

IDENTIFYING CHILDREN AT RISK: THE PREDICTIVE
VALIDITY OF KINDERGARTEN SCREENING MEASURES

By

S. Suzanne Jacobsen

B.Sc., California State Polytechnic University, 1973

M.A., California State Polytechnic University, 1978

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS OF THE DEGREE OF
DOCTOR OF EDUCATION

in

THE FACULTY OF EDUCATION

Department of Educational Psychology and
Special Education

We accept this thesis as conforming to
the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 1990

© S. Suzanne Jacobsen

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of *Education Psychology and Special Education*
The University of British Columbia
Vancouver, Canada

Date *Sept 18, 1990*

Abstract

Identifying Children at Risk: The Predictive Validity of Kindergarten Screening Measures

The early identification of children who are "at risk" of experiencing learning problems is of interest to educators and policy-makers. Conflicting evidence exists regarding the efficacy of screening measures for identifying children "at risk". The rationale for screening programs is that early identification of problems allows for treatment which may eliminate more severe problems from developing. If a student is identified as "at risk", school personnel may intervene with remedial programs. Subsequently, if the student succeeds, the earlier prediction is no longer valid. The identification of "at risk" would appear inaccurate because the intervention was successful in improving skills. Researchers often measure the prediction of "at risk" with a correlation coefficient. To the extent that the intervention is successful, the correlation of the identification of "at risk" with later measures of achievement is lowered. One of the problems with research on early prediction has been failure to control for the effects of the interventions which were implemented as a consequence of screening. An evaluation of "at risk" prediction is important because results of screening procedures are used to make decisions about retentions and the allocation of special services.

The purpose of this study is to investigate the relationship between kindergarten screening measures and grade three achievement for two entire cohorts enrolled in 30 schools in one school district. The

analysis employs a two-level hierarchical linear regression model to estimate the average within-school relationship between kindergarten screening measures and grade three achievement in basic skills, and determine whether this relationship varies significantly across schools. The model allows for the estimation of the relationship with control for individual pupil characteristics such as age, gender and physical problems. The study examines the extent to which the relationship between kindergarten screening and grade three achievement is mediated by children receiving learning assistance or attending extended (4-year) primary schooling. The study also examines differences among schools in the kindergarten screen/achievement relationships and the achievement of "at risk" pupils by including school characteristics in the analysis.

The results of this study indicate positive relationships between kindergarten screening measures and achievement outcomes, even after controlling for age, gender and physical conditions. The kindergarten screen/achievement relationship did not vary among schools. The study failed to demonstrate that controlling for interventions would improve the kindergarten screen/achievement relationship; in fact the effects were in the opposite direction.

Levels of adjusted achievement of pupils who obtained scores at the cut-off point for risk status varied significantly among schools. The "at risk" pupils performed better on all four achievement measures in schools with high school mean-ability than similar pupils in schools with low school mean-ability.

These results show that progress in the study of the predictive validity of screening measures can be made through the use of

hierarchical regression techniques. Researchers need to give consideration to the effects of educational interventions and the contextual effects of schools.

Table of Contents

	<u>Page</u>
Abstract	ii
Table of Contents	v
List of Tables	viii
List of Figures	ix
Acknowledgements	x
Chapter 1: Introduction	1
Background of the Problem	1
Purpose of the Study	6
Definition of Terms	6
Research Questions	8
Rationale	9
Limitations of the Study	10
Justification for the Study	11
Organization of the Study	12
Chapter 2: Review of Literature	13
Part 1: Early Identification of Risk	13
Underlying Assumptions of Risk	14
At Risk On Screening Measures	15
Advantages of Early Identification	17
Disadvantages of Early Identification	18
Status on Criterion Measures	20
Problems in Identification	23
Kindergarten Screening	24
Screening Instruments	28
Human Figure Drawing Test	29
Copying of Geometric Shapes	31
Tests of Language Development	33
Knowledge of Letters and Numbers	35
Summary of Screening Tests	37
Part 2: Factors Which May Affect Prediction Research	38
Age at Entry	38
Gender Differences	39
Health Problems and Physical Problems	41
Educational Interventions	43
Remedial assistance and its effect on achievement.	44
Retention and its effect on achievement.	47
Contextual Effects	50

Part 3: Prediction Studies	55
Reliability and Validity	55
Methodological Paradox of Prediction-Performance Research	58
Prediction-Performance Research	59
Correlation Analysis	59
T-tests and ANOVA	60
Discriminant Analysis	62
Prediction-Performance Matrices	63
Multiple Regression	68
Multilevel Modelling	69
Prediction Studies Summary	71
Summary	73
Chapter 3: Research Methodology	76
Introduction	76
Subjects	78
Procedures	79
Instruments	82
Draw-a-Person Test	83
Mann-Suiter Visual Motor Screen	83
Kindergarten Language Screening Test	84
Deverell Test of Letter and Numbers	84
Canadian Test of Basic Skills	85
Research Questions	85
Analysis of the Data	89
Data	89
Analyses	91
Preliminary Analyses	96
HLM Analysis	96
Threats to Validity	97
Summary	99
Chapter 4: Findings	100
Introduction	100
Correlation Matrix	100
Format of the Tables	102
Model I: Kindergarten Screening Measure/Achievement Relationships	103
Model II: Controlling for Pupil Characteristics	108
Model III: Controlling for Educational Interventions	112
Model IV: School-Level Variables	115
Parameter Variance Explained	118
Model V: Four Kindergarten Screening Measures in the Model	120
Model VI: Simplified Models Including Only Significant Variables	124
Summary	124

Chapter 5: Summary and Conclusions	125
Overview of the Study	125
Principal Findings of the Study	126
Limitations of the Study	140
Implications of the Study	141
Recommendations for Future Research Research	144
References	148
Appendix A: Descriptors of Physical Problems	175
Appendix B: Number of Pupils Identified with Physical Problems and Number Receiving Interventions	176
Appendix C: Technical Information - Draw-A-Person	177
Appendix D: Technical Information - Mann-Suiter Visual Motor Screen	178
Appendix E: Technical Information - Kindergarten Language Screening Test	179
Appendix F: Technical Information - Deverell Test of Letters and Numbers	180
Appendix G: Technical Information - Canadian Tests of Basic Skills (CTBS)	182
Appendix H: Characteristics of Schools in the Study	183
Appendix I: Use of Grade Equivalent Scores	184
Appendix J: Data Plots Reflecting Interventions	187
Appendix K: Graphic Representation of Predictive Utility	191
Appendix Table 1: Summary of Selected Prediction Performance Studies	195
Appendix Tables 2: HLM Results for Attrition	201
Appendix Tables 3-20: HLM Results	202
Appendix Table 21: Means and Standard Deviations of Outcome Measures For Four Samples	220
Appendix Table 22: Prediction-Performance Matrix Analysis	221

List of Tables

	<u>Page</u>
Table 1: Means, Standard Deviations, and Correlations of Student-Level Variables	101
Table 2: District and Sample Means and Intercepts for Pupils at the Risk Cut-off Score	103
Table 3: Estimates of the Effects on Grade Three Achievement of One Point (or one SD) Kindergarten Screening Measure Score	106
Table 4: Estimated Residual Parameter Variance of Mean Achievement for Pupils at the Cut-Off Score for Risk Status	111
Table 5: Estimated Coefficients for Kindergarten Screen/Achievement Relationships	114
Table 6: "Null-Model" Estimates of Variance in Grade Three Achievement	119
Table 7: Average Achievement Scores for Pupils Who Scored at the Cut-off Point for "Risk" Status	122

List of Figures

	<u>Page</u>
Figure 1: Effect of Changing the Cut-off Score of the Predictor or Criterion Measure	22
Figure 2: Prediction-Performance Comparison Matrix	64
Figure 3: Numerical Example of Prediction-Performance Matrix . . .	66
Figure 4: Administration of Kindergarten Screening Measures . . .	82

Acknowledgements

I would like to thank the members of the committee for their support and participation, Dr. D. Willms, Dr. P. Leslie, Dr. J. Conry and Dr. O.A. Oldridge.

I wish to extend my appreciation and gratitude to Doug Willms. He provided the opportunity to conduct the study. He patiently guided me through the hierarchical linear modelling analysis. He contributed ideas to the manuscript at every stage and extended support and encouragement throughout the study. I am indebted to him for his contributions to this work and to my education. I wish to thank him for being a better advisor than I had hoped for.

I wish to thank Steve Raudenbush for the long-distance support and HLM trouble-shooting. He willingly took phone calls at any hour and patiently provided advice on the analyses.

Lastly, I owe special appreciation to my husband John, for his support and encouragement; to our daughters, Sulynn and Teresa, for their untiring support, patience and help; and to our son Kirin, whose surprise arrival made being a student difficult but whose presence brings us great joy every single day.

Chapter 1

Introduction

Background of the Problem

Educators and policy makers are interested in the early prediction of children's school achievement because it is important to intervene with appropriate educational strategies for children "at risk" of experiencing difficulty in school achievement. Researchers estimate that between 15 and 25 percent of the children in school have learning difficulties and thus, are "at risk" of school failure (Satz & Friel, 1978; Norton, 1979). For the educational system to respond to the needs of "at risk" children, their early identification is a priority.

In recent years, screening programs have been implemented throughout Canada and the United States for the purpose of early identification and treatment of children who display signs of possible learning problems (Norton, 1979). Individual school districts are responsible for the selection and administration of particular screening instruments and provide the resources for subsequent intervention programs. These screening programs involve the evaluation of large groups of children with brief, low-cost procedures. Typically, a screening test or battery will include developmental measures and measures of specific skills related to academic performance. The findings are used to make some initial distinction between children who are expected to progress successfully and those who may need special services.

The rationale for screening programs is that early identification of

problems allows for treatment which may prevent more severe problems from developing. The value of identifying children "at risk" of school failure is determined by the intervention efforts which follow the identification (Mercer, Algozzine & Trifiletti, 1988). Screening, therefore, is only the first step in a process aimed at identifying specific skills prerequisite to successful academic performance for individual children. The results of screening should alert educators to general areas of delayed development and lead to preventive action to improve academic performance.

Some educators oppose the use of screening techniques because the outcome of the screening process involves labelling children "at risk" or "not at risk". The application of a label "at risk" may lead to further assessment and intervention which may eventually result in the child acquiring another label such as "learning disabled" or "mentally handicapped". Researchers have stated the negative effects of labelling include: the label has an adverse effect on teacher expectancy (Foster, Schmidt & Sabatino, 1976; Foster, Ysseldyke & Reese, 1975); classmates, teachers and parents perceive the child more negatively than their normal peers (Pullis & Smith, 1981) and the child's self-concept is reduced (Guskin, Bartel & McMillan, 1975).

Although educators are not in complete agreement regarding the value of kindergarten screening, legislation in the United States (Public Law 94-142) and in certain provinces in Canada (e.g., Bill 82 in Ontario), requires school districts to identify "high risk" children in kindergarten. Typically, the identification of these children involves the use of tests, but the validity of screening measures is frequently

questioned. For a screening program to be effective both content validity and predictive validity must be given consideration. Construct validity is also important. Construct validity is not necessary for prediction research in which the purpose of the study is to find predictors of a particular criterion (Borg & Gall, 1983). However, it is essential to prediction research which relates theory to research findings. Screening measures with construct validity strengthen the predictive validity of the measures in theory-based research. Some screening instruments fail to meet sufficient standards of technical adequacy for use in making decisions about children (Bracken, 1987). Shepard and Smith (1986) point out that the validity of a screening instrument depends on how the test is used and is entwined with the effectiveness of interventions that follow.

Most predictive validity studies express validity in terms of correlation coefficients which give no indication of the accuracy or efficiency of the tests. The result is that screening tests are commonly selected on the basis of inadequate data, face validity, testimonial evidence, or frequency of use by other screening programs (Lichtenstein, 1981).

One concern regarding the validity of screening programs is the lack of theoretical framework for conceptualizing the nature of handicapping conditions and the precipitating factors which lead up to them. Satz, Taylor, Friel and Fletcher (1979) state, "without a testable theory one lacks guidelines for the selection of a test battery which purports to identify the potentially high risk child" (p. 318). Jarman (1980) notes, however, that the lack of theoretical rationale is implicitly encouraged

by the exploratory nature of prediction research.

Policy makers and legislators are asking for better evidence to document the immediate and long-term effects and cost-benefits of early intervention (White, 1986). Judy (1986) suggests the long-term consequences of screening include significant savings to society in terms of services that will not be required, increased educational productivity and enhanced self-concept of children who otherwise might experience academic failure before assistance could be provided.

Many researchers have developed screening and readiness tests to assess facets of children's development assumed to be related to later school achievement (de Hirsch, Jansky & Langford, 1966; Book, 1974; Satz & Friel, 1978; Beery & Buktenica, 1982; Goldman, Fristoe, & Woodcock, 1970; Gauthier & Madison, 1973; Devereill, 1974; Harris, 1963). Despite many years of research directed at early identification the results are inconclusive. Researchers do not know which factors predict risk of failure, whether remedial interventions affect achievement or whether early screening tests reliably predict achievement for children "at risk".

Reviews of research provide conflicting evidence regarding the efficacy of screening measures for identifying children who risk learning failure. Factors which contribute to conflicting findings include: short time-frame prediction (the usual study is kindergarten to grade one); small samples (Glazzard, 1979; Meyers, Attwell, & Orpet, 1968); failure to consider the effects of gender (Badian, 1986); and failure to consider minor physical conditions which may affect academic performance. Wendt (1978) reports that there is considerable variability

in the purposes for screening and types of measures, and thus, not all investigations are directly comparable.

An important issue of prediction research concerns the confounding effects of interventions on the relationship between screening and subsequent achievement. If students are suspected of potential learning problems because of low kindergarten screening scores, school personnel may intervene with some remedial program. In schools with good remedial programs, children identified as "at risk" would, on average, attain higher achievement scores than children in other schools with comparable screening scores who received little or no remediation. The earlier identification of "at risk" would therefore appear inaccurate for some schools because the intervention was successful in improving their skills. In effect, successful interventions would lower the correlation of the identification of "at risk" with later measures of achievement.

The decision for intervention may be dependent on factors other than awareness of screening information. Also, individual schools may vary in teaching practices, allocation of resources, availability or intensity of interventions or class size and heterogeneity (Raudenbush & Bryk, 1986). Although there is not consensus among researchers, there is growing evidence that children perform better in schools with a high SES enrollment, even after family background characteristics are considered (Willms & Raudenbush, 1989; Willms, 1986). This is called contextual effects. These, and other factors, may also influence the relationships between screening scores and subsequent achievement.

Nearly all research on early identification has been concerned with the relationship between screening measures and subsequent achievement,

but few studies have given consideration to whether the relationship varies across schools, or to the effect that intervention efforts may have in mediating the relationships. This study examines the relationships between kindergarten screening measures and grade three achievement in basic skills. It investigates whether the relationships vary significantly across schools, and examines the extent to which the relationships are mediated by children receiving learning assistance or attending extended primary schooling. The study also examines whether the inclusion of school level variables in the analysis, such as school size, heterogeneity, rural or urban location or mean ability, provides explanation for between school differences in the relationships.

Purpose of the Study

The purpose of the present study is to examine the relationships between performance on four kindergarten screening instruments and achievement test scores obtained at grade three, for two entire age cohorts enrolled in 30 schools in one school district. The analysis employs a two-level hierarchical linear regression model to estimate the average within-school relationship between kindergarten screening measures and grade three achievement in reading, mathematics, vocabulary and language and determines whether these relationships vary significantly across schools. The analysis controls for the effects of individual student characteristics of age, gender and physical problems. The study also examines the extent to which the relationship between kindergarten screening and grade three achievement is mediated by children receiving learning assistance or attending extended primary schooling. The study also examines the extent to which the between

school differences can be explained by school level variables of school size, heterogeneity, rural or urban location and mean school ability.

Definition of Terms

The definitions of educational terms used in this study are listed below:

Screening refers to a systematic examination of supposed prerequisite skills and abilities necessary for learning in school. The absence of these is expected to identify educationally "at risk" children at an early stage (Stevens, 1987).

Screening tests are tests which provide a brief assessment of a child's developmental abilities, particularly those which are highly associated with future school success.

Learning difficulties is used to describe all conditions (i.e., learning disabled, developmental delay, experientially deprived, educationally handicapped, etc.) which result in the identification of children who may have difficulty in achieving the goals of education.

"At risk" is defined as the designation acquired from performance on a kindergarten screening measure. It is used in the literature to refer to children who have significantly low test scores on screening measures and who are likely to be retained in grade or referred for special educational services. It also is used to refer to children who fail to make adequate progress in school and therefore, are achieving below the level expected for their age and grade (Karweit, 1988). The literature is confusing because "at risk" refers both to the prediction of a child's subsequent status and to students who have failed to make adequate progress. In this study, "at-risk" is used in the former sense; it will

be used to describe children who score at or below a particular score on a kindergarten screening measure.

Learning assistance is individual instruction provided by a teacher other than the classroom teacher to remediate deficit skills. It is usually provided during regular classroom time in a room other than the classroom.

Extended primary consists of attending school four years to complete the three-year primary curriculum. It may represent a program of continuous progress or may be an alternate form of retention.

Research Questions

The purpose of this study is to examine the relationship between kindergarten screening measures and grade three achievement in reading, mathematics, vocabulary and language. The study also examines the extent to which the relationship between kindergarten screening and grade three achievement is mediated by the provision of learning assistance or extended primary. The study examines four research questions:

1. a) What is the average within-school relationship between grade three test scores in academic achievement and scores on kindergarten screening measures of perceptual-motor, language and cognitive skills?

b) To what extent do the relationships between achievement scores and kindergarten screening scores vary across schools?

2. a) What is the relationship between grade three achievement and kindergarten screening after controlling for the effects of gender, age at entry to kindergarten and whether the child has a physical problem?

b) Do the relationships between grade three achievement and kindergarten screening vary across schools after taking account of pupil

characteristics?

3. a) To what extent are the relationships between grade three achievement and kindergarten screening mediated by the provision of learning assistance or extended (4 year) primary schooling?

b) Does the extent to which the relationships are mediated vary across schools?

4. a) If there is significant variation between schools in their relationships between screening and outcome measures, to what extent can it be explained by school size, rural versus urban location or the school mean and variance of pupils' ability?

b) To what extent are the between-school differences in achievement explained by various school-level variables?

Rationale

The study addresses some of the shortcomings of previous research. Despite many years of research directed at early identification of children "at risk" for learning problems, the results are inconclusive. Prior research has been limited by factors that contribute to conflicting findings: short time-frame prediction, small samples, failure to consider the effects of gender, and failure to consider minor physical conditions that might affect academic achievement. The factors which predict risk of failure, whether early screening reliably predicts achievement for children "at risk" and in what ways intervention affects achievement are not known with certainty. Numerous researchers state that screening for the identification of "at risk" learners is of vital importance to help prevent children from exposure to recurrent failure and frustration (Norton, 1979; Keogh & Becker, 1973; Barnes, 1982). This

study extends the examination of prediction research by controlling for the effects of remedial interventions during the study and by employing techniques of analysis that investigate variation across schools.

