

Sigel, I.E. When do we know what a child knows? Human Development, 1974, 17, 201-217.

Sigel, I.E. The search for validity or the evaluator's nightmare. In R.A. Weinberg & S.G. Moore (Eds.), Evaluation of educational programming for young children: The Minnesota Roundtable in Early Childhood Education. Washington, D.C.: Child Development Associates Consortium, 1975.

Siever, R. Science: Observational, experimental, historical. American Scientist, 1968, 56, 70-77.

Silvernail, J.M. The design and validation of an instrument for identifying and measuring the coping behaviors of three-year-olds in preschool settings (Doctoral dissertation, University of Pittsburgh, 1974). Dissertation Abstracts International, 1974, 35, 1487A.

Skeels, H.M. Adult status of children with contrasting early life experiences. Monographs of the Society for Research in Child Development, 1966, 31, 3 (No. 105).

Skinner, B.F. Science and human behavior. New York: Macmillan, 1961.

Skinner, B.F. The technology of teaching. New York: Appleton-Century-Crofts, 1968.

Smilansky, S. The effects of socio-dramatic play on disadvantaged preschool children. New York: Wiley, 1968.

Smith, M.S. & Bissell, J.S. Report analysis: The impact of Head Start. Harvard Educational Review, 1970, 40, 51-104.

Soar, R.S. & Soar, R.M. Pupil subject matter growth during the summer vacation. Educational Leadership Research Supplement, 1969, 2, 577-587.

Soar, R.S. & Soar, R.M. An empirical analysis of selected follow-through programs: An example of a process approach to evaluation. In I.J. Gordon (Ed.), Early childhood education: The 71st yearbook of the National Society for the Study of Education (Part 2). Chicago: University of Chicago Press, 1972.

Southern, M.L. & Plant, W.T. Effects of cognitively structured small group activities upon cognitive functions in pre-first-grade programs for culturally disadvantaged children. Journal of Genetic Psychology, 1972, 121, 11-20.

Spaulding, R.L. Educational intervention in early childhood, Vols. I, II, III. Durham, N.C.: Duke University, 1971.

Sprigle, H. Inquisitive games: Discovering how to learn. Chicago: Science Research Associates, 1969.

Sprigle, H. The learning to learn program. Jacksonville, Florida: Learning to Learn School, Inc., 1970.

Sprigle, H., Van de Reit, V. & Van de Reit. A sequential learning program for preschool children and evaluation of its effectiveness with culturally disadvantaged children. Paper presented at the annual meeting of the American Educational Research Association, New York, 1967.

Stake, R.E. The countenance of educational evaluation. Teachers College Record, 1967, 68, 523-540.

Stake, R.E. Objectives, priorities and other judgement data. Review of Educational Research, 1970, 40, 171-212.

Stake, R.E. & Denny, T. Needed concepts and techniques for utilizing more fully the potential of evaluation. In R.W. Tyler (Ed.), The 68th yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press, 1969.

Stallings, J. Implementation and child effects of teaching practices in Follow Through classrooms. Monographs of the Society for Research in Child Development, 1975, 40, 7 & 8.

Stanley, J.C. & Hopkins, K.D. Educational and psychological measurement and evaluation. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Steinmitz, A. The ideology of educational evaluation. Educational Technology, 1975, 15, 51-58.

Stern, C. Evaluating language curricula for preschool children. Monographs of the Society for Research in Child Development 1968, 33, 49-61.

Stodolsky, S.S. Defining treatment and outcome in early childhood education. In H.J. Walberg & A.T. Kopan (Eds.), Rethinking urban education. San Francisco: Jossey-Bass, 1972.

Stott, D.H. An empirical approach to motivation based on the behavior of a young child. Journal of Child Psychology and Psychiatry, 1961, 2, 97-117.

Stott, D.H. & Albin, J.B. Confirmation of a general factor of effectiveness-motivation by individual tests. British Journal of Educational Psychology, 1975, 45, 153-161.

Stott, D.H., Williams, H.L. & Sharp, J.D. A general factor termed effectiveness motivation in the behavior of pre-school children. Educational Research, 1974, 18, 2.

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, 10.

Strike, K. Explaining and understanding: the impact of science on our concept of man. In L.G. Thomas (Ed.), The 71st yearbook of the National Society for the Study of Education (Part 1). Chicago: University of Chicago Press, 1972.