This study is a result of the need for further investigation into the area of early identification of "at risk" children using a statistical technique which considers the hierarchical nature of educational programs. In a sense it serves as a case study of one district's screening practices. The analysis allows for the examination of the relationships within schools, as well as the study of how the between-school relationships may vary. This study is an attempt to gain greater understanding of the relationships between early screening information and academic achievement, giving consideration to the effects of educational interventions and to school effects.

Limitations of the Study

This study uses extant data available from a school district screening program that was implemented for a period of ten years. The particular screening measures are not necessarily the best predictors of risk status, and the outcome measures are limited to the standardized achievement tests scores. Admittedly, achievement tests do not cover all the important goals of schooling. Nevertheless, the system was operating in this district and is better than most screening programs in that there are four measures administered at different times of the year. Also, the district established district norms for the cut-off scores on the screening measures.

This study is limited also in the following ways:

- the four screening instruments measure specific but limited

aspects of development;

- it is concerned exclusively with academic outcomes as measured on standardized tests;
- the physical problems are grouped as a single variable and thus, there is no discrimination between the effects of hearing, vision, speech or physical handicaps;
- although, the analysis controls for whether pupils received some intervention, there is no data describing the type or quality of the specific treatments administered, their timing or their duration;
- similarly, the school-level measures are proxies for processes and policies operating within the schools. Although the inclusion of these variables is useful for controlling for their effects, the analysis does not provide the kind of detailed information necessary for informing school policy.

Justification for the Study

If screening procedures determine whether children receive special services or are retained a year or more, then an evaluation of the accuracy and predictive validity of screening, giving consideration to the effects of interventions, is necessary. The primary emphasis of screening is to predict which children are likely to experience school problems and to provide intervention to children who require it to progress successfully. This study addresses the need for further investigation into the area of early identification of pupils designated "at risk" using hierarchical linear regression analysis. Analysis which allows for the investigation of the within-pupil and between-school

variation is an attempt to gain greater understanding of the relationships between early screening information and grade three achievement, giving consideration to the effects of educational interventions and school effects.

Organization of the Study

Chapter 2 is a Review of Literature which is divided into three major parts. Part one reviews the literature on early identification and screening practices. Part two discusses variables which may influence the predictive validity of screening programs. Part three describes prediction-performance methodology and concludes with a summary of the chapter. Chapter 3 describes the methodology for the study. Chapter 4 presents the findings of the study. Chapter 5 summarizes the findings, discusses the implications for educational policy and provides recommendations for future research. The bibliography and appendices follow Chapter 5.

Chapter 2

Review of Literature

To investigate the relationship between kindergarten screening measures and academic achievement a review of several important areas is necessary. The area of screening and early identification of children at risk for experiencing difficulty in learning is complex. Researchers apply the definition of "at risk" first to performance on screening measures and secondly, to subsequent performance on achievement outcome measures. The particular areas of development considered to underlie achievement and the means to measure those areas must be considered. Prediction research requires a time lag between screening and outcome thus, intervening factors which occur during the time under study and which may affect prediction findings merit discussion. The methods of analyzing prediction research vary greatly, and the means chosen affect the interpretation and application of the findings.

To address these areas the review of literature is divided into three principal parts. The first part is a review of early identification, screening practice and screening instruments. The second part discusses factors which may affect prediction research. The third part discusses prediction-performance studies and examines methods of analysis for prediction-performance research. The chapter concludes with a summary of the important considerations from the review of literature.

Part 1

Early Identification of Risk

The early identification of children "at risk" of experiencing learning difficulties receives wide support from professionals of varied

disciplines and from parents (Keogh & Becker, 1973). A variety of social, emotional, intellectual, biological, physical, linguistic, environmental or any combination of such factors may interfere with a child's optimum growth and normal development and result in the child requiring special attention. Before school age, the identification of children "at risk" generally rests with professionals in the fields of health and social welfare. When children reach school-age, the educational system assumes primary responsibility for the identification of students believed to be "at risk" of experiencing difficulty in school achievement. Many educators believe the identification of the school-age child with learning difficulties should be made as early as possible in the child's school career to prevent more serious learning problems from developing (Haring & Ridgway, 1967; Judy, 1986; Norton, 1979; Barnes, 1982). Most children enter the public school system in kindergarten, which provides the earliest opportunity for screening for potential difficulties.

Underlying Assumptions of Risk

The implementation of screening practices for identifying children is based on several common assumptions regarding models of risk. Keogh and Daley (1983) identify three important assumptions. First, the problems in development have their primary locus in the child. Second, development is assumed to be continuous such that early problems are precursors of subsequent problems. Third, differentiated treatments or interventions are assumed to be directly linked to risk or disability conditions. Two additional assumptions identified by Barnes (1982) are that the child at risk can be detected early in a reliable and predictable manner and that the earlier a child's potential disability is

detected, the more effective and beneficial treatment will be.

At Risk On Screening Measures

Describing educational risk and detecting children who fall into such a classification are not simple tasks. Numerous definitions of risk are used which confound explanations and contribute to inconsistent research findings and make comparisons of research findings difficult.

Researchers use poor performance in areas of developmental skills, maturity, or preacademic skills as indicators of risk status for young children. In classifying the "at risk" population some researchers identify a specific criterion, or "cut-off" score, on a particular instrument or battery of tests to indicate risk status (Fletcher & Satz, 1982; Book, 1974), while others use a vague definition (Jansky, 1978; Stevens, 1987; McCann & Austin, 1988; Karweit, 1988). Zeitlin (1976) defined the "at risk" population as those children, who because of problems of development and experience, are least able to meet the expectations of the school. Leigh (1983) stated that children who lack essential preacademic knowledge and skills are at risk for academic failure. Lerner, Mardell-Czudnowski and Goldenberg (1981) defined preacademic deficits as any deficit in cognitive, affective or psychomotor domains that might hamper the child's progress in learning and in school.

Opinions differ regarding the number of children at risk for experiencing difficulty in learning academic skills. Norton (1979) estimates that 25 percent of children entering school show some signs of developmental deviations in areas having a slower rate of development such as language, perceptual skills and attentional maturity. Mardell

and Goldenberg (1975a) used the lowest 10 percent of screening scores to identify high-risk children, children who were seriously behind other children of the same age, gender and location.

An important concept underlying screening of young children is that of readiness. The concept of readiness is built on the belief formulated during the 1920's and 1930's that a child developed in predetermined stages which could be measured by a readiness test (Durkin, 1974). Numerous tests were developed and used for early identification of risk status. A high score on a readiness test indicated the child was ready to learn while a low score indicated the child was not ready and instruction should be delayed. Children who encounter difficulty in school often were described as immature or lagging in development.

de Hirsch, Jansky and Langford (1966) found a close link between children's maturational status in kindergarten and reading and spelling achievement several years later. In a landmark project on screening, they developed a Predictive Index to be used as a screening device for identifying children who were likely to experience difficulty in learning to read. A later study by Jansky and de Hirsch (1972) concluded that behavior, maturation, language and reading readiness were the factors which best predicted academic performance.

After many years of research numerous individual screening tests and screening batteries have been developed. Most tests and batteries of tests include tasks to measure both developmental abilities and readiness skills. Meisels (1987) defined and analyzed the differences between developmental screening tests and readiness. He indicated that developmental tests provide a brief assessment of a child's developmental

abilities which are highly associated with future school success. Readiness tests are concerned with the curriculum-related skills a child has acquired, skills that are typically prerequisite for specific instructional programs. The primary areas included in early identification assessment are language, intelligence, motor skills, social-emotional development and preacademic skills. (For reviews see: Dykstra, 1967; Satz & Fletcher, 1979; Book, 1974; Paget & Bracken, 1983; Bracken, 1987; Mercer, Algozzine & Trifiletti, 1988.) Children are identified as potentially "at risk" of experiencing learning difficulties when their obtained scores are below a designated critical level of performance on a screening test or a battery of tests.

Advantages of Early Identification

Proponents of early identification of children "at risk" of experiencing school difficulties suggest several advantages. Keogh and Becker (1973) suggest that the sooner treatment begins, the greater the likelihood the treatment will have a positive impact and the treatment may prevent the development of other deleterious conditions or compounding problems. The early identification of handicapping conditions allows for family adjustment and acceptance which may result in additional support for intervention efforts (Hayden, 1974; Keogh & Becker, 1973). The implementation of well-designed early intervention can yield many positive benefits for children and their families. These may include: enhancing developmental progress as compared with children who are not provided with appropriate interventions (Casto & Mastropieri, 1986; Reynolds, Egan & Lerner, 1983); improved school performance as demonstrated by fewer grade retentions (Leinhardt, 1980), better academic

test scores and a reduction in numbers of school dropouts (Lazar & Darlington, 1978; Schweinhart & Weikart, 1986); and reducing the total number of years of special education required, resulting in a significant cost savings (Keogh & Daley, 1983; Lazar & Darlington, 1978).

The stated goals of screening programs include intention to provide appropriate interventions following screening, however the results of an extensive program of early childhood screening in five Head Start centers do not support the stated goals. Richardson-Koehler (1988) found that the screening data was used to identify special needs children, but was not used to provide feedback about individual or group needs, nor for planning and implementing instruction geared to meeting the identified needs of children. Teachers and aides had little understanding of the purposes and appropriate uses of the test results. They viewed the identification of special needs children as unrelated to other aspects of their work.

Disadvantages of Early Identification

The major concern expressed regarding screening programs is that the outcome of screening may lead to "labelling" children "at risk". The literature describing early identification and categorizing children as "at risk" or "not at risk" fails to support either negative or positive effects leading from the labels. Keogh and Becker (1973) suggest that when a child is identified "at risk", a set of expectancies, anxieties, and differential treatment patterns may develop both at home and at school which may be detrimental to the development of the child. Most concerns regarding the labelling of children stem from the belief that teacher or parent expectancy results in a "self-fulfilling prophecy

effect" as suggested by Rosenthal and Jacobson (1968).

Palady (1969) showed a strong relationship between teacher expectancy and student performance. In a study involving two groups of 22 elementary teachers, Foster, Schmidt and Sabatino (1976) concluded the label "learning disabled" generated negative expectancies which affect the teacher's objective observation of behavior and may be detrimental to the child's academic progress. They suggested that labelling may serve as self-fulfilling prophecy because of lowered expectations among parents and teachers. In reporting the findings of a rating task given to 38 teacher trainees, Foster, Ysseldyke and Reese (1975) reported that the teacher trainees held negative stereotypical expectations of children labelled "emotionally disturbed".

Bak, Cooper, Dobroth and Siperstein (1987) administered a questionnaire to 77 fourth-through sixth-grade children. Findings indicate that the children enrolled in regular classrooms who did not have special needs held higher expectations of the capabilities for the special needs children enrolled in a resource room than for the special needs children enrolled in the special class. The authors conclude that special class placements can act as de facto labels.

MacMillan, Jones and Aloia (1974) reviewed the literature and found few studies investigating the effects of labelling children. They concluded that the available data was inconclusive regarding the effects of labelling. They noted that children identified as handicapped received special education assistance which confounded the effects of labelling with the effects of special class placement. Rogers, Smith and Coleman (1978) contend that special class placement had a favorable, not

negative, impact on children's self-concepts although, their investigation did not investigate the effects on achievement or sustained positive self-concept if the children were returned to regular class placement.

The outcome of screening does not necessarily lead to a child being labelled as learning disabled, emotionally disturbed or some other such categorical label. Although categorical labels may be avoided, Keogh (1977) suggested that terms such as "high risk" or "at risk" may themselves assume the characteristics and consequent detrimental effects of labels.

Mercer, Algozzine & Trifiletti (1988) summarize the major disadvantages of early identification programs as follows: "Since measurement inadequacies and differential developmental problems make it difficult to accurately diagnose children "at risk", the major disadvantage of early identification becomes evident. Many children who are not disabled receive a disability label and the detrimental effects of that label present a problem to the child and his/her family."

Status on Criterion Measures

One cannot discuss prediction without an adequate definition of the criterion. The problem is the number of possible criterion measures is as varied as the number of schools using screening programs. The determination of risk status on a criterion measure is equally varied. Some researchers use a specific criterion, or "cut-off" score to indicate risk status. Fletcher and Satz (1982) used a designated number of years below grade level to discriminate mild and severe deficits from average performance. Jansky (1978) used the criteria of "failed in reading at

the end of second grade". The bottom quarter of the achievement distribution has also been described as representing the population not meeting with success (Book, 1974; Karweit, 1988). Other researchers use vague definitions or broad descriptors for students who fail to perform at a predetermined level such as: students failing in their first years (Stevens, 1987); and students who have increased probability of learning problems, adjustment difficulties, or dropping out of school (Scott, 1981).

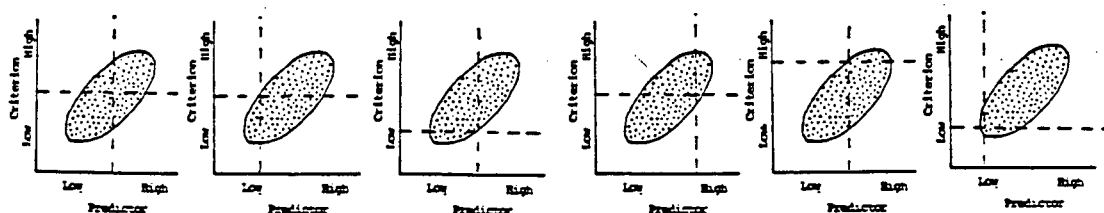
Most researchers attempt to dichotomize the outcome variable by selecting a particular cut-off score. Children who fall below the cut-off score are considered to be most vulnerable for dropping out or for longer term failure. Researchers often use the term "at risk" to refer to children who score below the cut-off point on a criterion measure. In fact, Richardson-Koehler (1988) pointed out that in many cases the term "at risk" has taken over from such descriptors as disadvantaged, low SES, underachieving and problem children.

In many studies researchers use screening measures to categorize children as "at risk" or "not at risk". They also dichotomize the outcome variable using a predetermined cut-off score. When this is done, the relationship between "at risk" status and criterion depends heavily on the selection of the cut-off scores. The decisions regarding the particular criterion measures and the cut-off scores selected for screening measures influence the validity of prediction of risk status. This problem makes it difficult to compare results across studies even if the same criterion measures are used because the cut-off scores may vary across studies.

Figure 1 illustrates the effect of moving the cut-off score for either the screening measure or the criterion measure.

Figure 1

Effect of Changing the Cut-off Score of the Predictor or Criterion Measure



The number of pupils identified as at risk can be increased or decreased by adjusting the cut-off score of either the predictor or criterion measures. The adjustment of the cut-off score of the screening measure has implications for the provision of services for pupils "at risk" and for subsequent costs to school districts. The adjustment of the criterion measure has implications for validating the effectiveness of the screening program. Thus, the absolute number of pupils identified as "at risk" on screening measures or of failing to achieve at a critical level, is an artifact of the cut-off scores selected.

Another important point which makes comparisons across studies difficult is that local standards vary, across districts and schools. McCann and Austin (1988) suggest "at risk" refers to students who, for whatever reason, are at risk of not achieving the goals of education, of not meeting local standards to complete their education. Beyer and Smey-Richman (1988) identified the "at risk" population as those students who are not meeting minimum standards of academic achievement as

determined by locally imposed standards. To the extent that local standards may vary, the identification of particular children "at risk" will vary.

Problems in Identification

A number of difficulties present in attempting to identify kindergarten children who are likely to experience learning difficulties. One of the difficulties when assessing young children is that differential developmental patterns make it difficult to determine if a particular child is truly at risk or simply needs more time to mature before becoming an efficient learner (Mercer, Algozzine & Trifiletti, 1988). Young children tend to produce widely discrepant results for the same test administered more than one time due to their changing physical, mental and emotional conditions. Judy (1986) explained that these changing conditions in young children result in large standard errors of measurement in their test results.

A further problem in attempting to identify kindergarten children who are likely to experience learning difficulties is that the conditions of learning difficulty or failure have not developed at the time of identification (Keogh & Becker, 1973). Thus identification is an hypothesis that a problem will develop, not a confirmation that it exists. A related difficulty is that educationally handicapping conditions, like reading disabilities or language disorders, have few, if any, well-known etiological components. Identification of "at risk" by screening therefore refers to the presymptomatic detection of a disorder which could interfere with the child's progress if left undetected (Barnes, 1982). Children whose kindergarten performance deviates

significantly from "normal" ranges may be easily identified; however, the deviation of performance for children with mild handicaps may not be great in kindergarten but may result in learning difficulties later in school (Paget & Nagel, 1986).

Despite these difficulties, investigators in psychology and education continue working to develop methods of predicting children's achievement and identify those children "at risk" (Stevenson, Parker, Wilkinson, Hegion, & Fish, 1976). Screening for children "at risk" is not a precise or exact science, partially because many aspects of early development are subject to considerable variation in time and across individuals and many intervening variables during the time under study cannot be totally controlled.

Kindergarten Screening

Barnes (1982) defines screening as a process of early detection to identify those children in the general population who may be at risk for a specific disability or who may otherwise need special services. The stated purposes for early screening and identification include: identifying those children who have special learning needs (Lerner, et al., 1981; Salvia & Ysseldyke, 1985); describing individual strengths and weaknesses, particularly as these relate to programming (Meisels, 1985); developing appropriate recommendations for interventions tailored to each child's individual needs (Bricker, 1986); attempting to provide preventive education, rather than to wait for problems to crystallize in later grades which would then require more costly and less effective remediation strategies (Wendt, 1978); and facilitating delivery of early intervention services, thereby, enhancing eventual adjustment (White,

1986).

Social and educational policies concerning screening children have been guided by the notion that prevention is preferable to remediation (Evans, 1976; Glazzard, 1982). The long-term consequences of screening include significant savings to society in terms of services that will not be required, increased educational productivity and the enhanced self-concept of children who otherwise might have experienced academic failure before assistance could be provided (Keogh & Becker, 1973; Hayden, 1974; Keogh & Daley, 1983; Lazar & Darlington, 1978). The importance of early identification of children "at risk" is based on solid evidence showing that early identification coupled with remedial assistance can reduce the risk of school failure and subsequent grade retention (Simner, 1983; Becker & Gersten, 1982; Lazar & Darlington, 1978).

Screening is only the first step in a process to provide appropriate services to children. Meisels (1987) suggested that testing in kindergarten should only be used to make better and more appropriate services available to the largest number of children. According to Leigh (1983), when screening is not followed by provision of either a thorough diagnostic evaluation or some type of intervention within a reasonable period of time, early identification efforts serve no useful purpose. Keogh and Daley (1983) state that unless identification leads to differentiated services, the screening is wasteful.

The main purpose of screening is the examination of large age groups of children with brief, low-cost procedures, to identify those children who appear to fall above or below certain critical levels of performance

(Gulliford, 1976). Single variable and multiple-variable predictive batteries are used to predict academic achievement. Some of the single variables studied as predictors of academic achievement include: visual acuity (Griffin & Eberly, 1971); hearing acuity (Goetz, 1971); chronological age (Dykstra, 1966); gender (Weintraub, 1966); and intelligence (Black, 1971).

Numerous studies examine combinations of variables or screening batteries for predicting achievement or "at risk" status (Glazzard, 1979; Book, 1974; Satz & Friel, 1974; Adelman & Feshbach, 1971; de Hirsch, Jansky & Langford, 1966). The major variables considered to be among the best predictors of first-grade performance when children were tested in kindergarten include:

- performance on tasks of cognitive development (Kaufman & Kaufman, 1972);
- facility at using a pencil (Eaves, Kendall, & Crichton, 1974);
- pre-academic skills such as letter recognition (Colligan & O'Connell, 1974; Telegdy, 1975; Badian, 1986; Keogh & Becker, 1973; Stevenson, et al. 1976; Mercer, Algozzine, & Trifiletti, 1988);
- level of perceptual development (Morency & Wepman, 1973);
- level of language development (Jansky & de Hirsch, 1972; Eaves, Kendall & Crichton, 1974; Stevenson et al., 1976).

Some studies have investigated teacher observations and behavioral or academic ratings of children as a screening tool or as a component of a battery, but the findings are inconclusive. Kirk (1966) found the ratings of three teachers on 112 kindergarten children favored older

children as being bright and younger children as being slow. She concluded the ratings had reached an acceptable level in identifying slow children, but were only marginal at identifying bright children. Meyers, Atwell and Orpet (1968) found that the behavior rating of attention in kindergarten was as predictive of reading words, comprehension and spelling in fifth grade as was a picture vocabulary test.

Feshbach, Adelman and Fuller (1974) concluded from a study of 888 kindergarten children that a kindergarten teacher's ratings could predict first grade reading achievement as efficiently as a psychometric battery. Wells and Peterson (1978) studied 111 kindergarten children in four classes. They found the Kindergarten Teachers' Checklist (KTC) was a good predictor of scores in the Iowa Tests of Basic Skills in first grade accounting for 30% of the variance. Keogh's (1977) study found that trained observers using a systematic observation of children's behavior were in strong agreement with kindergarten teachers' perceptions of risk or non-risk. In a second related project involving 250 kindergarten children and 20 teachers, Keogh found teachers' ratings of children in kindergarten and first grade consistently favored girls, although objective measures did not yield significant gender differences.

In a review of research Simner (1983) found that many traditional warning signs of school failure had little actual bearing on later school achievement. He identified five signs more likely to be evident among kindergarten children who are truly at risk for school failure than among children who are not at risk for failure. The five signs are: in-class attention span, distractibility or memory span; in-class verbal fluency; in-class interest and participation; and letter or number identification

skills and printing errors. He concluded that children at risk are "...not necessarily lacking in many basic motor, language, drawing, and copying skills when compared to the average kindergarten child" (p.24). The results of this, and other studies, suggest that, to the extent that the criterion measures are related to school performance or school success, predictors based on school-related tasks are significantly more reliable than measures of other tasks (Meisels, 1984).

Screening Instruments

Measurement instruments, other than behavior scales, used for kindergarten screening may be described as developmental tests and school-readiness tests. Developmental screening tests are designed to identify children who may have a learning problem or a handicapping condition that could affect their overall potential for success in school. Such tests focus on performance in a wide range of areas including speech, language, cognition, perception, affect and gross and fine motor skills. Readiness tests focus on current skill achievement and performance rather than on a child's developmental potential. Thus, readiness tests and developmental screening tests sample different, although potentially overlapping areas of measurable behavior (Meisels, 1984). The primary areas included in early identification assessment are language, intelligence, motor skills, social-emotional development and preacademic skills. Individual screening tests and screening batteries have been developed to assess these areas. (For reviews see: Dykstra, 1967; Satz & Fletcher, 1979; Book, 1974; Paget & Bracken, 1983; Bracken, 1987; Mercer, Algozzine & Trifiletti, 1988.)

Human Figure Drawing Test

Children's drawings have been investigated as a measure of developmental status since the late 1800's. Many researchers have described the changes in children's drawings over the course of development. (For reviews see: Goodenough, 1926; Harris, 1963.) Goodenough (1926) developed the Draw-A-Man Test based on the assumption that the intelligence of children could be estimated from their drawings of the figure of a man. Harris (1963) revised and refined the Goodenough test and the Goodenough-Harris Drawing Test became the principal rating approach applied to children's drawings to estimate intellectual ability (Naglieri, 1988). The purpose of the test is to measure intellectual maturity which was defined by Harris (1963) as the ability to form concepts of an abstract character. Evaluation of the child's drawing of a human figure serves as a way of measuring the complexity of his or her concept formation ability (Sattler, 1988).

In a study in which children were asked to draw a person, Ferinden, Jacobson and Linden (1970) found that a high-risk drawing correctly identified 99% of the children who had difficulty with reading in first grade. Eaves, Kendall and Critchton (1974) found the draw-a-man subtest of the Modified Predictive Index (de Hirsch et al., 1966) was strongly correlated with word analysis at the grade two level.

The Draw-a-person test has been shown to correlate with intelligence tests such as the Wechsler (Dunn, 1967; Pikulski, 1972; Tramill, Edwards & Tramill, 1980) and the Stanford Binet (Ritter, Duffy, & Fischman, 1974). Correlations range from .24 to .88. Interrater reliabilities are satisfactory ranging from .80 to .90 for the point scale and .70 to .90

for the quality scale (Sattler, 1988; Naglieri & Maxwell, 1981). Sattler described the Draw-A-Man Test as an acceptable screening instrument for use as a nonverbal measure of cognitive ability.

Dunleavy, Hansen, Szasz and Baade (1981) administered a human figure drawing test to 141 kindergartners. In comparing a group of students who passed the Metropolitan Readiness Test with students who failed the test, the human figure test had identified 42% of the non-ready children. The researchers concluded the test was useful for the early identification of the academically not-ready child.

Goldman and Velasco (1980) investigated the relationship between human figure drawings and risk for experiencing emotional problems. Their results suggest that drawings which omit important body-parts are predictive that the child is a high risk for developing emotional problems.

Duffy, Ritter and Fedner (1976) conducted a study of 80 children who were administered a battery of tests including the Draw-A-Man test in kindergarten. The test was a statistically significant predictor of academic success as measured by the Total Stanford Achievement Battery. However, because the test accounted for only 9.3% of the variance, the researchers concluded the Draw-A-Man test had little practical utility as a predictor of school performance.

In a recent study in which the scoring system was altered to utilize an empirically derived subset of items, Simner (1985) found the overall predictive validity of the Draw-A-Man test equalled or exceeded that achieved with many other school readiness tests. He concluded that if scoring was confined to certain key items, the drawings could identify

five-year old children who are at risk for school failure.

The Draw-a-person test has been found to be predictive of school performance. The purpose of the test is to measure intellectual maturity. The areas of performance measured by the test include nonverbal cognitive ability and concept formation. The test requires fine motor ability to draw with a pencil and attention to complete the task.

Copying of Geometric Shapes

Tests of copying of geometric shapes have long been used as developmental tests for children. Developmental factors which may be involved in copying simple geometric designs include appropriate motor development, perceptual discrimination and the ability to integrate perceptual and motor processes (Sattler, 1982).

de Hirsch (1966) stated that pattern copying and human figure drawing tasks require a relatively high degree of integrative competence. Both tasks, according to de Hirsch, are like reading, writing and spelling, in that they require the ability to organize parts into a meaningful whole.

Two form copying tests have been widely researched as predictive instruments, the Bender-Gestalt Test (Bender, 1938) and the Developmental Test of Visual Motor Integration (Beery, 1989). Research on these tests indicates that the age at which figure-copying tests are administered may affect their predictive power.

Duffy, Ritter and Fedner (1976) reported the Developmental Test of Visual-Motor Integration administered to 182 kindergartners was a significant predictor of reading and mathematics subtests of the Stanford

Achievement Test in second grade. They also indicated, however, that the practical utility of the test was limited.

Ferinden, Jacobson and Linden (1970) reported the Bender-Gestalt administered in grade one was a better predictor of reading ability in grade one than when it was administered in kindergarten. Duffy, Keogh and Becker (1966) found negligible correlations between scores on the Bender and grade three reading ability when the effect of intelligence was held constant. In a longitudinal study conducted by Stevenson et al., (1976), the Bender-Gestalt test proved to have little predictive value for achievement measured in grade two.

In a review of research on traditional warning signs of school failure, Simner (1983) reported on fourteen studies involving copying tasks. The relatively low correlations (.00 to .54) led Simner to advise caution when using the results of copying tasks for the purpose of individual prediction.

Park (1978) suggested the power of drawing tests was in identifying children at risk for learning problems. Because these tasks require attention, fine motor control, attention to detail and the ability to follow instructions, poor performance might reflect disruption of the processes of focal attention which could lead to difficulty in learning.

Tests of copying geometric designs have been found to be predictive of school performance. The predictive power may vary with the age at which the test is administered. The skills required include visual perception, motor production and visual motor integration. The task also requires attention and ability to follow directions, skills which are necessary for success in school.

Tests of Language Development

Researchers have provided empirical support for the hypothesis that competence in oral language is predictive of satisfactory performance in academic skills (Ilg & Ames, 1964; Jansky & de Hirsch, 1972; Zeitlin, 1976; Stevenson et al., 1976; Steinbauer & Heller, 1978; Book 1980). General belief and acceptance of this relationship has influenced the design of curriculum and development of primary level materials (Gray, Saski, McEntire & Larsen, 1980).

Although it seems logical that various language processes and skills underlie subsequent academic achievement, a strong cause-and-effect relationship has not been established. A review of research illustrates that findings vary considerably depending on the particular oral language process being considered and the outcome measure analysed.

Jansky and de Hirsch (1972) used a predictive battery developed from their previous research. They reported that the Oral Language subtests accounted for the greatest proportion of variance (14%). The picture naming subtest was reported to be highly predictive of reading status at the end of grade two ($r=.53$).

Similar findings were reported by Satz, Friel and Rudegeair (1976) following a factor analysis of kindergarten abilities and their relationship to grade three achievement. They found that the oral language factor contributed most to reading while visual-motor ability contributed most to spelling.

Stevenson et al. (1976) found that attention span and verbal fluency were the best overall in-class indicators of future academic achievement. Verbal fluency included the spontaneous use of precise

words and the capacity to convey abstractions when asked to describe events.

Groff (1977) reviewed the research related to oral language and reading. His analysis identified ten studies which concluded there was a significant degree of correlation between oral fluency and reading achievement. At least eleven studies indicated a significant relationship between the complexity of oral language and reading achievement. In contrast, six studies indicated no significant relationship between syntax complexity and reading achievement. (See Groff for details and references.)

Simner (1983) reviewed research studies and reported that basic language skills such as defining common words, naming colors and body parts, and identifying pictures of common objects, show only marginal relationship with subsequent school achievement.

Basic assessment and child development texts often state that language and intelligence are closely related and it is impossible to indicate where one ends and the other begins (Salvia & Ysseldyke, 1985; Anastasia, 1976; Papalia & Olds, 1975). Language cannot be measured without measuring intellectual abilities to some degree. Gray, Saski, McEntire and Larsen (1980) illustrated the close relationship between language and intelligence. Their study of 74 five- and six-year old children indicated that a strong and statistically significant correlation between oral language and readiness existed when age was controlled. However, the language test did not discriminate readiness groups when intelligence was entered as a covariate. The researchers concluded there was little relationship between oral language and school

readiness. They suggested that pervasive effects of IQ were a significant determinant of a child's performance on both measures of oral language and school readiness.

Hammill and McNutt (1980) synthesized the results of 89 correlational studies of the relationship of various language constructs to measures of reading. Their results indicated a low relationship ($r=.39$) between oral receptive language and reading and practically no relationship between oral expressive language and reading. Their review of literature focused on hypothetical constructs of various language processes and their relationship to reading, not on specific tests or subtests. They suggest that individual subtests or tests vary greatly in their predictive power.

The strength of the relationship between oral language proficiency and school achievement is not clear. The effects of age and intelligence confound findings when not controlled. The close relationship between intelligence and language makes measurement of "pure" language difficult. The demands of curriculum materials and expectations for oral participation within the classroom make a measure of the child's oral language desirable as a screening measure.

Knowledge of Letters and Numbers

Chall (1967) conducted a thorough review of the research on the relationship between knowledge of letters and reading. She concluded that a child's ability to identify letters by name in kindergarten or the beginning of grade one was an important predictor of reading achievement in grade one and two (r 's from .3 to .9). (See Chall for details and references.) Wide support for the strength of this relationship led to

the inclusion of tasks of letter recognition in many readiness tests (Jansky & de Hirsch, 1972; Deverell, 1974; Adelman & Feshbach, 1971).

In a study in which Dykstra (1967) isolated the components of readiness tests, he found that tests of letters and numbers were the best single predictor of reading success in grade one. The addition of other tasks on readiness tests added little to the predictive value of the test.

Numerous researchers also found tests of letter recognition to be important factors for predicting reading success (Durrell, 1958; Askov, Otto & Smith, 1972; Jansky & de Hirsch, 1972; Klein, 1977; Butler, 1979; Badian, 1986). Stevenson et al. (1976) reported that the number and letter skills which children knew before entering kindergarten were good predictors of their learning during their first three years of school. Busch (1980) conducted a study of 1000 grade-one students and concluded that the ability to recognize upper- and lower-case letters and beginning sounds was the best single predictor of reading achievement.

Lesiak's (1978) data provides guidelines for cut-off scores. He indicated the average kindergarten child could name 14 to 15 of the 26 upper-case letters. Children likely to experience later learning problems could only name one to five letters.

The primary expectation for schooling is literacy. Pupils are expected to learn to read and write. Letter recognition has been found to be predictive of reading success. The educational goals that children learn to read and write makes a measure of letter recognition a logical choice as a screening measure.

Summary of Screening Tests

The literature describing four types of screening measures has been reviewed; draw-a-person tests, copying geometric designs, expressive language and recognition of letters and numbers. The tests measure particular areas of development but there is overlap in the skills they measure. For example, the draw-a-person is administered as a measure of cognitive ability but it requires visual-motor integration and fine-motor skill to draw. The test of expressive language requires a level of receptive language for the child to perform. All the tests require receptive language, ability to follow directions and attention.

These four tests cover three areas of development generally assumed related to early school learning: cognitive development; visual-motor development; and language development (Lesiak, 1978). The implementation of a screening program is intended to identify children who may experience difficulty learning and who require intervention to alleviate or eliminate the source of the difficulty (Keogh & Becker, 1973). These particular screening tests have been found to be predictive of school achievement, and therefore they are consistent with the stated purpose of screening.

Part 2

Factors Which May Affect Prediction Research

Many factors may have an effect on the outcomes investigated by prediction-performance research of young children. Three areas which deserve consideration are individual pupil characteristics, the effects of interventions which occur during the time under study, and contextual effects of schools. Three characteristics of the child which may affect the relationship between prediction and performance are reviewed: age at entry to kindergarten; gender; and physical problems. Two educational interventions, learning assistance and attending an extended primary program, are also discussed. Contextual effects of schools is discussed briefly.

Age at Entry

Researchers and reviewers have addressed the question of within-grade age effects because chronological age is the major criterion for admitting children to school. When children who are the youngest in their grade are compared with their older classmates, they are usually less successful (Beattie, 1970; Davis, Trimble & Vincent, 1980; Ames, 1963; Hall, 1963; Miller & Norris, 1967; Gredler, 1978).

Hedges (1977) reviewed in detail the research literature related to screening and early identification. One area focussed on age at entrance as it relates to school success. The conclusions he reached from the literature on age at entry follow: the older children are at entrance, the greater their chances of academic success; when comparing the achievement of an older child of comparable intelligence, the younger child's scores remain inferior; younger children do not seem to have the

social maturity desirable for successful performance; and chronological age has slightly more effect on boys in younger and normal age groups than on girls of comparable intelligence (see Hedges for details and references). Other studies noted that children who were youngest in their class had the highest incidence of failure (Walsh, 1988); were more likely to repeat a grade (Lloyd, 1978); were more likely to be referred to special education (Di Pasquale, Moule & Flewelling, 1980); and were more likely to be labelled as learning disabled (Diamond, 1983).

Two important points regarding research on age at entry were made by Shepard and Smith (1987). The first point they make regarding the research on age at entry is that most researchers fail to control for the effects of intelligence or gender in analysing age effects. After analysing the age trend by the ability status in one study, Shepard and Smith (1987) suggested that the low achievement reported for some younger children was more accurately a combination of youngness and low ability than of age alone. The second major point was the consideration of practical rather than statistical significance. They found differences in percentile between the oldest three month children and the youngest three month children was only nine points, a difference of little practical significance.

Gender Differences

Research findings vary regarding the existence, cause and significance of gender differences in academic achievement. In a frequently cited review, Maccoby and Jacklin (1974) reported that differences existed between males and females in measured verbal and quantitative ability. Females tended to score higher on verbal ability

and males tended to score higher on quantitative ability. The differences were apparent during elementary years and increased into adolescence and adulthood. Numerous studies from many countries report sex-related differences in achievement (Walden & Walkerdin, 1985; Hanna & Kuendiger, 1986; Brandon, Newton & Hammond, 1987; Johnson, 1987; Shuard, 1986). Aiken (1972) reported sex-related differences as early as kindergarten, with males performing at higher levels than females. Fox and Cohn (1980) reported differences in performance between males and females in early elementary school to be small, but increasing through grade seven.

Researchers suggest a number of possible explanations for the observed differences in performance. Although it is generally acknowledged that gender differences in intelligence are non-existent (Sattler, 1988; Hyde, 1981; Stockard, 1980), some researchers interpret selected research findings to suggest there may be biological differences in intellectual aptitude (Fox & Cohn, 1980). Other researchers seek to explain the differences by social and psychological factors. Parental views and behaviors towards children of different sex which are used as possible explanatory variables include: providing gender-specific toys and different opportunities for play (Fennema & Peterson, 1985); providing active, vigorous play for boys which encourages the development of spatial and constructional skills for boys (Burton, Drake, Ekins, Graham, Toplin & Weiner, 1986); and holding lower educational expectations for girls than for boys (Maccoby & Jacklin, 1974). School settings and culture may provide differential opportunities which lead to differential performance (Hieronymus, King, Bourdon, Gossling, Grywinski,

& Moss, 1976) as illustrated in the following ways: males are more likely to be assigned to high-ability groups for mathematics (Hallinan & Sorenson, 1987); males often receive more hours of formal instruction in primary years than do females (Eccles & Jacobs, 1986); teachers have lower expectations for girls and make less academic demands (Burton, et al., 1986); and teachers may promote confidence, flexibility, risk-taking and rule breaking, behaviors found more often in males than females (Walden & Walkerdine, 1985). Student attitudes and motivations may explain differences in performance in various academic subjects (Good & Slavings, 1988; Burton et al., 1986; Pattison & Grieve, 1984).