Strodtbeck, F.L. The hidden curriculum of the middle class home. In C.W. Hunnecutt (Ed.), Urban education and cultural deprivation. Syracuse: Syracuse University Press, 1964.

Stufflebeam, D.L. An introduction to the Phi Delta Kappa book "Educational evaluation and decision-making". Address at the 11th Annual Phi Delta Kappa Symposium on Educational Research, Ohio State University, 1970.

Stufflebeam, D.L., Foley, W.J., Gephart, W.J., Buga, E.G., Hammond, R.L., Merriman, H.O. & Provus, M.M. Educational evaluation and decision-making in education. Itasca, Ill.: Peacock, 1971.

Suchman, E.A. Evaluating educational programs. The Urban Review, 1969, 3, 4.

Suchman, E.A. Action for what? A critique of evaluative research. In R.O'Toole (Ed.), The organization, management and tactics of social research. Cambridge, Mass.: Schenkman, 1970.

Thirton, A. Evaluation of action research. In Council of Europe (Documentation Center for Education in Europe). Problems in the evaluation of preschool education. Strasbourg: Author, 1975. (ERIC Document Reproduction Service No. ED 117 158).

Tinbergen, N. The study of instinct. Oxford: Oxford University Press, 1951.

Tinsley, H.E.A. & Weiss, D.J. Interrater reliability and agreement of subjective judgments. Journal of Counselling Psychology, 1975, 22, 358-374.

Torrence, E.P. Education and the creative potential. Minneapolis: University of Minnesota Press, 1963.

Tripodi, T., Fellin, P. & Epstein, I. Social program evaluation. Itasca, Ill.: F.E. Peacock, 1971.

Van Alstyne, D. Play behavior and choice of play materials of preschool children. Chicago: University of Chicago Press, 1932.

Vane, J.R. Problems in and strategies for evaluating preschool programs, Journal of School Psychology, 1976, 14, 39-45.

Van de Riet, V. & Resnick, M.B. Learning to learn: An effective model for early childhood education. JSAS Catalog of Selected Documents in Psychology, 1976, 6, 48-49. (Ms. No. 1255)

Walker, D.K. Socioemotional measures for preschool and kindergarten children. San Francisco: Jossey-Bass, 1973.

Washington, W.O. Toward a theory of social competence: Implications for measuring the effects of Head Start programs. Urban Education, 1975, 10, 73-85.

Weber, L. The English infant school and informal education. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Webb, E.J., Campbell, D.T., Schwartz, R.D. & Sechrest, L. Unobtrusive measures. Chicago: Rand McNally, 1971.

Weikart, D.P., Deloria, D.J., Lawson, S.A. & Wiegerink, R. Longitudinal results of the Ypsilanti Perry Preschool Project. Ypsilanti, Mich.: High/Scope Educational Research Foundation, 1970.

Weikart, D. The cognitively oriented curriculum. Urbana, Ill.: ERIC - National Association for the Education of Young Children, 1971a.

Weikart, D.P. Early childhood special education for intellectually subnormal and/or culturally different children. Paper prepared for the National Leadership Institute in Early Childhood Development, Washington D.C., 1971b.

Weikart, D.P. Organizational scheme for preschool curriculum models. Ypsilanti, Mich.: High/Scope Educational Research Foundation, 1971c.

Weikart, D.P. Relationship of curriculum, teaching and learning in preschool education. In J.C. Stanley (Ed.), Preschool programs for the disadvantaged. Baltimore: Johns Hopkins University Press, 1972.

Weikart, D.P., Kamii, C.K. & Radin, M. Perry Preschool progress report. Ypsilanti, Mich.: Ypsilanti Public Schools, 1964.

Weiss, C.H. Evaluating action programs. Boston: Allyn & Bacon, 1972a.

Weiss, C.H. Evaluation research. Englewood Cliffs, N.J.: Prentice-Hall, 1972b.

Weiss, C.H. Utilization of evaluation research: Toward comparative study. In C.H. Weiss, Evaluating action programs. Boston: Allyn & Bacon, 1972c.

Weiss, R.S. & Rein, M. The evaluation of broad aim programs: A cautionary case and a moral. In C.H. Weiss, Evaluating action programs. Boston: Allyn & Bacon, 1972.