One important fact noted in a study conducted by Martin and Hoover (1987) was that there is greater variability in the skills of males than females across all subtests and all grades three to eight on the Iowa Tests of Basic Skills. Sabers, Cushing and Sabers (1987) also noted that the size of the differences between the sexes was not very great when compared with the differences within the sexes. Males were found to be more variable than females in both reading and mathematics. Willms and Kerr (1987) found that social-class differences were far greater than gender differences. They reported differences in mean levels of performance between working class and middle class groups to be between 1.25 and 1.5 standard deviations, compared with sex differences of about .25 of a standard deviation.

Health and Physical Problems

Educators must deal with the effects of physical disease and impairment on a regular basis. Academic progress can be negatively affected by chronic illness (Stehbens, Kisker & Wilson, 1983), speech and

language difficulties and motor production difficulties (Gubbay, 1975), and the side effects of medications may influence attention and concentration (Rapoport & Flint, 1976). The relationships between various disabilities, allergies and chronic illness and a variety of adverse academic, social and emotional problems have been demonstrated by Cowen, Weissberg and Gisare (1984), Larter (1982), Kornberg and Kaplan (1980), and Rawls, Rawls, and Harrison (1971).

The implications for meeting the needs of students who are blind, deaf or physically handicapped are apparent in school settings. However, mild to moderate physical conditions which are less obvious may also affect a student's academic performance. The range of possible physical handicapping conditions includes, but is not limited to, problems of vision or hearing, speech and language difficulties, motor production difficulties, allergies and physical illness. These difficulties may appear insignificant when compared with the difficulties of children with obvious physical disabilities, but the impact on school performance may be great. Grimley and McKinlay (1977) stated that children with subtle difficulties of learning can be in desperate need of help and if their needs are not recognized, secondary emotional problems are bound to arise.

The estimated prevalence of communication disorders is five percent of school age children (Frisch & Handler, 1974). Research studies of children with speech difficulties or motor performance difficulties frequently refer to subjects demonstrating difficulty in both speech and motor areas (Yoss & Darley, 1974; Jenkins & Lohr, 1964; Gubbay, 1975; Gordon and McKinlay, 1980; Crary, 1984). Difficulties in speech and

motor production may negatively affect acquisition of skills in spoken and written language and the children may be perceived as lazy or unmotivated because they fail to complete daily assignments (Gubbay, 1975).

One in five children has a major allergic disease (Rapoport, 1976). Reaction to food is one of many variables which may combine and interact to give rise to learning and behavior problems (Hammond, 1980). Allergy related problems may complicate learning problems. For example, comprehension deficits may be intensified by otitis media resulting from allergy, and the side effects of allergy medication may increase attentional disorders and hyperactivity in some children (Mc Loughlin, Hall, Isaacs, Petroski, Karibo & Lindsey, 1983).

The incidence rate of medical problems which may influence performance in school is large enough to be given consideration. Gortmaker and Sappenfield (1984) estimated that 10 to 20 per cent of all children have a chronic medical disorder. Perrin (1986) reported that two percent of all children suffer from a severe chronic illness that regularly interferes with daily activities including school attendance and performance. The effects of various medical problems are usually studied in isolation. Researchers investigating prediction-performance on educational outcome measures rarely control for the effects of physical problems on performance.

Educational Interventions

Educators continually question how best to help students experiencing academic difficulty. Interventions are implemented in the belief that they will help improve the academic achievement of students.

Remedial assistance is a widely accepted practice, but there is little controlled assessment of the effectiveness of remedial programs.

Retention is also a widely accepted practice. Retention policies and rates vary greatly between schools and across districts (Holmes & Matthews, 1984; Jackson, 1975).

Where educational interventions are provided during the time between the administration of a screening measure and an outcome measure, the intervention may affect the student's performance on the outcome measure. Two educational interventions which may affect student performance on outcome measures are remedial assistance and retention in grade.

Remedial assistance and its effect on achievement.

When students have not progressed within the regular instructional program, educators may intervene with a remedial program for individual students. Resources are allocated to provide instructional intervention to increase student success (Deno, 1986). Remedial programs are usually intended to supplement the regular educational program. Most often, students are taken out of their regular classrooms for remedial instruction in specific academic areas, often reading or mathematics (Madden & Slavin, 1987; McNutt & Friend, 1985).

The research literature is ambiguous regarding the efficacy of educational interventions. The evidence is not strong for positive, negative or neutral effects. Comparisons across studies are difficult because intervention models vary in many ways such as, setting, instructional strategies, types of pupils served and goals for instruction. One difficulty in interpreting research findings is that it

is not clear whether differences are related to materials studies, setting, grouping or the effects of the remedial label.

Few studies compare the progress of students participating part-time in resource room programming with students in regular class placement. Smith and Kennedy (1967) studied educable mentally handicapped students assigned randomly to either daily part-time resource room instruction or full-time regular class placement. They found no significant differences in academic achievement between the groups. In contrast, Glavin, Quay, Annesley and Werry (1971) found behavior disordered students participating in resource room programming gained significantly in reading and mathematics achievement as compared with behavior disordered students in regular class placement.

The important issue is not the setting in which remedial instruction is provided, but the effectiveness of the remediation on academic achievement. Variables which have been identified as having strong relationships to the acquisition of academic skills include: time and opportunity to learn (Gettinger, 1984); level of academic engaged time (Haynes & Jenkins, 1986); opportunities for a student to make correct responses (Greenwood, Dinwiddie, Terry, Wade, Stanley, Thibadeau & Delquadri, 1984); and implementation of a specific reinforcement contingency plan (Shapiro, 1987). An individual or small group intervention program may include some or all of these variables. Shapiro (1988) stated that interventions which incorporate these potent variables have been shown to be powerful and effective in remediating academic skills.

Numerous researchers have criticized resource room programs for

reasons such as: failing to increase academic learning time (Haynes & Jenkins, 1986); failing to coordinate instruction with that of the classroom (Johnstone, Allington, & Afflerbach, 1985); and failing to produce transfer to the regular program (Anderson-Inman, 1986).

Gallagher (1984) considers resource rooms ineffective and Affleck, Madge, Adams & Lowenbraun, (1988) found them to be more costly than other alternative interventions.

Thurlow, Ysseldyke, Graden and Algozzine (1983) observed eight students receiving instruction in the classroom and in a resource room. Although opportunities for differentiated instruction were available in the resource rooms, no practical differences were noted in the amount of time the students were actively engaged in instruction in the two settings.

Some researchers have reported positive findings for resource room programs. Leinhardt (1980) reported the findings of a study of low-achieving kindergarten students who were promoted to first grade, but were given a special remedial instructional program. At the end of grade one, the low-achievers performed at higher levels than promoted students given conventional instruction or students who were placed in a transition room with special instruction.

Wolfenden (1980) conducted a longitudinal study of 108 students over four years. He reported that remedial intervention, started in kindergarten, reduced the number of grade retentions and individual assistance programs that would have been required if the intervention had not occurred.

Other researchers report positive effects of intervention programs

on reading (Boehnlein, 1987) and mathematics (Peterson, 1989). Madden & Slavin (1983) examined effective remedial programs and determined that the achievement of students identified as "at risk" can be significantly increased, by either extensive modifications in the regular program or by intensive remedial pull-out intervention.

Retention and its effect on achievement.

Retention is a common educational practice but research has failed to validate its effectiveness. A critical review of research on retention by Jackson (1975) concluded there was no reliable evidence that grade retention resulted in higher achievement for pupils having difficulty learning than did grade promotion for similar pupils. Holmes and Matthews (1984) analysed eight studies in which retained students were matched with promoted counterparts on the basis of achievement. They concluded that the research did not support that retention improves basic skills.

In contrast, some researchers report positive achievement gains for retainees. McAfee (1981) reported on three groups of students: those retained in grade one; those who were in a compensatory education program during grade one; and those promoted to grade two. McAfee's analysis of the data revealed retention appeared to be beneficial in early grades, one to four, but had no effect in intermediate grades, five to seven.

Sandoval and Hughes (1981) defined successful retention as one in which the retained child completed the retained year ranking in the top third of the class. They found that students who made academic and social-emotional gains after repeating grade one lacked serious academic deficits in the year prior to retention, had strong self-esteem and

social skills and showed signs of difficulty in school because of lack of exposure to the material.

Peterson, DeGracie and Ayabee (1987) studied first-, second-, and third-grade retainees matched on several variables as same age students not retained. Retained students improved their relative class standing by the end of the retained year, but after three years there were no differences between retained and promoted students.

Baenen (1988) conducted a five-year study of 243 students and a comparison group matched on several variables. She reported the following: retention did not meet its goals of helping students catch up to grade level and stay there; there was no significant difference in growth trends in those retained in grade one versus a later grade; and those promoted showed better growth in both reading and math than those retained.

The research on retention has been generally criticized as being flawed and of poor quality (Jackson, 1975; Medway & Rose, 1986). Some problems with the research include: more stringent retention policies exist in some schools than in other schools; control groups are sometimes age peers and sometimes grade peers. The major concern regarding research on retention is the effect of selection bias. That is, selection bias may favor promotion because, at the time of the decision to promote or retain, the promoted students were performing better than retained students in ways not captured by the control variables.

One concern regarding studies of interventions or retentions is the threat to internal validity of selection bias which may affect the findings. However, the elimination of selection bias in studies of the

effects of retention would be difficult to achieve. Shepard and Smith (1987) point out that random assignment of children who are candidates for retention into retained or not-retained groups is unethical. The desire for random assignment also lacks feasibility in that parents, or teachers, rather than researchers, often control the decision of whether a child is retained.

Holmes and Matthews (1984) conducted a meta-analysis of 44 studies on retention in which they investigated the effects of selection bias. They calculated 575 effect sizes for variables within the studies. The mean effect size was $-.37$. This indicated that on average, the retained pupils scored $.37$ standard deviation units lower on various outcome measures than promoted pupils. Eighteen of the 44 studies had matched subjects, that is a retained group, and a promoted group matched on several variables. A mean effect size was calculated for the matched group studies to see if it differed from overall effect sizes. The effect size for the matched groups was $-.38$, similar to the effect size for all the studies of $-.37$. The consistency between the two measures supported their conclusion that differences in designs of studies resulted in no significant amount of bias in the results. They concluded that the cumulative research evidence shows the potential for negative effects consistently outweighs the positive outcomes. Their findings suggested that retention had a negative effect on pupil's personal adjustment, self-concept and attitude toward school. They also found that retained students performed 0.44 standard deviations below their promoted counterparts on various measures of academic achievement.

Contextual Effects

Contextual effects is the term used by researchers to describe the effects of the collective properties of a school. These collective properties within a school have an effect on individual pupil achievement over and above the effects of the personal characteristics or attributes of the pupils (Willms, 1985). Researchers have attributed contextual effects to the teaching environment, the disciplinary climate, curriculum patterns, course content and "peer group" influences (Willms, 1986; Summers & Wolfe, 1977; Winkler, 1975, Clifford & Heath, 1984).

The literature describes two alternative points of view which attempt to explain school effects. The first position reflects the organizational view of school effectiveness. The effects of family influences and experiences at school determine the learning outcomes. The school experiences are shaped by the organizational structures and practices of the classroom, school and district. (For reviews of the literature see Anderson, 1982; Murnane, 1981; Rutter, 1983). In general, the findings of these reviews are contradictory, and suggest weak organizational effects (Willms, 1987).

The second viewpoint suggests that the most important determinants of school effects are institutional (Meyer, 1977, 1980). The institutional view suggests that schooling outcomes are determined by elements of the schooling system. These elements are defined by certain rules, roles and definitions. These elements include educational levels, types of schools, curricular topics and the specific roles of instructors and students. They derive their meaning from societal definitions rather than organizational circumstance (Willms, 1987). Meyer (1980) contends

that the structures and practices, the organizational aspects that affect student outcomes are relatively homogeneous within a school, and their effects are small compared with effects of rules, roles and definitions and the institutional effects.

Many researchers have abandoned the search for school effects explained solely by institutional or organizational elements and have turned to examining differences inside schools based on the research concluding that student achievement varies as much, or more, within schools as between schools (Bidwell & Kasarda, 1980).

An explanation which allows for a relationship between organizational or institutional elements and within-school elements as an explanation for school differences is one which incorporates a hierarchical view of educational processes. Two models which conceptualize learning as a multi-level process have emerged in the literature, the additive model and the interactive model of schooling (Gamoran, in press).

Barr and Dreeben (1977, 1983) describe an example of a hierarchical model. They view schools as "nested layers" in which the outcomes of one hierarchical level constitute the inputs at the next level. They suggest that district and school administrators allocate resources to classrooms; key resources include time, curricular materials and the competencies of teachers and students. They emphasize the collective nature of schooling. Students receive instruction in groups (such as classes or within-class groups), so it is the characteristics of the group that must be most closely tied to the instruction that occurs in a given context. Thus, instruction is a group-level outcome with consequences for the

individual-level process of learning. This additive model (Barr & Dreeben, 1983) views classroom instruction as the crucial force in achievement. Learning is seen as a consequence of interaction between individual characteristics and features of instructional opportunity.

Sorensen and Hallinan (1977) suggest that the opportunities for learning apply to classes, not individual students in isolation. In their formulation, class-level variables (opportunities for learning) affect the relation between individual level inputs (ability and effort) and outputs (achievement). Their interactive model of schooling views learning as the result of student ability and effort, but depends on the opportunity to learn (Sorensen & Hallinan, 1977). The additive model and the interactive model each have distinct elements, however, they share a common view of education as a hierarchical model in which processes at one level have an effect on outcomes at another level (Gamoran, in press).

A number of class-level variables have been examined in the research on contextual effects. Instructional time is one class-level variable which has been found to contribute to achievement. It has been measured in many different ways including: the length of the school year (Wiley & Harnischfeger, 1974); daily time teachers devote to instruction (Gamoran & Dreeben, 1986); and time students spend engaged in academic work (Denham & Lieberman, 1980). Instructional practices have been found to contribute to achievement. Instructional practice has been demonstrated in the following ways: the more words taught during first grade reading, the more students learn (Barr & Dreeben, 1983; Dreeben & Gamoran, 1986); the more curriculum covered during the year, the higher the attainments

at the end of the year (Tizard, Blatchford, Burke, Farquhar & Plewis, 1988); the more content coverage in math, and the higher quality of instructional discourse in English, the greater the achievement in the respective subjects (Gamoran, 1988).

School mean SES or school mean ability has been shown to have an effect on pupils' academic achievement, even after controlling for the individual effects of pupils' family background (Willms & Raudenbush, 1989; Willms, 1986; Summers & Wolfe, 1977; Henderson, Mieszkowski, & Sauvageau, 1978; Brookover, Sweitzer, Schneider, Beady, Flood & Wisenbaker, 1978; Rutter, Maughan, Mortimore, Ouston, & Smith, 1979). Attempts to measure directly some within-school processes have shown that a number of variables are associated with school mean SES (Brookover et al., 1978; Alexander, Fennessey, McDill & D'Amico, 1979). Therefore, in the absence of a study that includes a wide range of variables describing administrative and teaching practices, curricula, and school climate, school-level aggregates of pupil-level characteristics such as school mean-ability, may act as proxies for variables describing certain school processes (Willms, 1986).

Few studies have examined the impact of school composition on pupils with below-average ability. In a study of a sample of Scottish secondary pupils, Willms (1985) found that the average ability level of a school was associated with higher exam performance at the secondary level, for pupils of differing levels of ability, even after controlling for individual pupil ability and family background characteristics. Summers and Wolfe (1977) found that elementary school pupils, who tested at or below the average for their grade, scored higher if they attended schools

with high achieving students, but students scoring above average for their grade were not particularly affected.

Recent advances in statistical estimation have shown that single-level methods are not optimal for estimating multi-level models. Problems of aggregation bias and mis-estimation of standard errors have distorted single-level estimates of multi-level processes (see Bryk & Raudenbush, 1987). Statistical methods are now available that permit one to estimate data at more than one level simultaneously, so that each variable can be measured at its own level (Raudenbush & Bryk, 1988; Aiken & Longford, 1986; Goldstein, 1986).

In the examination of the relationship between kindergarten screening measures and achievement, the inclusion of contextual effects as a component of the analysis may provide explanatory power not available from a pupil-level analysis. Hauser (1970) suggests that contextual effects may only be artifacts of an underspecified model. This study provides a good test for an elementary age sample because the model includes several measures of ability in kindergarten and the outcomes are measured in grade three. Consideration of contextual effects is pertinent to this study because if there are significant contextual effects, and if they are stronger for low ability pupils, they would have the effect of lowering the kindergarten screen/achievement outcome relationship.

Part 3

Prediction Studies

The implementation of early intervention should be based on a valid and efficient detection program. The primary concern regarding screening of young children is the potential for misdiagnosis. The false labelling of a child as being at risk may have negative effects on the child and family (Salvia, Clark, & Ysseldyke, 1973; Foster, Schmidt, & Sabatino, 1976; Algozzine, Mercer, & Countermine, 1977) and result in wasted expenditures for unnecessary services (Gallagher & Bradley, 1972).

A screening program may result in two types of misclassifications. One is the identification of children "at risk" who are not actually "at risk" of school difficulties. This can occur if the screening measures are not valid predictors of future academic success, or because children performed poorly on the screening measure due to extraneous factors. Some children may have an accurate screening score indicating they are "at risk", but they are only slow in development. After kindergarten, or even after grade one, they might make rapid gains. The second misclassification is when children are not identified as being "at risk" when they actually could benefit from remedial services. This kind of error can also stem from invalid tests or from measurement error at the time of screening. Children misclassified in this way require intervention but do not receive it. Thus the reliability and validity of screening measures are of primary concern in prediction studies.

Reliability and Validity

Some screening instruments show satisfactory levels of reliability

and predictive validity but others do not (Lindsay & Wedell, 1982). Appendices A-E provide technical information regarding the specific screening measures included in this study. Reliability and validity are related. A test cannot be valid if it is not reliable, but reliability is not sufficient to make a test valid (Gronlund, 1975).

Reliability refers to the consistency and stability of measurement, and reflects the degree to which examiners can rely upon the score (Goodwin & Driscoll, 1980). A reliable test should yield similar results when administered two or more times during a short period to the same students. In young children, development is uneven and thus, measures of their performance tend not to be as reliable as those designed for older children and adults (NAEYC, 1988).

There are several types of reliability and several ways of deriving estimates of reliability which are discussed in texts of educational and psychological measurement (e.g. Anastasi, 1976; Salvia & Ysseldyke, 1985; Glass & Hopkins, 1984; Gronlund, 1985).

Test reliability is determined through statistical procedures and is estimated using correlation methods. Essentially the various methods determine how much error is present under different conditions. In general, the more consistent the test results are from one measurement to another, the less error there will be and the greater the reliability. Different types of consistency are determined by different methods and thus, the reliability coefficient must be interpreted according to the type of consistency being investigated. The major methods of estimating reliability are test-retest, which is an index of stability, equivalent forms, split-half and the Kuder-Richardson method, which are indices of

the internal consistency of items on the test. Without evidence of consistency in the screening measures, the results may simply be products of chance.

Validity refers to the degree to which the instrument measures what it is purported to measure. Validity is an indicator of the accuracy of a test and of the inferences that may be drawn from it; the stronger the validity of a screening test, the more credible its results (Meisels, 1984). A screening measure is valid to the extent that it differentiates between those students who are at risk for experiencing difficulty in school, and those who are not at risk.

There are two ways of determining validity of screening measures: logical and empirical (Zeitlin, 1976). Logical validity refers to a judgement about the adequacy and appropriateness of the content of a test. The test instrument is inspected to determine that the content and format are consistent with the domain of skills, abilities or behaviors that the instrument purports to measure. Empirical validity is determined through statistical procedures. To determine if the test works as it is intended to, the results are compared to a criterion measure that is a meaningful indicator of the target problem (Lichstein & Ireton, 1984).

Validity may be concurrent or predictive. Different ways to determine a test's validity include comparing results with scores derived from other measures given at the same time (concurrent validity) or at a later time (predictive validity). Validity, whether concurrent or predictive, is measured by the strength of association and is frequently expressed by Pearson correlation coefficients. Correlations indicate the

strength of relationship between two instruments and reflect the accuracy with which one measure can be used to predict a second measure. With regard to kindergarten screening, validity is measured by the strength of association between findings identified in screening and the presence of difficulty in school performance as confirmed in subsequent assessment.

The predictive validity of a screening process depends first on the reliability and construct validity of both the screening and outcome measures. If either of the measures are unreliable, the predictive validity of the screening process will be jeopardized. Also, if either the screening measure or the outcome measure does not adequately reflect the constructs they are meant to represent, the judgement about predictive validity will be inaccurate.

Methodological Paradox of Prediction-Performance Research

There are unique features of prediction research during early childhood and kindergarten which may contribute to a methodological paradox. The nature of the research is long-term but there can be political and economic pressures to release findings of predictor measures before the outcome measures are collected. This may lead to a methodological paradox which has implications for the predictive validity of screening measures (Keogh & Becker, 1973; Zeitlin, 1976). If early identification and diagnosis is accurate and remedial interventions are successful, a child at risk of experiencing difficulty receives help which results in successful school performance. Subsequently, the child's score on a criterion measure is improved and the predictive validity of the screening instrument appears to be low. Having identified the child as at risk, the educator is obligated to intervene

and the effects of the intervention limit the predictive validity of the instruments by raising the scores on the criterion measure so that the prediction that the child was at risk appears inaccurate. When the screening "at risk" prediction is accurate, failure to provide remedial intervention would guarantee high predictive validity, however, the purpose of identification is to provide intervention to children who require it to be successful. Where a random sample is selected, findings of predictor measures guarded and no intervention provided, there would be no paradox.

Prediction-Performance Research

Several models for validating screening instruments have been discussed in the literature. Detailed descriptions of the statistical procedures and the ways various methods can be applied to research designs can be found in texts of statistics and measurement (Pedhazur, 1982; Tabachnick, 1983; Glass & Hopkins, 1984). The remainder of this chapter is a review of techniques most frequently used in prediction-performance research. Reference to measurement scales include the following: interval: number represent rank order of observations; ordinal: numbers indicate rank order of observations; nominal: numbers represent categories. For discussion of Scales of Measurement, see Glass and Hopkins (1984) and Ghiselli, Campbell and Zedeck (1981). Appendix Table 1 presents a number of a studies which used multiple-instrument batteries as predictors and includes information regarding the subject sample, time-frame of study, analysis used and correlations obtained.

Correlation Analysis

The first approach to establishing predictive validity of a

screening measure or screening battery is the validity coefficient model. In longitudinal studies, interval data from a criterion measure is correlated with screening test interval data. The resulting correlation coefficient is used as an indicator of the effectiveness of the screening measure. When statistically significant correlations are obtained, an indication of the screening test's predictive validity is inferred.

The results of a correlational analysis describe the degree of overlap between two measures of the same phenomenon, but not the number of correct and incorrect decisions concerning children at risk (Wilson & Reichmuth, 1985). The limitation of correlational analysis was illustrated by Lichtenstein (1981) in a comparison of two screening instruments. The correlation coefficient between the total scores of the two tests was high (.82) which indicated a strong linear relationship, but the classification analysis illustrated the tests identified different children at risk. He concluded that making assumptions about the predictive validity of screening measures on the basis of correlational statistics is tenuous and ill-advised (p.68). Although many researchers report correlational statistics as a component of their analysis, several studies have reported correlational coefficients as their primary means of analysis in prediction research studies. (See: Ferinden, Jacobson & Linden, 1970; Book, 1974; Buttram, Covert & Hayes, 1976; Duffy, Ritter & Fedner, 1976; Rubin, Balow, Dorle & Rosen, 1978; Goldman & Velasco, 1980; Lindquist, 1982; Simner, 1985.)

T-tests and ANOVA

Another approach to establish the predictive validity of a screening

test is to use t-tests (Barnes, 1982; Borg & Gall, 1983; Miller, 1988). Scores obtained on a screening test are used to classify children as "at risk" or "not at risk", that is, interval data is collapsed into dichotomous data. At a later date, the children are tested on a criterion measure to ascertain the validity of the prediction. The criterion data is interval data. T-tests show whether there are statistically significant differences between the mean criterion score of the children identified as "at risk" and the mean score of those identified as "not at risk". Analysis using the means of each group may illustrate that the mean performances are different, but gives no consideration to the distributions of the groups or of any overlap in the distributions. This approach also fails to indicate the number of correctly or incorrectly identified students. Many prediction-performance studies report t-tests as one component of the analysis which also may include correlations and regression analysis. (See: Eaves, Kendall & Crichton, 1972; Hartlage & Lucas, 1973; Stevenson, et al., 1976; Wells & Peterson, 1978; Butler, 1979; Miller, 1988.)

In some prediction-performance research the interval data from the criterion measure is classified into three or more groups such as high, middle or low scoring. Analysis of Variance (ANOVA) is an inferential technique which can be used to determine whether the differences among three or more sample means are greater than would be expected from sampling error alone. If multiple t-tests are used for three or more means, the probability of error increases as the number of groups increases. ANOVA is appropriate because the chance of error is reduced. ANOVA also may be used to compare subgroups that vary on more than one

factor. For example, the differences between males and females within the high, middle and low groups can be investigated using ANOVA. The ANOVA analysis results in an omnibus F value which indicates if means differ significantly. If the F-ratio is significant, special t-tests are used to specify which particular means differ significantly. There are special t-tests for multiple comparisons including: Duncan's, Newman-Keuls, and Tukey and Scheffe.

One-way ANOVA may be used when the subgroups differ on one factor, two-way ANOVA when the subgroups differ on two factors. More complex variations of ANOVA and Analysis of Co-variance (ANCOVA) are discussed in most texts of statistical analysis, but are not used frequently in prediction-performance research. (See Hartlage & Lucas, 1973; Badian & Serwer, 1975; Stevenson et al., 1976; Dunleavy, Hansen, Szasz & Baade, 1981.)

Generally the T-test and ANOVA approaches have less power than a correlational approach; essentially the interval or ordinal data obtained by screening is collapsed to a dichotomous, nominal measure (at risk vs not at risk). If screening could only provide a dichotomous score, the t-test and correlational approaches would be identical.

Discriminant Analysis

Another approach to establishing validity of a screening test is Discriminant Analysis. It is essentially an adaptation of the regression analysis technique, designed specifically for situations in which the criterion variable is categorical rather than quantitative. Discriminant Analysis involves two or more predictor variables and a single criterion variable which reflects an individual's group membership (i.e., "at

risk", "not at risk"). The analysis equation uses the individual's score on the predictor variables in an attempt to predict the group of which the individual is a member. (See Satz & Friel, 1974; La Torre, Hawkhead, Kawahua & Bilow, 1982; Fletcher & Satz, 1982.)

Discriminant Analysis is a useful technique when the criterion variable is in the form of categories reflecting discrete groups. The criterion in prediction-performance research is often based on a continuous variable (e.g. achievement) with a selected cut-off score which, in effect, creates a dichotomous variable. This technique is valuable when the criterion variable is a category such as "drop-out". However, when the criterion is continuous such as academic achievement, one loses power by creating a dichotomous variable based on some arbitrary cut-off score.

Prediction-Performance Matrices

The major model for evaluating the utility of educational screening instruments is a prediction-performance matrix (Meehl & Rosen, 1955). The matrix is composed of two levels of performance on the screening measure, that is, interval data is dichotomized to be "at risk" and "not at risk". The performance on the criterion measure is dichotomized into two levels. The levels are usually designated as poor or good performance.

Figure 2 presents an example of a prediction-performance matrix and lists several formulas created to evaluate the effectiveness of a screening test. Effectiveness is measured as attaining higher rates of accurate identification and prediction than would be possible without the test (Satz & Fletcher, 1979).

Figure 2

Prediction-Performance Comparison Matrix

		Criterion Measure Performance	
		Poor	Good
Screening Measure Performance	Poor	True Positives A	False Positives B
	Good	False Negatives C	True Negatives D

NOTE: Sensitivity= $(A/A+C)$; specificity= $(D/B+D)$; overreferral= $(B/A+B)$; underreferral= $(C/C+D)$; predictive utility of screening positive= $(A/A+B)$; predictive utility of screening negative= $(D/C+D)$.

Quadrant A shows the true positives. These are students who were predicted to be "at risk" by the screening instrument and who performed poorly on the criterion measure. Quadrant D shows the true negatives, students predicted to perform well who performed well on the criterion measure. Quadrant B shows the false positives, students identified as "at risk" by screening measure but "not at risk" on the criterion measures. These students are seen to be misclassified "at risk" by the screening measure because they obtain successful scores on the criterion measure. Quadrant C shows the false negatives, students not identified as "at risk" by screening but identified by later poor performance on the criterion measure. These students are seen to be misclassified "not at risk" by the kindergarten screening measure because they experience difficulty in learning as indicated by low scores on the criterion measures. The two-by-two matrix displays two groups correctly identified

by the screening measure, the true positives and true negatives (Quadrant A, true "at risk", Quadrant D, true "not at risk") and two groups incorrectly identified, the false positives and false negatives (Quadrant B, students identified "at risk" but who perform successfully on the criterion and Quadrant C, students identified "not at risk" but having difficulty).

The outcomes of prediction-performance comparisons can be evaluated by different approaches. Classificational analysis, also called cross-tabulations, evaluates the accuracy of the screening instrument in terms of the correspondence between the screening outcome and the status of the child on the criterion measure. Mercer, Algozzine and Trifiletti (1988) illustrate the outcomes of vertical and horizontal analysis in a review of single-instrument and multiple-instrument prediction studies.

Classificational analysis allows for the comparison of false inclusions and exclusions (B & C) to true positives and true negatives (A & D). Figure 3 illustrates a numerical example of a prediction-performance matrix. By applying a horizontal analysis method, percentages of correct and incorrect outcomes can be obtained. For example, $A/A+B$ and $D/C+D$ give the percent of correct outcomes, $C/C+D$ and $B/A+B$ give the percent of incorrect outcomes. A vertical analysis allows for the consideration of the relationship between prediction (i.e., within the cells) and actual performance. For example, $A/A+C$ and $D/B+D$ give the percent of students who performed as predicted, conversely, $C/A+C$ and $B/B+D$ give the percent of students for whom the prediction was inaccurate. The percent of correctly identified students, also called the overall hit rate, is computed $A+D/A+B+C+D$.

The proportion of students with special needs who are identified accurately as "at risk" by the screening instrument is reported as sensitivity and may be computed as $A/A+C$. Specificity indicates the proportion of children not in need of special services whose scores on the screening measure were above the cut-off score and may be computed as $D/D+B$. Together, sensitivity and specificity permit comparisons to be made between the base rate, or prevalence, of a physical problem and classificational decisions derived from a screening test (Harber, 1981).

Figure 3

Numerical Example of Prediction-Performance Matrix

		Criterion Measure Performance Reading		
		Poor	Good	
Draw-A-Person	Poor	30 19% (32%)	135 81% (14%)	165
	Good	65 8% (68%)	743 92% (86%)	808
		95	878	973

Overall hit rate=79%; Sensitivity=32%; Specificity=86%
 Overreferral=81%; Underreferral=8%
 Predictive utility of screening positive=19%
 Predictive utility of screening negative=92%

These types of analyses allow for the observation of numbers of correctly and incorrectly identified children utilizing the performance score on the prediction and criterion measures. They are called the

predictive utility of screening positive or negative. Policy makers like this analysis because it identifies proportions of the students requiring special services. However, all of these proportions depend on an arbitrary cut-off score for both the screening measure and the criterion measure. By raising or lowering the cut-off point on either the screening measure or the criterion measure, it would be possible to reduce the number of false positives. The consequence of reducing the false positives in this way is that the number of false negatives is increased. The ideal screening instrument would refer all children in need of special services, but minimize the number of false referrals (Lichtenstein, 1981). (For review and comparisons of horizontal and vertical analysis see Mercer, Algozzine & Trifiletti, 1988.)

If one collects data on the screening measure and the outcome measure at the interval level, one can judge the validity of the screening process using a correlational method, which is more powerful than classificational analysis. The data could also be used to report statistics derived from a number of prediction-performance matrices on various cut-off scores for both the criterion and the outcomes.

An important consideration regarding classificational analysis is that the analysis fails to consider effects of interventions which may have occurred during the time under study. Classificational analysis does not consider which students identified "at risk" perform well on the criterion due to the effects of the intervention. If interventions are effective, failure to consider the intervention effects may inflate the false positives because the prediction "at risk" was accurate and the intervention accomplished the desired goal of the students experiencing success.

Multiple Regression

Multiple regression is a statistical technique for determining the correlation between a criterion variable and some combination of two or more predictor variables. The multiple correlation coefficient (R) is a measure of the relationship between a criterion variable and a predictor variable or combination of predictor variables. R^2 is the coefficient of determination and expresses the amount of variance in the criterion variable that is accounted for by all the predictor variables combined. There are several variations of multiple regression analysis: forward, backward and stepwise. Each variation uses a different procedure for selecting predictor variables to obtain the best prediction of the criterion variable.

In multiple regression, the beta coefficients are sometimes referred to as partial regression coefficients. They express the correlation between two variables under the condition that all other concomitantly measured variables are held constant.

The raw score form of the regression equation is useful for predicting the effects on the criterion variable of a unit increase in each predictor variable. The standardized form is needed to interpret the relative importance of various predictor variables.

In kindergarten prediction, multiple regression may be used to identify the amount of variance in the criterion variable accounted for by the predictor variables taken as a group. A variation in the analysis, step-wise regression, may be used to indicate the rank-ordering of the predictor variables in terms of their efficacy in accounting for variance in the criterion variable. This analysis allows the researcher

to identify the screening measures or subtests which account for the greatest variance and may be helpful in selecting batteries of tests for screening programs. (See: Randel, Fry & Ralls, 1977; Rourke & Orr, 1977; Glazzard, 1982; Schmidt & Perino, 1985; Badian, 1986; Jacob, Snider & Wilson, 1988.)

Multilevel Modelling

Multilevel modelling is an extension of simple linear regression. The interpretation of multilevel modelling is similar to the interpretation of ordinary regression with the extension of being able to say whether relationships vary between schools. Multilevel models have been developed which can simultaneously estimate the effects of variables at three or more levels, such as the pupil, school and district levels (Aikin & Longford, 1986; Goldstein, 1986; Raudenbush & Bryk, 1986).

The major reason for using multilevel techniques in educational research is that multileveling techniques allow the researcher to take schools into account in the analysis. Schools introduce an extra random component - an extra degree of uncertainty. In using the multilevel technique the standard error of estimates is a more precise estimate because it includes the extra random component introduced by the school variable. The (two-level) multilevel technique disaggregates the relationship between variables into two components, within-school and between-school. The within-school component describes the relationship inside the school, thus, it compares the Kindergarten screen/achievement relationship of individuals who attend the same school. The between-school component takes account of differences between schools. The multilevel model estimates an overall pupil-level relationship

between outcomes and background factors, taking account of the nested structure of the data. It provides estimates of the relationships within each school that are differentially "shrunk" towards the overall pupil-level relationship (Raudenbush & Bryk, 1986). The multilevel estimates are biased, but consistent, and less variable than OLS regressions which are unbiased, but may have large standard errors when the sample sizes within schools are fairly small.

The result of fitting the multilevel model is an estimated regression line for each school, analogous to the single line for a whole sample. Multilevel regression provides a test of whether observed differences in intercepts and slopes could have occurred by chance, or whether there really are differences in the population of schools. It also provides a test of whether a general tendency in the slopes, is likely to be present in the population, or whether it is a random artifact of the sampling. Third, it provides a test of whether differences in the slopes between schools could have arisen by chance.

In kindergarten prediction research multilevel modelling may be used to identify the relationships between kindergarten screening and achievement outcomes at the pupil-level and determine if the relationships vary significantly among schools. Where findings indicate that schools differ significantly in the kindergarten screen/achievement relationships or the achievement levels of pupil's "at risk", one can attempt to explain differences between schools in terms of school characteristics. In the present study it is possible to observe the relationships between kindergarten screening measures and grade three achievement for the entire sample and to observe between-school variation

in the relationships. After controlling for the effects of student characteristics and educational interventions, it can be determined if these variables have an effect of mediating the relationships between screening measures and achievement.

The parameters (i.e., intercepts and slopes) that specify the relationships between kindergarten screening and achievement within each school can become the dependent variables in a school-level regression that attempts to determine the importance of certain school-level variables (i.e., school size, school mean ability) in explaining between-school variation in the kindergarten screen/achievement relationship or achievement of pupils designated "at risk".

Hierarchical linear modelling allows one to examine what relationships exist within the entire pupil-level sample and between schools represented in the sample. The simultaneous estimation of the parameters at both levels results in more accurate estimates than can be obtained by single-level analysis techniques.

Prediction Studies Summary

Three important factors which influence the strength of prediction using kindergarten screening measures are: reliability of both the screening and criterion measures; validity of the screening and criterion measures; and the analyses employed. The predictive power is weakened if any of the measures have poor reliability or validity.