Wellman, B.L. I.Q. changes of preschool and nonpreschool groups during the preschool years: A summary of the literature. Journal of Psychology, 1945, 20, 347-368.

Wenar, C. Competence at one. Merrill-Palmer Quarterly, 1964, 10, 329-342.

Wenar, C. Executive competence and spontaneous social behavior in 1-year-olds. Child Development, 1972, 43, 256-260.

Werner, H. Comparative psychology of mental development. Chicago: Wilcox & Follett, 1948.

Werner, H. The concept of development from a comparative and organismic point of view. In D.B. Harris (Ed.), The concept of development. Minneapolis: University of Minnesota, 1957.

Whitbread, N. The evolution of the Nursery-Infant School. London: Routledge & Kegan Paul, 1972.

White, B.L. Informal education during the first months of life. In R.D. Hess & R.M. Baer (Eds.), Early education. Chicago: Aldine, 1968.

White, B.L. & Watts, J.C. Experience and environment (Vol. 1). Englewood Cliffs, N.J.: Prentice-Hall, 1973.

White, R.W. Motivation reconsidered: The concept of competence. Psychological Review, 1959, 66, 297-233.

44

OF / DE

Willems, E.P. Planning a rationale for naturalistic research. In E.P. Willems & H.L. Raush (Eds.), Naturalistic viewpoints in psychological research. New York: Holt, Rinehart & Winston, 1969.

Willems, E.P. & Raush, H.L. (Eds.), Naturalistic viewpoints in psychological research. New York: Holt, Rinehart, Winston, 1969.

-Willems, E.P. & Willems, G.J. Comparative validity of data yielded by three methods. Merrill-Palmer Quarterly, 1965, 11, 65-71.

Williams, W. & Evans, J.W. The politics of evaluation: The case of Head Start. The Annals of the American Academy of Political and Social Science, 1969, 385, 118-132.

Winnett, R.A. & Edwards, S.M. An evaluation plan for educational innovations. Journal of Community Psychology, 1974, 2, 345-351.

Wisler, C.E., Burns, G.P. & Iwamoto, D. Follow Through redux: A response to the critique by House, Glass, McLean and Walker. Harvard Educational Review, 1978, 48 (2), 171-185.

Wohlwill, J. The study of behavioral development. New York: Academic Press, 1973.

Wolff, M. & Stein, A. Head Start six months later. Phi Delta Kappa, March, 1967. (Reprinted in J. Frost (Ed.), Early childhood education rediscovered. New York: Holt, Rinehart, & Winston, 1968).

Wright, M.J. Should we rediscover Blatz? Was his psychology American, European or uniquely Canadian? Canadian Psychologist, 1974, 15, 140-144.

Wright, M.J. Competence in preschool children: II. Two years in the laboratory preschool. (Research Bulletin No. 348). London, Ontario: Department of Psychology, University of Western Ontario, 1975a.

Wright, M.J. Changes in the social competence of Canadian pre-school and day nursery children of low and high socio-economic status. Interchange, 1975b, 6, 4.

Wright, M.J. The laboratory preschool: program development in the first three years (Research Bulletin No. 355). London Ontario: Department of Psychology, University of Western Ontario, 1976.

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

Yawkey, T.D. & Silvern, S.B. Evaluating instructional goals of developmental programs in early childhood education. Education, 1975, 96, 36-39.

Zigler, E., Abelson, W.D. & Seitz, V. Motivational factors in the performance of economically disadvantaged children on the PPVT. In E. Flaxman (Ed.), Educating the disadvantaged. New York: AMS Press, 1976.

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

Zigmond, N.K. & Cieci, R. Auditory learning: Dimensions in early learning series. San Rafael, California: Dimensions, 1968.

APPENDIX A

Development-Achievement Measures

Title ^a	Reference	Area Assessed	Norms	Notes
Intellectual Development Measures	1. Peabody Picture Vocabulary Test	Dunn, L.N., Expanded manual for the Peabody Picture Vocabulary Test. Circle Pines, Minnesota: American Guidance Service, 1965.	Verbal intelligence	Standardized IQ of 4012, selected to represent U.S. population with an approximately normal distribution on a widely used group IQ test (Kuhlmann-Finch). Extensive reliability validity research reviewed in manual (pp 30-42) and elsewhere. Short term retest reliability in range of $r = .85$.