The technique of analysis has an effect on the power of prediction. Most prediction and criterion measures result in interval data. Frequently the data is collapsed into two categories and analysed as a dichotomous variable. These analyses have less power than analyses which

utilize the full range of the data. Several different analyses were described which use combinations of dichotomous and interval data.

The most frequently used prediction-performance analysis is classificational analysis which utilizes two dichotomous variables. The advantage of classificational analysis is that the percentages of correct and incorrect predictions are identified. This information is of interest to policy-makers because it has direct implications for services provided and expenditures.

The application of a multilevel analysis has several advantages over other analysis: interval data is utilized for both the predictor and criterion measures; student background characteristics can be controlled; the effects of interventions can be controlled to determine if they mediate the kindergarten screen/achievement relationships; and the relationships can be examined at the pupil-level and the between-school level. If schools differ in the kindergarten screen/achievement relationships or achievement levels of "at risk" pupils, the differences can be investigated by including school characteristics in the analysis.

The application of a multilevel analysis allows for investigation of kindergarten prediction consistent with the hierarchical structure of education. Children are nested within schools. Schools may respond differentially to screening data and may allocate resources differently. An investigation of the predictive validity of kindergarten screening measures is strengthened by examining the relationships at both the pupil-level and the between-school levels.

Summary

The review of literature in Chapter 2 has addressed three main areas: kindergarten screening, variables which may affect prediction; and the most common methodological approaches used in prediction-performance research.

Virtually all screening programs are implemented with the stated purpose to identify children who may experience difficulty learning and who require educational interventions to assist them to learn. Educators and researchers are not of unanimous opinion regarding the value of screening. The primary concern involves the possible negative consequences of "labelling" the child. Lack of screening programs may result in failure to recognize a child at risk of experiencing difficulty learning in school. Failure to provide appropriate intervention may be costly to the child who does not experience success in school, to the school system in terms of retentions, and to society in terms of numbers of school dropouts and subsequent costs for the outcomes brought about by lack of education. Despite the lack of consensus on the value of screening, wide-spread legislation has required the development of screening programs in many school districts in North America.

The particular difficulties in prediction-performance research include: differential developmental patterns of young children making reliable measurement uncertain; the goal of screening is to hypothesize the existence of a problem before the symptoms of the problem are present; there is a proliferation of screening instruments, some of which are not technically adequate; and remedial interventions are implemented to correct or alleviate the predicted problem before the measurement

occurs which is meant to confirm the existence of the predicted problem.

These difficulties are confounded in the research by the fact there is no singular definition of the term "at risk". "At risk" is used to describe pupils performance at screening level and performance at the time of subsequent outcome measure; it is used to identify pupils who obtain a score below a particular cut-off point and it is used generally to describe any pupils who have difficulty for any of a number of social, emotional or physical reasons.

The variables included in a study affect the validity of the study as do the variables which are not included. Many studies have failed to include important pupil characteristics such as age at entry and gender. Failure to include these variables may result in masking findings of a developmental nature related to age, gender or a combination of both and also weaken the power of the analyses.

Numerous methodological approaches have been utilized for prediction-performance research. Appendix Table 1 provides an overview of a selection of prediction-performance research studies. Eighteen studies cover only a short-time frame, end of kindergarten or first grade; five used sample sizes of less than one hundred subjects; seventeen studies failed to consider gender, most of the studies failed to consider chronological age at entry which may be representative of important developmental levels and may interact with gender as an identifier of "risk" status. All the prediction-performance studies reviewed are flawed in at least one major way. Also, they fail to consider the hierarchical nature of educational effects. Children learn within-classes, classes function within schools, schools allocate

resources which affect learning including opportunities for educational interventions. If educational interventions are allocated differentially or if class size, peer group, or instructional practices vary among schools, it is necessary to view the predictive validity of screening measures from a multilevel perspective and look to both within-school and between-school variables to explain the relationships between kindergarten screening measures and outcome measures.

This study is an attempt to fill a gap in the prediction-performance research literature by applying a statistically appropriate hierarchical model and by controlling for pupil's background characteristics and for the effects of interventions which occur during the time under study.

Chapter 3

Research Methodology

Introduction

Kindergarten screening refers to the administration of screening measures, or tests, to kindergarten pupils. The purpose of screening is to discover those pupils who are at risk of experiencing difficulty in learning academic skills. The underlying assumption for identification is that the early provision of educational interventions will eliminate or alleviate the predicted learning difficulties. Chapter 2 is a review of the advantages and disadvantages of screening practices. The primary concern regarding screening practices is the possible negative effects which may result from labelling a child "at risk".

The important variables which might have significant effects on the predicted outcomes merit consideration. The purpose and underlying assumptions for screening include the intention to provide educational intervention to pupils identified as "at risk". Interventions are made available because educators believe they have a positive effect on academic outcomes. Therefore, a study of the predictive validity of screening measures should include controls for effects of the interventions. Failure to control for the effects of interventions may result in low predictive validity, because if the intervention alleviated or eliminated the factors which were predicted to lead to poor performance, and the pupil performed well on the outcome measures, the initial prediction of "at risk" would appear inaccurate.

This study examines the relationship between four kindergarten screening measures and grade three achievement in reading, mathematics,

vocabulary and language. I expect that the relationships between individual screening measures and different outcome scores may vary; some screening measures may be better predictors of particular outcome scores than other screening measures. I also expect that the relationships may vary across schools as the allocation of resources may be different for individual schools.

I expect pupil characteristics of age, gender and physical problems have an effect on the relationship between screening measures and grade three achievement. The normal developmental differences related to age and gender may result in pupils of similar age and gender attaining scores on screening measures which are different from older or younger pupils. Maturation may mediate these initial differences by grade three. When age and gender are controlled, the kindergarten screen/achievement relationships may be weaker. The effects of a physical problem may depress the screening or outcome scores.

I expect the kindergarten screen/achievement relationships will increase after controlling for the effects of educational interventions. If schools respond to the screening information and provide effective instructional intervention to students identified as "at risk", the outcome scores of those at risk would be higher than without intervention. The effect would be to lower the relationship between the screening scores and outcome scores. Therefore, by statistically controlling for this effect through the inclusion of interventions in the model, one would expect the kindergarten screen/achievement relationship to increase.

This study extends the traditional predictive validity approach by

investigating the kindergarten screen/achievement relationships at two levels, the pupil-level and the school-level. This allows for the investigation of between-school differences which may exist in the kindergarten screen/achievement relationships and in the average levels of achievement of pupils who obtained scores at risk on screening measures. The contextual effects of schools may contribute to school differences. I expect that school-level variables will explain some of the between-school differences in the kindergarten screen/achievement relationships and in the average levels of achievement of pupils identified as "at risk".

This chapter describes the research methodology for the study. The chapter begins with a description of the subject population and the procedures for data collection, followed by a brief description of the four kindergarten measures and the outcome measures. The research questions and hypotheses for each question are presented. The analysis is discussed in three sections: the variables used in the study; a description of the application of hierarchical linear modelling; and the preliminary analyses. The last section of the chapter discusses the threats to validity. The chapter concludes with a brief summary.

Subjects

The subjects for this investigation include all pupils enrolled in one Canadian school district who were born in 1975 or 1976, and who were enrolled in the school district in 1987-88. Thus, the subjects represent two age cohorts, not grade cohorts. The school district enrolled approximately 15,000 students with about 1000 students at each grade level. The school district included over 30 elementary schools

serving the municipality. The smallest schools were two-room, rural schools with enrollments of 40-60 students. The largest schools were within the city core, with enrollments ranging from 300-500.

The municipality includes two cities in which there is multi-family housing for low-income families. Large suburban areas accommodate working class and middle-class families. Areas exist where upper-middle and high-income professionals live and many large rural, agricultural farms are part of the community. The population is of mixed SES and several racial and ethnic groups are represented in the community. The municipality is a growing area with a light-industrial and agricultural base. It is located within easy driving distance of a major city.

The 1975 and 1976 cohorts were comprised of 1030 and 1035 students respectively. Subjects were selected who had been administered four screening measures in kindergarten and the Canadian Test of Basic Skills (CTBS) in grade three. One hundred and twenty students who were administered kindergarten measures were enrolled in special class placements and were not administered grade three CTBS in their grade three year. These students were excluded from the study. Attrition of students resulted largely from movement out of the district. The achieved sample included 957 students, 497 in the 1975 cohort and 460 in the 1976 cohort. A discussion of the analysis of attrition bias is presented in the Threats to Validity section of this chapter.

Procedures

The district granted permission to review the pupil records for all children enrolled in 1987-88 who were in grades four through eight, or in special class placement. Each pupil record card (PR Card) was reviewed

to identify all children who were born in 1975 or 1976. The majority of the pupils comprising the 1975 and 1976 cohorts were in grades five and six, respectively, at the time of data collection. Therefore, I was able to obtain records of grade three achievement for most pupils, even those who had repeated one or two grades. Each student was assigned an identification number and pertinent information was recorded from the PR Card: birthdate, retention in grade, extended primary, learning assistance and known medical conditions. The pupil cumulative record of each student was then reviewed for test scores, medical information and educational interventions. All information was recorded on fortran data record sheets.

The following test scores were obtained, from the original protocols:

Kindergarten Screening:

Draw-a-person;

Mann Suiter Visual Motor Test;

Kindergarten Language Screening Test;

Deverell Test of Letters and Numbers.

The following test scores were obtained from computer generated reports of test scores administered in grade 3:

Canadian Tests of Basic Skills:

Test scores (each of four years)

- Vocabulary

- Reading

- Total Language

- Total Mathematics.

Canadian Cognitive Abilities Test:

- Verbal Test
- Non Verbal Test
- Quantitative Test.

School records were reviewed to obtain evidence of the following educational interventions:

- extended primary (retention)
- acceleration
- learning assistance intervention
- speech and language therapy
- E.S.L. instruction
- special class placement.

A data sheet was provided to each learning assistance teacher on which the names of the subject pupils were listed. Learning assistance teachers placed a check in columns labelled with the following individual physical conditions which may have a effect on academic achievement during the primary years:

- history of ear infections
- conductive hearing losses
- hearing impairment
- vision problem (wears glasses)
- allergies
- physical handicap
- chronic illness

(See Appendices A and B for descriptors of physical problems and number of pupils reported to have physical problems.)

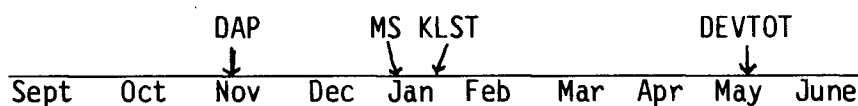
Upon completion of collection of pupil information, data were entered on the University Mainframe computing system. Once the data were entered, a comprehensive data cleaning procedure was employed to ensure the information entered was accurate. The cleaning procedure entailed the computing of frequencies and other basic descriptive statistics for each of the student- and school-level variables. The goal was to check for inconsistencies which might indicate an entry error. The hierarchical linear model statistical program requires that the data files be set up in a particular way. Data collection, computer entry, cleaning and set up of the raw data in preparation to begin analysis required approximately one year.

Instruments

The kindergarten screening measures were four individual tests administered at designated times throughout the school year to screen for exceptionalities in cognitive, language, visual-motor and pre-academic areas. The district provided inservice sessions and written instructions to all kindergarten teachers to ensure standard administration and scoring of each of the instruments. Figure 4 illustrates the administration of the various screening instruments across the school year.

Figure 4

Administration of Kindergarten Screening Measures



Draw-a-Person Test

The kindergarten teachers administered the Draw-a-Person Test (DAP) in November. Teachers administer the test by asking pupils to draw a person as well as they can. The teachers make no suggestions regarding how to complete the figure. The score of the DAP is an indicator of non-verbal cognitive ability. The district selected it to identify the maturation level at which the child was functioning (Harris, 1963). Researchers have demonstrated and discussed extensively the reliability and validity of the Draw-a-Person Test (Dunn, 1967; White, 1979; Naglieri & Maxwell, 1981). Appendix C presents the technical characteristics of the test. The district adapted the scoring from the Goodenough-Harris scale to include the items on the scale which primary and younger children were likely to draw. The district developed local norms from the first administration of the instrument. The test has a total score of 31. A score of 7 or lower was the district's cut-off score for indicating students were "at risk".

Mann-Suiter Visual Motor Screen

The kindergarten teachers administered the Mann Suiter Visual Motor Screen (MS) in January as an indicator of visual perception and fine motor skills. The Mann Suiter is a simple, normed, screening test which consists of copying four geometric figures: circle, square, triangle, and diamond. Successful completion of the four figures represents minimal standards for success in handwriting (Mann, Suiter & McClung, 1987). Appendix D presents the norms for completion of each figure. One mark is given for each correct figure. Two or more errors indicated the student was "at risk".

Kindergarten Language Screening Test

The kindergarten teachers administered the Kindergarten Language Screening Test (KLST) individually to each pupil in January as an indicator of the child's language development. The KLST is a normed screening test which investigates several aspects of language; previous studies have demonstrated its reliability and validity (Gauthier & Madison, 1973). Appendix E presents technical information about the test. It consists of seven parts: a) first and last name, age; b) identification of four colors - red, yellow, blue, green; c) counting with pointing 1-4, 5-10; d) identification of body parts - chin, knee, elbow, ankle; e) three part oral sequential command, knowledge of prepositions; f) sentence repetition; g) spontaneous language sample from a three-picture representation. The total score is 29. A score below 21 indicated "at risk" status.

Deverell Test of Letter and Numbers

The kindergarten teachers administered the Deverell Classification Test at the end of the school year when the children have had a common experience base. The purpose of the test is to measure the child's ability to recognize upper and lower case letters and numerals to 12. Appendix F presents technical information about the test. The district chose this task because researchers have shown consistently that knowledge of letters and numbers is one of the best predictors of academic success (Dykstra, 1967; Jansky & de Hirsch, 1972; Simner, 1982). The total score possible is 64 and a score of less than 56 indicated the student was "at risk".

Canadian Tests of Basic Skills

The Canadian Tests of Basic Skills (CTBS) is a group-administered, norm-referenced achievement test, derived from the Iowa Tests of Basic Skills. Technical characteristics of the tests are presented in Appendix G. The district had administered the CTBS to all pupils in May of grade three, and thereafter, annually through grade seven. Classroom teachers administered the CTBS to the majority of pupils after they had attended 39 months of primary school (including 10 months of kindergarten). Approximately 10 per cent of the students remained in primary grades for four years, and therefore these pupils did not complete the CTBS until after 49 months of primary schooling. The outcome measures in this study were grade three levels of reading, mathematics, vocabulary and language.

Research Questions

The purpose of this study is to examine the relationship between kindergarten screening measures and grade three achievement in reading, mathematics, vocabulary and language. The study also examines the extent to which the relationship between kindergarten screening and grade three achievement is mediated by the provision of learning assistance or extended primary. The study examines four research questions:

1. a) What is the average within-school relationship between grade three test scores in academic achievement and scores on kindergarten screening measures of perceptual-motor, language, and cognitive skills?
- b) To what extent do the relationships between achievement scores and kindergarten screening scores vary across schools?

I hypothesize that for each measure of grade three achievement, the relationship between achievement scores and screening scores will vary across various screening measures; some screening measures may be better predictors of particular outcomes scores than other screening measures. I expect also there will be significant variation across schools in these relationships because schools vary in their allocation of resources to pupils with differing levels of ability. For example, some schools may be more successful in bolstering achievement of low ability pupils than of high ability pupils, or vice versa.

2. a) What is the relationship between grade three achievement and kindergarten screening after controlling for the effects of gender, age at entry to kindergarten, and whether the child has a physical problem?

b) Do the relationships between grade three achievement and kindergarten screening vary across schools after taking account of pupil characteristics?

I expect that the average within-school relationships will be weaker after controlling for pupil's personal characteristics, and that there will be less variation across schools in these relationships. The literature reviewed in Chapter 2 suggested that younger, male children are likely to perform worse on the screening measures. Physical problems may interfere also with the pupils performance. Therefore, when the effects of these characteristics are controlled, I expect the kindergarten screen/achievement relationships to be weaker and the between-school differences in the relationships to decrease. The

decrease in variation between schools will be minimal if the pupil characteristics are distributed equally across schools because the effects of controlling the variable would be similar for all schools.

3. a) To what extent are the relationships between grade three achievement and kindergarten screening mediated by educational interventions of learning assistance or attending extended (4 year) primary schooling?

b) Does the extent to which the relationships are mediated vary across schools?

My hypothesis is that remedial interventions depress the relationships between kindergarten screening and grade three achievement. Therefore, I expect that the average within-school relationship between screening and achievement will be greater after taking account of the effects of learning assistance and extended primary schooling. Because the effects of these interventions may vary across schools, there may be greater variation in kindergarten screen/achievement relationships after removing the effects of the intervention.

4. a) If there is significant variation between schools in their relationships between screening and outcome measures, to what extent can it be explained by school size, rural versus urban location or the school mean and variance of pupils' ability?

b) To what extent are the between-school differences in achievement

explained by various school-level variables?

I expect that if there is significant variation between-schools in their kindergarten screen/achievement relationships, some of the variation may be explained by school size. In smaller schools, low ability pupils may have greater opportunities to benefit from regular instruction, and may have a better chance of receiving remedial instruction. If so, this would result in shallower kindergarten screen/achievement slopes for smaller schools. The same processes may apply in rural schools compared with urban schools. Allocation of resources to pupils of varying ability may also be related to the distributions of ability within and across schools. Several studies have shown that school mean-ability or SES can have an effect on students' outcomes over and above the effects associated with students' individual backgrounds (Willms & Chen, 1989; Brookover et al., 1978; Summers & Wolfe, 1977).

I hypothesize that the school-level variables will explain some achievement differences between-schools because some schools may be more successful in bolstering achievement of low ability pupils. These differences may result from various processes which contribute to pupils' achievement. This study did not include variables for measuring processes such as teaching practices, curriculum coverage or parent and teacher press for academic success (Anderson, 1982). Therefore, the school-level variables in this study may act as proxies for other variables of school processes and provide explanatory power in the analysis (Willms, 1986).

Analysis of the Data

Data

The goal of the analysis is to examine the relationships between kindergarten screening measures and grade three achievement in reading, mathematics, vocabulary and language, and to determine the extent to which the relationships are mediated by educational interventions. The analysis includes examination of the relationships within schools and between schools, and investigates whether the relationships vary across schools.

The within-school variables include kindergarten screening measures, student characteristics, and interventions. They are:

DAP This is a continuous variable measuring the score obtained on the Draw-A-Person Test. The variable was centered on the cut-off score of 7, which indicated "at risk" status.

To center a variable on a particular value, one subtracts that value from each individual's score. After centering, therefore, children with a score of zero on DAP had scored at the "at risk" cut-off score. Those with negative DAP scores scored below the cut-off score, and those with positive scores scored above the cut-off score. Centering facilitates interpretation of the intercepts in the within-school equations (see Willms, 1984).

KLST This is a continuous variable measuring the score obtained on the Kindergarten Language Screening Test. The variable was centered on the cut-off score of 20, which indicated "at risk" status.

MS This is a continuous variable measuring the score obtained on the Mann-Suiter Visual Motor Screening Test. The variable was centered

on the cut-off score of 2, which indicated "at risk" status.

DEVTOT This is a continuous variable measuring the score obtained on the combined subtests of the Devere'll Test of Letters and Numbers. The variable was centered on 55 which indicated "at risk" status.

AGE Age at entry was based on the month born. It was coded in unit increments from -5.5 to 5.5: students who were relatively young for their cohort received negative values (e.g., birth dates in December and November were assigned values of -5.5 and -5.4 respectively); students who relatively were old for their cohort received positive values.

GENDER A dummy variable representing pupil's sex was coded zero for males and one for females.

HANDICAP was a dummy variable representing whether or not the pupil was affected by one or more of the following physical conditions: visual impairment, chronic ear infections, hearing impairment, allergies, physical handicap or chronic illness. A zero indicates no physical impairment present, a one represents the presence of a physical impairment.

LEARNING ASSISTANCE is a dummy variable representing whether or not the pupil received remedial instruction. A zero indicates no assistance was provided; a one indicates the pupil participated in remedial instructional activities.

EXTENDED PRIMARY is a dummy variable representing whether or not the pupil attended four years of primary schooling (after completing kindergarten) to complete grades one to three. A zero indicates the pupil completed the primary school in three years; a one indicates the pupil attended four years of primary schooling.

The between-school variables are:

MEANCCAT This is a continuous variable representing the mean ability score within each school calculated from the grade three level Canadian Cognitive Abilities Test scores of the combined cohorts of pupils. Before aggregating to the school level, the variable was centered on 100, which is the national norm for the test.

SDCCAT This is a continuous variable, the standard deviation of the CCAT scores for the combined cohorts within each school. This is used to represent the heterogeneity of pupil ability within each school.

SCHOOLSIZE This is a continuous variable describing the size of the grade three enrollment.

RURAL This is a dummy variable representing geographic location of the school; it is coded Rural = 0, Urban = 1.

ATTRITION This is a continuous variable representing the effect of attrition within each school. The computation of the attrition variable is described under Attrition Bias.

(See Appendix H for Characteristics of Schools.)

Analyses

The analyses of the relationships between kindergarten screening measures and grade three achievement employ a two-level hierarchical linear regression model (Bryk & Raudenbush, 1987). The model estimates the average within-school relationship for each screening measure with reading, mathematics, vocabulary and language and the extent of variation in these relationships across schools (questions 1a & 1b).

The model is represented by the following equations:

$$\text{Pupil-level: } (\text{Achievement})_{ij} = \beta_{0j} + \beta_{1j} (\text{K-Screen})_{ij} + e_{ij} \quad (1)$$

$$\text{School-level: } \beta_{0j} = \theta_{00} + U_{0j} \quad (2)$$

$$\beta_{1j} = \theta_{10} + U_{1j} \quad (3)$$

where the subscripts $_{ij}$ denote: pupil $i(i=1,2,\dots,n)$ in school j ($j=1,2,\dots,30$).

The first level of the model comprises 30 separate within-school regressions, represented by equation 1. The parameters of interest are β_{0j} and β_{1j} , the intercepts and slopes for the 30 schools. K-Screen refers to one of the kindergarten screening measures, or to a combination of screening measures. Because the screening measures are centered around the cut-off score for "at risk" status, the estimate of β_{0j} for a particular school is an estimate of how well a pupil with a kindergarten screening score at the cut-off would score on the outcome variable. Estimates of β_{1j} specify the outcome/kindergarten screen relationship for each school.

The second level of the model, represented by equations 2 and 3, expresses the β_{0j} and β_{1j} as a grand mean (θ_{00} and θ_{10} respectively) and a school-level residual term (U_{0j} and U_{1j} respectively). This is the simplest school-level model; it contains no school-level variables. An estimate of θ_{0j} therefore represents the average achievement score for the entire district for a pupil with a K-screen score at the cut-off. An estimate of θ_{1j} represents the average achievement/kindergarten screen relationship for the district (Question 1a).

HLM combines the equations and estimates the parameters at both levels simultaneously (Raudenbush & Bryk, 1986). HLM also provides

estimates of the variance of school-level residual terms, that is, $\text{Var}(U_{0j})$ and $\text{Var}(U_{1j})$. In this model, $\text{Var}(\beta_{0j}) = \text{Var}(U_{0j})$, and $\text{Var}(\beta_{1j}) = \text{Var}(U_{1j})$. By examining whether $\text{Var}(\beta_{0j})=0$, one can determine whether the observed differences in adjusted means across schools could have occurred by chance. The examination of whether the $\text{Var}(\beta_{1j})=0$ determines whether the observed differences in the outcome/kindergarten screening relationship across schools could have occurred by chance. A test of this hypothesis addresses question 1b. I expect significant variation across schools in the outcome/kindergarten screening relationship because the schools probably vary in their allocation of resources to pupils with differing levels of ability.

Questions 2a and 2b are addressed by adding variables describing pupil-level characteristics to the first level of the model:

$$\begin{aligned} (\text{Achievement})_{ij} = & \beta_{0j} + \beta_{1j} (\text{K-Screen})_{ij} + \beta_{2j} (\text{Gender})_{ij} + \\ & \beta_{3j} (\text{Age})_{ij} + \beta_{4j} (\text{Handicap})_{ij} + e_{ij} \end{aligned} \quad (4)$$

The second level of the model now includes five equations, similar to Equations 2 and 3, which model the between-school variation in the intercepts and in the four first-level parameters.

In this hierarchical model, the β_{0j} for a particular school is an estimate of how well a pupil with a screening score at the cut-off point would score on the outcome variable after controlling for gender, age on entry to kindergarten and handicapping conditions. I expect that the average within-school relationships would be weaker after controlling for

pupil characteristics, that there would be less variation across schools, and that variation between schools in the outcome/kindergarten screening relationships would be influenced minimally.

To address questions 3a and 3b, two dummy variables are added to the pupil-level model;

$$\begin{aligned} (\text{Achievement})_{ij} = & \beta_{0j} + \beta_{1j} (\text{K-Screen})_{ij} + \beta_{2j} (\text{Gender})_{ij} + \\ & \beta_{3j} (\text{Age})_{ij} + \beta_{4j} (\text{Handicap})_{ij} + \beta_{5j} (\text{Extended} \\ & \text{Prim})_{ij} + \beta_{6j} (\text{Lrn. Asst})_{ij} + e_{ij} \end{aligned} \quad (5)$$

The school-level model now includes seven equations.

The main interest of this model is the size and direction of the average β_{5j} and β_{6j} across schools. If the estimates of these parameters were positive, the results would suggest that those who received remedial interventions scored higher than their peers with comparable scores on the screening measure. If this were the case, I would expect the estimate of β_{1j} for this model to be larger than the estimate for the model given by Equation 4. This would lend support to the hypothesis that remedial interventions have the effect of lowering observed relationships between screening measures and subsequent academic achievement.

To address questions 4a and 4b, school-level variables are added to the model. The school level model regresses the parameters from the student-level (within-school) model on particular school-level variables, such as school mean-ability or school size. The full model includes the

pupil-level equation (Equation 5) and two school-level equations:

$$\beta_{0j} = \theta_{00} + \theta_{01}Z_{1j} + U_{0j} \quad (6)$$

$$\beta_{1j} = \theta_{10} + \theta_{11}Z_{2j} + U_{1j} \quad (7)$$

where Z_{1j} is a school-level variable. (The model can include more than one school-level variable.) θ_{01} and θ_{11} are the regression parameters of interest. They indicate the strength of the effects of the school-level variable on the average levels of achievement and on the outcome/kindergarten screen relationship. The error terms U_{0j} and U_{1j} are school-level error terms. They specify the unique contribution of each school not explained by the school-level variables in the model. I am interested in whether the estimates of the adjusted slopes of outcome/kindergarten screen, β_{1j} , and the estimates of adjusted school performance, β_{0j} , are a function of particular school-level variables.

I expect that the inclusion of one or more school-level variables would explain some of the between-school differences in the outcome/kindergarten screen relationships. This would result in a decrease in the variance of U_{1j} . Similarly, I expect that one or more school-level variables would explain the between-school differences in achievement for students who obtained scores "at risk" on kindergarten screening measures. If this were the case then the variance of the school-level residuals, U_{0j} , would decrease significantly with the addition of the school-level variables.

Preliminary Analyses

The SPSS-X program was used in the following ways:

- a) to prepare the raw data for analysis;
- b) to calculate means and standard deviations of all variables for each cohort and for the combined cohorts and achieved sample;
- c) to compute a Pearson-product-moment correlation matrix of all variables;
- d) to create new variables from raw data, including interaction terms for all kindergarten screening measures with age, gender and intervention variables, to calculate the mean CCAT score for each student; to calculate the mean ability score for each school;
- e) to plot the mean achievement score at each point of the kindergarten screening measures to ensure a linear relationship between the kindergarten screening and outcome measures;
- f) to run ordinary least squares regression to investigate the effects of some variables for selection of variables to be included in hierarchical model analysis, including a model with a quadratic term, to check for linearity; subsequent to these exploratory regressions, physical conditions were combined because individual problems lacked explanatory power in this analysis. Educational interventions of ESL, speech/language therapy and special class placement were dropped as they also lacked explanatory power.

The Tell-A-Graf graphics program was used to generate plots and graphs.

HLM Analysis

Preliminary HLM analyses tested whether regression coefficients

varied significantly from school to school. No significant variation in the effects of age, gender or handicapping conditions occurred and therefore regression coefficients were modelled as fixed effects. Now the β_{2j} , β_{3j} and β_{4j} in Equations 4 and 5 become simply β_2 , β_3 and β_4 ; therefore there are no school-level equations for these parameters. I tested also whether there were significant differences between the cohorts in the adjusted levels of achievement. Differences were small and statistically insignificant, so cohort was dropped from the model. Interaction terms between kindergarten screening measures and age or gender were tested in preliminary analyses. Most of the interaction terms were not significant. When they were significant the effects were small, and thus not included in further analyses. The regression coefficients which varied significantly (intercepts, slopes, and interventions) were modelled as random effects.

School-level variables were included in exploratory models to examine their effects on both kindergarten screen/achievement relationship and achievement. Only school mean ability was found to contribute significantly so it remained in the model. The attrition variable was not significant in preliminary analysis when modelled on both the intercept and the slope, but because it had a consistent, small, negative effect on the slope, it was retained in the equation which modelled variations in slopes.

Threats to Validity

The question of validity is always of concern in predictive research. The quality of the screening and outcome measures, the characteristics of the sample and selection bias may influence the

validity of the findings. Care was given to address some of these concerns in the following ways:

- a) Appendices C-G present the technical information for the kindergarten screening measures and outcome measures;
- b) Characteristics of the sample:
 - i) age cohorts were selected to prevent inflation of performance scores by "over-age" pupils enrolled in grade level;
 - ii) male-female representation is similar for each total cohort and for the achieved sample;
 - iii) mean scores on criterion measures for the total data set and for each cohort are similar and although they are slightly higher than national norms, the cohort and sample means are representative of the district performance during the years the CTBS and CCAT were administered (Appendix Table 21).
- c) Attrition Bias

One concern was that the achieved sample would be biased because it represented a less transient population than those who had not been in the district continuously. A second concern was the loss of the grade-three outcome scores for the 120 pupils in special class placements. Although the kindergarten screening scores for these pupils were included in the total data, there were no CTBS scores for them, and therefore they were not included in the hierarchical analysis.

I was able to estimate the extent of attrition bias by comparing the kindergarten screening scores of pupils in the achieved sample with scores of pupils who were tested in kindergarten, but were not tested in grade three, either because they had left the district, were in special

programs or for some other reason. The comparison was accomplished in the following way. Standard scores were computed for each of the kindergarten screening measures. A principal components analysis was computed on the kindergarten scores to determine the factor loading of the four kindergarten screening measures. The first principal component was used as a composite screening measure. A dummy variable identified pupils with both kindergarten screening scores and outcome scores (coded 1), and those with only kindergarten screening scores (coded 0).

An hierarchical linear regression was run on the entire sample using the following equation:

$$(\text{Composite})_{ij} = \beta_{0j} + \beta_{1j} (\text{Study})_{ij} + e_{ij}$$

The results presented in Appendix Table 2 suggest that the mean scores of the achieved sample were upwardly biased, but the differences were small. This does not mean that attrition would necessarily have affected the observed Kindergarten-screen/achievement slopes. However, the finding did suggest that it would be worthwhile to model explicitly the effects of attrition on slopes to test their significance. Thus a measure of the extent of attrition for each school was included in the school-level equations of the HLM (see Equations 6 and 7). The estimates of the parameter β_1 for each school are used as the Attrition variable.

Summary

This chapter has described the subject sample, research procedures, methodology and threats to validity. Chapter 4 presents the findings of the analyses. Chapter 5 discusses the findings and includes recommendations for future research.

Chapter 4

Findings

Introduction

This chapter presents the findings of the analyses. The first section presents a correlation matrix of the major dependent and independent variables. The second section explains the format of the tables which display the findings of the HLM analyses. The next four sections present the results of the fitted models, Models I to IV, which address the research questions and investigate the relationships between the outcome measures on each of the kindergarten screening measures. The seventh section discusses the estimated parameter variance explained within-schools and between schools. The eighth section presents Model V which extends the investigation by including all four screening measures in the same model. The ninth section presents Model VI, which includes only those variables found to be significant in prior models. Some tables are presented in the text. Tables describing the findings for particular HLM models may be found in the Appendices. The discussion of the findings and recommendations for further research are presented in Chapter 5.

Correlation Matrix

Table 1 shows the means and standard deviations of the pupil-level variables and their correlations. Most of the correlations were statistically significant and relatively low: they range from $-.002$ to $.704^{**}$. (**Indicates correlations are significant at the .01 level) The highest correlations were between the grade three achievement measures, which range from $.589^{**}$ to $.704^{**}$. Correlations between educational

Table 1

Means, Standard Deviations, and Correlations of Student-Level Variables

	DAP	KLST	MS	DEVTOT	READ3	MATH3	VOCAB3	LANG3	AGE	GENDER	HANDI- CAP	EX PRIM	LRN ASST
Means (Centered)	5.05	4.43	6.40	.29	41.45	41.62	41.52	43.89	.150	.005	.086	.093	.095
Standard Deviation	(4.57)	(3.12)	(5.29)	(.80)	(8.61)	(7.40)	(8.17)	(8.33)	(3.43)	(.509)	(.409)	(.291)	(.293)
Total (N=957)													
DAP	1.00												
KLST	.252**	1.00											
DEVTOT	.249**	.374**	1.00										
MS	.280**	.231**	.216**	1.00									
READ3	.292**	.329**	.292**	.200**	1.00								
MATH3	.203**	.296**	.287**	.237**	.651**	1.00							
VOCAB3	.283**	.357**	.297**	.204**	.685**	.589**	1.00						
LANG3	.309**	.346**	.360**	.269**	.699**	.704**	.627**	1.00					
AGE	.220**	.148**	.090**	.164**	.087**	.126**	.118**	.124**	1.00				
GENDER	.231**	.299**	.133**	.058*	.142**	.047	.082**	.196**	-.027	1.00			
HANDICAP	-.047	.025	-.010	.014	.053	.012	.010	-.013	.016	-.002	1.00		
EX.PRIM	-.256**	-.307**	-.434**	-.234**	-.286**	-.290**	-.262**	-.346**	-.088**	-.099**	-.147**	1.00	
LRN ASST	-.168**	-.190**	-.219**	-.176**	-.265**	-.270**	-.263**	-.291**	-.056	.028	.085**	.375**	1.00

Note: * Significant at the .05 level.

**Significant at the .01 level.

interventions and screening and achievement measures were negative and statistically significant. On the kindergarten screening and grade three achievement measures, girls had an advantage over boys, and older students had an advantage over younger students. The correlations between educational interventions with sex and age favored boys and younger students.

Format of the Tables

Appendix Tables 3 to 20 present the results for the regressions of reading, mathematics, vocabulary and language achievement on each of the kindergarten screening measures. The format is essentially the same for all the tables although not all models include all of the variables listed.

The table is divided into three parts. The first section of the top part of the table shows the average within-school equation (see Chapter 3, Equation 1). Parameters were fixed for variables which did not vary across schools in the exploratory analysis. The second section of the top part of the tables shows the effects of the between-school variables. These are estimates of parameters that were allowed to vary across schools, which were included only in Models IV, V, and VI. The middle part of the table shows the estimates of the extent to which the parameters vary across schools. The bottom section of the tables show the maximum likelihood of σ^2 , and the two estimates of R^2 at the pupil and school levels: the total variance explained and the residual parameter variance explained on achievement.

Model I

Kindergarten Screening Measure/Achievement Relationships

The Model I regressions address the questions "What are the average within-school relationships between screening measures and achievement measures?" and "Do the relationships vary across schools?"

Table 2 displays the district and achieved sample means and standard deviations for the grade three test scores. The district statistics were computed on complete data for two cohorts combined, the study sample statistics were computed for subjects having scores on all variables included in the analysis. The means are expressed as grade equivalent measures in months of schooling; for example, a score of 41.4 is equivalent to a grade equivalent of grade 4.14. The tests were taken at the end of grade 3, that is, after 39 months of schooling for most pupils, and after 49 months for those who attended the extended primary program. (See Appendix I for discussion on Grade Equivalent Scores.)

Table 2

District and Sample Means and
Intercepts for Pupils at the Risk Cut-off Score

	Reading	Mathematics	Vocabulary	Language
District Means (2 Cohorts)	41.1	41.2	41.3	43.7
Study Sample Means	41.5	41.7	41.5	43.9
Intercepts "at risk"				
DAP	38.31	39.48	38.62	40.17
KLST	37.52	38.54	37.49	39.77
DEVTOT	38.23	38.38	38.39	39.93
MS	40.64	40.67	40.78	42.62

The achievement scores for reading, mathematics, vocabulary and language were regressed on each kindergarten screening measure. The figures were taken from Model I of Appendix Tables 3 to 18. The intercepts represent the average achievement score for pupils who obtained a score on a screening measure at the cut-off point for "risk" status. The expected average score for all pupils is 39 based on the norms of the CTBS. The intercept scores for pupils obtaining a screening score at the cut-off point reported in Table 2 are near the expected average score. The district means are higher than the expected score thus the average performance of the pupils at the cut-off point for risk status is below the district average.

One reason for the relatively high scores may be the loss of the 120 pupils due to placement in special class. These pupils required intensive intervention in response to their learning difficulties. The scores on achievement measures for these pupils would probably be low and would have lowered the average score for the pupils "at risk" if they had been administered the CTBS.