2. Goodenough-Harris Drawing Test	Harris, D.B., Children's Drawings as measures of intellectual maturity. New York: Harcourt, Brace & World, 1963.	Intellectual maturity (non-verbal measure of intelligence)	Standard scores from 3 years to 15 years; mean standard score at each age level (1 year intervals) = 100, standard deviation = 15.	Scale items validated with SES-stratified U.S. sample of 1000. Extensive subsequent reliability validity research reviewed in referenced book containing test manual (pp. 21-22, 90-99) and elsewhere. Retest reliability in range of $r = .70$; split-half and interscorer reliabilities in range of $r = .90$.
-----------------------------------	--	--	--	--

^a Numbering of measures corresponds to that used in Figure 1 (see Method).

Title	Reference	Area Assessed	Norms	Notes
Perceptual-Motor Abilities Measures				
3. Slingerland Prereading Screening Procedures (subtests 1 to 5)	Slingerland, B.H. Teacher's manual to accompany Pre-reading Screening Procedures. Cambridge, Massachusetts: Educators Publishing Service, 1968.	Visually-based skills of discrimination, memory, fine motor control and perceptual motor integration	Five-step rating (High, Medium +, Medium, Medium -, Low) of each subtest score for readiness level children (late kindergarten-early grade 1).	Subtest ratings established in relation to concurrent scores on widely used tests of readiness (Metropolitan and of IQ (Pintner-Cunningham). NB. Ratings for the five subtests used in this study were quantified from 1 to 5, thus deriving a low total score of 5 and a high total score of 25.
4. Slingerland Screening Tests for Identifying Children with Specific Language Disabilities: Level A (for grades 1 to 2)	Slingerland, B.H., Teacher's manual to accompany Slingerland Screening Tests. Cambridge, Massachusetts: Educators Publishing Service, 1969, 1970.	Academically-relevant visually-based (subtests 1 to 5) and auditory-based (subtests 6 to 8) skills of discrimination, memory, fine motor control and perceptual-motor integration.	A total error score of 15 on subtests 3 to 8 of appropriate level Slingerland Tests established as the breakoff point for identifying children with specific language disabilities. Error scores above the breakoff point identify the lowest quartile of the screened population.	Approximately 3000 suburban U.S. children included in breakoff point research; approximately 10,000 U.S. children currently involved in SES-stratified research.
5. Slingerland Screening Tests for Identifying Children with Specific Language Disabilities: Level B (for grades 2 to 3)				

(see measure 4 above)

Title	Reference	Area Assessed	Norms	Notes
6. Gross-Motor Observations	University City, Missouri. Re-printed in <u>Hand-book of pupil experiences</u> . Tulsa Public Schools, Oklahoma, 1973.	Locomotion, large muscle coordination and balance (walking, hopping, jumping, skipping)	Age equivalents from 3 years-11 months to 5 years-5 months (1 to 5 month intervals), for boys and girls separately	
7. Circus 7 (How Words Sound)	Anderson, S.B., and others, <u>Circus manual and technical report</u> . Princeton, New Jersey: Educational Testing Service, 1974.	Ability to discriminate among phonemes heard in meaningful words	Nursery school and kindergarten level means and standard deviations established for each Circus sub-test. Both levels of norms broken down by sex, race, SES, geographic region, and preschool vs. no preschool experience. (Item analyses also provided.)	Standardization sample of 300 nursery school and 644 kindergarten children, stratified by geographical region, city size, program sponsorship and ethnic group. Internal consistency and split half reliabilities in range of $r = .87$ to $.92$. NB. For the third summer assessment in the present study (1976), the project language specialist reduced the Circus 7 item pool by half (from 44 to 22) on the basis of local vocabulary and dialect.

Title	Reference	Area Assessed	Norms	Notes
8. Sentence Repetition	<p>Meckham, M.J., Rex, J.L. & Jones, J.D., Utah Test of Language De- velopment: Manual. Salt Lake City: Communication Research Asso- ciates, 1967; and Terman, L. M. & Merrill, M.A., Stanford- Binet Intelli- gence Scale. Boston: Houghton- Mifflin, 1960</p>	<p>Rote verbal auditory memory (12 syllable sentence)</p>	<p>Three errors (additions, omissions or substitutions) established as breakoff point indicating ser- ious verbal mem- ory difficulty at readiness level (late kinder- garten)</p>	<p>Readiness breakoff score based on diagnosis-as- sessment in several rural northeastern Quebec school regions, including Re- gional School Board of Gaspésia which extends to the Magdalen Islands. This measure adopted from the Stanford Binet Year 4 (Memory for sentences I: Alternate Test) and Utah Test of Language Develop- ment age equivalent for 4-5 years.</p>
9. Wechsler Intelligence Scale for Children (Digit Span sub-test)	<p>Wechsler, D., Wechsler Intel- ligence Scale for Children: Manual. New York: Psycho- logical Corpora- tion, 1949</p>	<p>Rote verbal auditory memory (digit series)</p>	<p>Age equivalents from 5 years-2 months to 15 years-10 months (4 month inter- vals)</p>	<p>Standardization sample of 200 children selected to represent the U.S. white population - stratified by geographical region, urban-rural-insti- tutional residence, and father's occupation - plus 2200 "feebleminded" cases. Digit Span reliability based on correlation be- tween digits forward and digits backward using Spearman-Brown formula: at 7 years $r = .88$, at 10.5 years $r = .96$, at 13 years $r = .96$.</p>

Title	Reference	Area Assessed	Norms	Technical-Procedural Notes
10. Basic Concepts Inventory (item 20 - digit repetition)	Engelmann, S. <u>Basic Concepts Inventory</u> . Chicago: Follett, 1967	Rote verbal auditory memory and pattern awareness (digit series)	Criterion-referenced measure for preschool-kindergarten level; no norms available	Reliability and predictive validity studies in progress (see reference, p.45).
11. Illinois Test of Psycholinguistic Abilities (Auditory Sequential Memory subtest).	University of Illinois, <u>ITPA Manual</u> . Urbana, Illinois: University of Illinois Press, 1969	Rote verbal auditory memory (digit series)	Psycholinguistic age for auditory sequential memory from 2 years to 10 years-3 months (6 month intervals)	Standardization sample of 962 children with in average range of IQ, school achievement personal-social adjustment, and sensory motor integrity; selected to approximate U.S. SES distribution. Extensive reliability-validity research reviewed in manual and elsewhere. Subtest internal consistency in range of $\bar{r} = .89$ (4 years-7 months to 5 years-1 month) to $\bar{r} = .95$ (9 years-7 months to 10 years-1 month); 5-month retest reliability from $\bar{r} = .75$ (4 years) to $\bar{r} = .89$ (8 years)
12. Tapping	Sheeran, N., Project language-speech-hearing specialist. Faculty of Education, McGill University, Montreal, Quebec	Nonverbal auditory memory (number and rhythm)	None established previous to this study	Tapping measure being developed by project language specialist.

Title	Reference	Area Assessed	Norms	Technical-Procedural Notes
Language Acquisition Measures				
13. Circus 8 (How Words Work)	(see measure 7 above)	Ability to understand sentences that emphasize structure and functional words (receptive language)	(see measure 7 above)	Standardization sample similar to measure 7. Internal consistency and split half reliabilities in range of $r = .71$ to $.81$. NB. For third summer assessment in the present study (1976), Circus 8 scores were prorated for two items excluded by the project language specialist on the basis of local dialect.
14. Metropolitan Readiness Test (Listening subtest)	Nurss, J.R. & McGauvran, M.E., Metropolitan Reading Test. New York: Harcourt, Brace & Jovanovich, 1974	Ability to recall, integrate and interpret the content of messages of language (comprehension)	No norms available for Listening subtest alone	
15. Circus 10-II (Say and Tell)	(see measure 7 above)	Ability to use appropriate forms of functional words (expressive language)	(see measure 7 above)	Standardization sample similar to measure 7. Internal consistency and split half reliabilities in range of $r = .89$ to $.93$. NB. For third summer assessment in the present study (1976), the project language specialist reduced Circus 10-II item pool from 38 (at 2 points each) to 25 (at 1 point each) on basis of local language usage.

Title	Reference	Area Assessed	Norms	Technical-Procedural Notes
Conceptual-Factual Learning and Academic Achievement Measures	16. Circus 2 (How Much and How Many)	(see measure 7 above)	Quantitative concepts (relational and numerical)	Standardization sample of 934 nursery school and 1883 kindergarten children, stratified as for measure 7 above. Internal consistency and split half reliabilities in range of $r = .86$ to $.88$. NB. For second summer assessment in the present study (1975), Circus 2 items containing digits were dropped from forms used with preschool children, thereby reducing item pool from 40 to 30.
17. Opposites	Sheeran, N., project language-speech-hearing specialist. Faculty of Education, McGill University, Montreal, Quebec	Verbal concepts	None established previous to this study	Opposites measure being developed by project language specialist
18. Circus 11 (Do You Know)	(see measure 7 above)	General knowledge of facts important to functioning in and out of school	(see measure 7 above)	Standardization sample similar to measure 7. Internal consistency and split half reliabilities in range of $r = .77$ to $.80$. NB. Circus 11 scores prorated throughout the present study for two items rejected on the basis of specifically U.S. content.

Title	Reference	Area Assessed	Norms	Notes
19. Circus 5 (Finding Letters and Numbers)	(see measure 7 above)	Ability to recognize and discriminate letters and numbers	(see measure 7 above) Norms provided for each of the three and split half reliability of the three subparts in range of $r = .79$ to $.80$ for capital letters, lower case letters, $r = .42$ to $.44$ for numbers, numbers	Standardization sample similar to measure 7. Internal consistency
20. Early Childhood Inventory (Identifications of Body Parts, Shapes, and Colours)	Coller, A. and Victor, J., <u>Early childhood inventories</u> project. Institute for Developmental Study, New York University, n.d.	Receptive (non-verbal) and expressive (verbal) identifications of body parts, shapes and colours	Criterion-referenced measure for preschool level; no norms available	used total of capital letter and number (but not lower case letter) scores (total of 14 rather than all 20 items). Therefore normative means; but not standard deviations, are available relevant to project Circus 5 scores.
21. Stanford Achievement Tests: Primary I Battery (for grades 1.5 to 2.4)	Kelley, T.L. and others, Stanford <u>Achievement Tests</u> , New York: Harcourt, Brace and World, 1964.	Academic skills - reading, language arts, arithmetic areas	Grade equivalents in decimal form corresponding to months for all grade-school levels. (Stanine and percentile derived scores also provided.)	Standardization sample of over 200,000 representative of the U.S. population in terms of the geographical region, city size, SES, and public vs. private schools. Normalized distribution on a widely used group IQ test (Otis-Lennon), administered concurrently with Stanford tests, used in establishing derived scores at each grade

Title	Reference	Area Assessed	Norms	Technical-Procedural Notes
22. Stanford Achievement Tests: Primary II Battery (for grades 2.5 to 3.9)	(see measure 21 above)	Academic skills - reading, language arts, arithmetic, and science areas	(see measure 21 above)	level. Split half reliabilities for Primary I Battery tests in range of $\bar{r} = .81$ to $.95$.
23. Slosson Oral Reading Test	Slosson, R.L., Slosson Oral Reading Test, East Aurora, New York: Slosson Educational Publications, 1963.	Ability to pronounce written words (mechanics of oral reading)	Reading level grade equivalents, in decimal form corresponding to months of the school year, for all grade school levels.	Norms based on standardized school readers For 108 grade 1-to-8 students, correlation with Gray Standardized Oral Reading Paragraphs (Bobbs-Merrill) $\bar{r} = .96$; one week retest reliability $\bar{r} = .99$

APPENDIX B

Items for Teacher Rating of Child-Environment Interaction Processes

Ratings: 1. rarely or never
2. occasionally
3. often
4. almost always

1 2 3 4

1. He/she stands ground and stays in novel situations (compared with withdrawing, resistant or fearful reactions).

2. He/she engages in solitary play when given free choice of activity.

3. Freedom to explore or manipulate the environment for its own sake, without others' approval or attention, appears rewarding to him/her.

4. He/she waits turn, shares and otherwise cooperates in group activities.

5. He/she can attend to two things simultaneously or in rapid alternation (e.g., listening or talking while working at another task with hands).

6. He/she fails to follow adults' directives or initiatives (verbal or non-verbal); is withdrawn or aloof with adults.

7. He/she uses language freely in social or information-focused communications with others (compared with action, gesture or emotional-display forms of communication unaccompanied by language).

8. He/she shows sense of social rules and justice in his/her activities (e.g., respects others' and public property, pressures peers who break rules, asks permission, thanks).

Ratings: 1. rarely or never
2. occasionally
3. often
4. almost always

1 2 3 4

9. He/she is somewhat adventurous and experimental in his explorations of the environment -- including moderate risk-taking and/or hazarding of failure.

10. He/she uses peers as resources (in seeking information, explanation, judgement in dispute, help with clothing or other task or need).

11. He/she uses adults as resources (in seeking information, explanation, judgement in dispute, help with clothing or other task or need).

12. He/she selects activities independently; decides without difficulty or procrastination.

13. He/she actively refuses to follow peers' suggestions, directions, game rules; resists peers' demands (verbally or physically).

14. He/she appears to enjoy talking and listening to speech; participates in story sharing, question and answer, other verbal activities.

15. He/she is successful in obtaining and defending materials and advantages from others.

16. He/she uses materials and equipment appropriately and effectively, showing appreciation of their purpose and care requirements.

17. He/she is concerned about adult opinion; seeks adult approval

18. He/she persists at tasks with reasonable tolerance of frustrations and limitations.

19. He/she leads group; secures group attention, initiates activities, is imitated.

Ratings: 1. rarely or never
2. occasionally
3. often
4. almost always

1 2 3 4

20. He/she manipulates or attempts to manipulate peers (control primarily for its own sake).

21. He/she manipulates or attempts to manipulate adults (control primarily for its own sake).

22. He/she uses language to reproduce dialogue, songs, rhymes.

23. He/she is well received as a playmate; is generally accepted by the group.

24. He/she uses materials in a variety of ways, showing originality.

25. He/she expresses affection to peers, either verbally or by smile, touch, sharing or other gesture.

26. Task completion (e.g., of puzzle, assignment, other structured activity) appears rewarding to him/her without others' approval or attention.

27. Praise by adults in response to his/her behaviour appears rewarding to him/her.

28. He/she follows verbal directions in a work or game activity (e.g., "Simple Simon", craft project).

29. He/she expresses to others pride in his products or attributes (e.g., shows his work to others).

30. He/she shows flexibility (e.g., in change of routine -- he/she is not upset; in "shifting gears" -- he/she switches easily from active to quiet activity or from humour to seriousness

APPENDIX C

Relative Change in Development-Achievement Measures Pre- to Post-SKG as Determined with Multiple Regression (Group 2)

Note. Squared multiple correlation coefficient (R^2) indicates the proportion of score variance due to assessment occasion after removal of variance due to individual children.

^aCircus 8 variance due to occasions attains .05 level of statistical significance only after removal of variance due to age (although age variance alone is statistically nonsignificant).

D-1

APPENDIX D

GROUP 2: Pre-SPS through Post-GR2.

GROUP 3: Pre-SPS and Post-GRI.

APPENDIX D

Profiles of Development-Achievement Measures (Standardized Scores) for Highest- and Lowest-Rank Children (Groups 2 & 3).

Note. See Appendix A for full titles and descriptions of measures.

APPENDIX E

Rating Distributions for Three Selected Interaction Process Items (7, 19, 30) by Age-Grade Level and Group, with Individual Off-Mode Ratings Indicated for Case Study Children (Highest, Lowest, Most Changed).

Note 1. For purposes of cross-sectional comparison, ratings by teacher D are also entered separately on occasions where she is one of two raters.

Note 2. Shaded entries indicate modal rating.

Note 3. Individual off-mode ratings for case study children are indicated as follows:

highest achiever ▲	lowest achiever ▼
second highest ,	second lowest
achiever ↑	achiever ↓
most positively	most negatively
changed in	changed in
achievement rank ①	achievement rank ①

ITEM 30: SHOWS FLEXIBILITY

	Group 1				Raters	Group 2				Raters	Group 3				Raters	Group 4				Raters		
	1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4		1	2
Age-grade																						
Pre-SPS																						
Post-SPS/ Pre-KG																						
Pre-SKG/ Post-KG																						
Post-SKG/ Pre-GR1																						
Post-GR1																						
Pre-GR2																						
Post-GR2																						
Pre-GR3																						
Post-GR3																						

END

24.12.80

FIN