Two approaches to data selection may also have contributed to the fact that the intercepts are relatively high for pupils identified as "at risk". The pupils lost through attrition had slightly lower kindergarten screening scores than the remaining pupils. If their subsequent achievement was low, the failure to include the scores would have the result of a higher mean score for achievement than if their scores had been included. The selection of the study sample required listwise deletion of subjects. To be included in this study, a pupil had to have

scores for eight test measures. Students who were absent during the testing may have had lower achievement scores than students for whom there were eight scores.

There are at least three other reasons why the scores may be high. Some pupils may have performed poorly on one test only and thus may not have truly been at risk of experiencing learning difficulties. The outcome scores for these pupils may have been relatively high when compared to pupils who had learning problems during their primary years.

Another possibility is that the cut-off score was not the most appropriate score for identifying the true "at risk" pupil. If the cut-off score was set too high, normal measurement error may have resulted in the inclusion of pupils who were not truly at risk. The selection of a lower cut-off score would lower the level of the average performance (intercept).

A third possibility is that the achievement of pupils identified as "at risk" was bolstered by interventions and therefore, their grade three achievement is similar to that of other pupils who had performed acceptably on the screening measures.

Table 3 displays the estimates of the coefficients for all of the within-school relationships between the grade three achievement measures and the kindergarten screening measures. All relationships were statistically significant at the .01 level of significance. (All 16 relationships were statistically significant at the .001 level of significance; the probability of a Type I error is .003 thus I am not concerned about the Type I error rate being inflated because of the large number of statistical tests.)

The metric across kindergarten screening measures is different; therefore, the within-school effects must be interpreted independently for each measure. For example, the coefficient for Reading on Draw-a-person test is .57; this means that each point earned on the Draw-a-person test represents .57 of a month growth in reading. The coefficient for Language on Mann-Suiter test represents 2.57; this means that each point earned on the Mann-Suiter test represents 2.57 months of growth in Language. The number of items on Draw-a-person is 31 and the number of items on the Mann-Suiter is 4, therefore, the interpretation of the coefficients must be considered with regard to the individual measures. The standardized coefficients are presented in parenthesis. These coefficients may be compared across predictor and criterion measures.

Table 3

Estimates of the Effects on Grade Three Achievement of
One Point (or one SD) Kindergarten Screening Measure Score

	DAP (31 items)	KLST (28 items)	DEVTOT (64 items)	MS (4 items)
Reading	.57** (2.60)	.86** (2.68)	.47** (2.49)	1.89** (1.51)
Mathematics	.35** (1.60)	.66** (2.06)	.40** (2.12)	2.29** (1.83)
Vocabulary	.53** (2.42)	.90** (2.81)	.46** (2.43)	1.92** (1.54)
Language	.61** (2.79)	.86** (2.68)	.55** (2.91)	2.57** (2.06)

** Significant at the .01 level.

The relationships between the kindergarten screening measure of language ability (KLST) are strong for all four outcome measures of

achievement. The relationship between the kindergarten screening measure of cognitive ability, the DAP, was stronger for reading, vocabulary and language (.57, .53, .61 resp.) than for math (.35).

The predictive relationships for the test of letters and numbers, the Deverell, are strongest for language, followed by reading and vocabulary and weakest for mathematics. The relationship between the visual motor test, Mann-Suiter, is stronger for language and mathematics than for reading and vocabulary. Because the Mann-Suiter had only four items, the coefficients appear large compared with the other screening measures. One difficulty in interpreting the relationships of the DEVTOT and MS with achievement outcome measures is the ceiling effect for both tests caused by a limited number of items. These two tests did not measure the entire range of performance possible on visual-motor performance or on recognition of letters and numbers. The relationships between these screening measures and achievement might be different if more difficult items had been included which measured the full range of ability with respect to visual-motor performance or letter and number recognition.

The standardized coefficients can be compared. In general, the strongest predictive relationships for all four outcome measures were found for the KLST and the Deverell Test of Letters and Numbers. The weakest predictor is the Mann-Suiter. The low number of items in the Mann-Suiter resulted in ceiling effects which probably limited the predictive power of the test.

The reliability of the estimates of adjusted school mean-achievement for the first model ranged from .4 to .6. The Kindergarten-

screen/achievement relationships did not vary across schools. The slopes could not be estimated reliably; the range was from .022 to .252.

Reliable estimation of the slopes for individual schools would require larger numbers of subjects within schools. In most cases this is not possible because enrollment in elementary schools is small. An alternate approach to increase the reliability would be to collect data on the same schools over a number of years and to estimate the average within-school slope over time (Willms & Raudenbush, 1989).

Model II

Controlling for Pupil Characteristics

The second model (see Appendix Tables 3-18) examines the kindergarten screen/achievement relationships after controlling for the effects of student-level characteristics, and examines the extent to which the relationships vary across schools (Questions 2a and 2b). The second model includes age, gender and physical problems as control variables. The within-school coefficients for age, gender and physical problems were constrained to be identical across schools. The intercepts and the coefficients for kindergarten screening measures were allowed to vary across schools.

The age-at-entry effect ranged from .07 to .25 across the 16 regressions (Model II), with a median value of .18. The effect was statistically significant for mathematics and language across all regressions, and significant for vocabulary in three of the four regressions. Age-at-entry was not a statistically significant predictor of grade three reading scores. An age-at-entry effect of .18 means that for every month a child is older than his or her peers, the average grade

three achievement score is .18 months of schooling higher, after taking account of the child's kindergarten screening score, sex, and whether the child has a handicapping condition. Thus, by grade three there remains, on average, approximately, a two-month gap between those who are relatively young for their grade (i.e., those born in November and December), and those who are relatively old for their grade (i.e., those born in January or February).

The estimated gender effect ranged from .06 to 2.94 across the 16 (Model II) regressions, with a median value of 1.09. The estimates were significant for reading and language on all four kindergarten screening measures and significant for vocabulary in one regression (MS). Gender was not a statistically significant predictor of grade three mathematics, nor of vocabulary in three of the four regressions. A gender effect of 1.09 means that for females, the average grade three achievement score is 1.09 months of schooling higher, after taking account of the child's kindergarten screening score, age, and whether the child has a physical problem. Thus, by grade three, females achieve, on average, more than one month of schooling higher than males with similar characteristics. The gap between males and females is larger, on average, for achievement in reading and language than for math and vocabulary (1-2 months in reading and 2-3 months in language).

If the primary interest of research was specifically age or gender, one would not control for kindergarten screening measures. The unadjusted gap for age and gender would be larger, suggesting that females make greater progress in these areas between kindergarten and grade three.

The estimated effects of physical problems ranged from $-.80$ to $.82$, with a median value of $.04$. The effects were not significant for any measures; however, the estimates for math and language were small and negative, while the estimates for reading and vocabulary were small and positive.

Question 2a asks what the relationships would be after controlling for pupil characteristics. The average within-school relationships of outcomes on kindergarten screening scores ranged from $.40$ to 2.57 in Model I and from $.32$ to 2.29 in Model II. The estimated coefficients all declined minimally, the difference between the Model I and Model II coefficients ranged from $.02$ to $.28$, with a median of $.04$. Although the coefficients declined, they remained statistically significant and nearly as strong as in Model I.

Question 2b asked whether the kindergarten screen/achievement relationships varied across schools after controlling for pupil characteristics. The range of the estimated parameter variance across schools was $.01$ to $.16$. Most of the slopes were found not to vary significantly across schools. The failure to reject the hypothesis of significant variation across schools may be a Type II error. Difference between schools in their within-school slopes could not be reliably estimated because the number of subjects within the school was relatively small at the primary level. The estimated parameter variance was significant for five slopes: reading, vocabulary and language on the Draw-a-Person test ($.04^*$, $.04^*$, $.05^{**}$ respectively), mathematics on KLST ($.08^*$), and language on the Devereil ($.03^*$). As an example, the language with DAP shrunken estimates of slopes were about $-.28$, $-.25$, $-.25$ for the

bottom three schools and .23, .23, .21 for the top three schools.

The significant differences in the relationships of three achievement measures on Draw-a-Person and of mathematics on KLST may indicate differential achievement outcomes across schools for pupils who obtained low scores on Draw-a-Person.

Table 4

Estimated Residual Parameter Variance of Mean Achievement
for Pupils at the Cut-Off Score for Risk Status

	DAP	KLST	DEVTOT	MS
Reading	7.20**	6.84**	5.02**	4.06**
Mathematics	8.40**	8.35**	2.52	3.56**
Vocabulary	5.16**	4.43**	.51	2.26**
Language	12.46**	13.86**	8.81**	6.52**

** Significant at the .01 level.

Table 4 presents the estimates of the residual parameter variance of mean achievement scores for pupils at the cut-off score that indicates risk status on kindergarten screening measures. Fourteen of the sixteen estimates of the parameter variance of achievement varied significantly across schools with a range from .51 to 13.86**. In most cases the variation among schools is large in substantive terms. For example, the estimated parameter variance of achievement (intercepts) for language on DAP was 12.46. The standard deviation of the estimate is therefore about 3.5 (Square Root of 12.46). Therefore, the difference in performance between an average school and a low (or high) scoring school could be as much as six months of schooling.

Model III

Controlling for Educational Interventions

The third model examines the mediating effects of educational interventions on the relationships (Questions 3a and 3b). It allows for the examination of whether the kindergarten screen/achievement relationships are mediated by educational interventions and investigates if the mediating effects vary across schools.

The estimated coefficients for the two interventions are negative, large and statistically significant across all models. The range of the effects of attending an extra year of primary schooling is from -2.22** to -6.74**, with a median effect of -4.16**. The range of the effects of receiving learning assistance is from -4.92** to -6.50**, with a median effect of -5.52**. An extended primary effect of -4.16 means that a pupil who obtained a score at the cut-off point on a kindergarten screening measure and who attended an extra year (four years total) of primary school, scored on average, 4.16 months of schooling below pupils with similar kindergarten screening scores who attended only three years of primary school. Similarly, the pupils who scored at the cut-off score for risk and who received learning assistance, achieved on average, 5.52 months of schooling below similar pupils who did not receive the assistance. This suggests that on average, pupils who received special education interventions had lower achievement scores, even after controlling for their individual pupil characteristics of age, gender and physical problems, than did their peers who obtained similar screening scores but who did not receive special educational interventions.

There are at least three possible explanations for the large,

significant, negative effects. One is that the interventions were ineffective or insufficient. Second, placement in intervention programs and subsequent progress may have been influenced by factors not included in the model. Third, opportunities to receive assistance may have been determined by factors other than screening information and may have occurred later for some pupils than for others and thus, the effects may vary for different children. These possibilities are discussed in greater depth in Chapter 5.

After controlling for the effects of educational interventions, the effects of age and gender declined minimally. The range of the estimated effects for age was .06 to .21*. Age effects remained significant for mathematics and language on all four kindergarten screening measures and for vocabulary on the Deverell and Mann-Suiter. The range of the estimated effects for gender was 1.29* to 2.41**. The gender effects remained significant for language on all four screening measures and for reading with the Deverell and Mann-Suiter. This means that advantages in achievement for older pupils and for females described under Model II findings remain even after controlling for the effects of special educational interventions.

The average within-school estimated coefficients for the relationships between outcome measures and kindergarten screening measures decreased across all measures. Table 5 presents the estimated coefficients for the outcome measures on screening measures from Models II and III.

Table 5

Estimated Coefficients for Kindergarten Screen/Achievement Relationships

	DAP		KLST		DEVTOT		MS	
	Models II	III	II	III	II	III	II	III
Reading	.51**	.39**	.80**	.61**	.44**	.33**	1.70**	.98*
Mathematics	.32**	.21**	.64**	.48**	.38**	.26**	2.16**	1.56**
Vocabulary	.50**	.40**	.87**	.72**	.44**	.34**	1.73**	1.16**
Language	.52**	.37**	.76**	.54**	.51**	.36**	2.29**	1.52**

* Significant at the .05 level.

** Significant at the .01 level.

Every estimated coefficient declined. The range of the estimated coefficients in Model II was .32** to 2.29**, with a median of .52**. The range of the estimated coefficients in Model III was .21** to 1.56**, with a median of .48**. This means that the estimated coefficients of the relationships between achievement measures on screening measures were lower after controlling for the effects of educational interventions.

I hypothesized that participation in special educational interventions would result in improvement of achievement scores for students identified as "at risk" on kindergarten screening measures during the time under study. Higher achievement scores for pupils with low scores on screening measures would have the effect of flattening the slope; that is, lowering the estimated coefficient. If this were the case, when the effects of the educational interventions were controlled, the slope would become steeper; that is, the estimated coefficient would increase. The results of Model III indicate that the slopes were

mediated by the educational interventions, but in the opposite direction of my hypothesis. Controlling for interventions decreased the estimated coefficients for the outcome on screening relationships. This means that the predictive power of the screening measures decreased after controlling for the effects of the interventions.

Differences between schools in the estimated parameter variance for achievement for pupils with screening scores at the cut-off point for risk status decreased after controlling for educational interventions. The estimated variances between schools remained significant for reading on DAP (4.33*), and MS (2.73**), and for language on DAP (10.41**), KLST (9.98**) and MS (6.23**). This means that the average achievement of pupils who received scores at the cut-off score for risk status on kindergarten screening measures remained different across schools even after controlling for the effects of educational interventions. The slopes did not vary across schools.

The estimated parameter variances for extended primary and for learning assistance were significant for language on KLST (13.86*) and for reading on Draw-a-person (13.49*). This means that the effects of these educational interventions differed significantly among schools for the pupils who obtained scores at the cut-off point for risk status on the KLST and Draw-a-person.

Model IV

School-Level Variables

The fourth model allows for the examination of whether differences between schools in the kindergarten screen/achievement relationships or between-school differences in achievement can be explained by

school-level variables (Questions 4a and 4b). The primary interest of this study was the relationships between kindergarten screening measure and grade three achievement. The research questions explored the extent of the relationships, the effects of pupil characteristics and educational interventions on the relationships, and how the relationships varied across the schools. The results of the analyses indicated the kindergarten screen/achievement relationships did not vary, however, the average achievement of pupils who scored at the cut-off point for risk status varied significantly across schools, even after controlling for pupil characteristics and the effects of educational interventions.

The collective properties of a school which have an effect on individual pupil achievement over and above characteristics or attributes the pupils bring to the learning situation are called contextual effects (Willms, 1986). Researchers have shown that school mean-ability has an effect on pupils' academic achievement, even after controlling for the individual effects of pupils' family background (Summers & Wolfe, 1977; Henderson, Mieszkowski & Sauvageau, 1978; Brookover et al., 1978; Rutter, et al., 1979; Willms, 1986). Willms and Raudenbush (1989) noted that in studies which fail to include a wide range of variables that describe school policies and practices, an aggregate variable of pupil-level characteristics may act as a proxy for variables describing policies and practices. In this study, school mean-ability may act as a proxy for other variables not available for investigation.

Model IV includes two school-level variables, school mean-ability modelled only on the intercepts and attrition modelled on the kindergarten screening/achievement relationship slope. During

preliminary analysis school mean-ability was modelled on the slope but was found not to be significant. There were no between-school differences in the kindergarten screening/achievement relationships, and therefore, school mean-ability was not modelled on the slopes in the final model. The attrition variable remained in the model because it had a small negative effect on the slope. As the primary interest of this study is the kindergarten screening/achievement relationship (slope), it was appropriate to include this negative effect which represented the attrition bias, even though the effects were not significant.

The age and gender effects declined minimally, or remained the same as in Model III. The age effect remained small but significant for Mathematics and Language on all four screening measures and for Vocabulary with DEVTOT and MS. The range of the age effects was .08 to .23**. The gender effect remained significant for Language on all four screening measures and for Reading on DEVTOT and MS. The range of the gender effects was -.63 to 2.32**. These effects may be interpreted as described in Model II.

The estimated coefficients for the effects of interventions remained significant, large and negative across the 16 regressions. The range of the coefficients for extended primary was -2.25* to -6.65**. The range of the coefficients for learning assistance was -4.33** to -5.81**. These coefficients also did not change appreciably, and may be interpreted as described in Model III.

The estimated coefficients for school mean-ability modelled on achievement are statistically significant across all models. The range of the coefficients is .30** to .46**, with the median .34**. This

suggests that pupils in schools with higher school mean-ability had higher achievement scores than similar pupils in schools with low school mean-ability.

After controlling for school mean-ability, the estimated parameters of variance for mean achievement for pupils at the cut-off point for risk status remained significant for half the models: reading on DEVTOT (3.48*); mathematics on DAP (3.76**), KLST (5.36*), and DEVTOT (3.58**); language on DAP (6.57**), KLST (6.40**), DEVTOT (3.48*), and MS (7.77**). The estimated parameter variance for the remaining models declined to a level which could have occurred by chance. This means that differences between schools in the mean achievement of pupils at risk on certain kindergarten screening measures was lowered by controlling for school mean-ability for eight of the models. This suggests that pupils who scored at the cut-off score for risk status earned higher achievement scores on average in schools with high school mean-ability than similar pupils in schools with low school mean-ability. For example, the mean language achievement of pupils "at risk" within the highest school was about five months of schooling ahead of the mean achievement of similar pupils in the lowest school, even after controlling for school mean-ability.

Parameter Variance Explained

The estimated variance in grade three achievement at the pupil-and school-levels, based on a "null" model (i.e. without any pupil or school-level variables in the model), are as follows:

Table 6

"Null-Model" Estimates of Variance in Grade Three Achievement

	Pupil-Level	School-Level		Total	OLS
Reading	69.13	5.49	=	74.62	74.13
Mathematics	50.89	4.28	=	55.17	54.76
Vocabulary	63.69	3.58	=	67.27	66.75
Language	62.64	7.85	=	70.49	69.39

Thus about 3 to 8 percent of the variance in achievement is between schools.

R-squared values for the various models can be expressed in a number of ways. One statistic of interest is the proportion of the total variance explained by the inclusion of all variables in the model; which

I will denote R^2_T :
$$R^2_T = \frac{\text{Var}(Y) - \hat{\sigma}^2}{\text{Var}(Y)}$$

where $\hat{\sigma}^2$ is the maximum likelihood estimate of the residuals. Estimates of the proportion of (null model) pupil-level variance explained by pupil-level variables, and (null model) school-level variance explained by school-level variables, are also of interest. I will represent them as R^2_p and R^2_s :

$$R^2_p = \frac{\hat{\sigma}_0^2 - \hat{\sigma}^2}{\hat{\sigma}_0^2} \qquad R^2_s = \frac{T_0 - T}{T_0}$$

where σ_0^2 and T_0 are the estimates of pupil- and school-level variances for the null model, and σ^2 and T are estimates of variance after inclusion of relevant variables (e.g. Zuzovsky & Aitkin, in press). In models where all pupil-level variables are fixed, these estimates are

straight forward. However, when within-school regression slopes are allowed to vary, the covariance between school-level variables and within-school slopes complicates the estimates of both R^2_p and R^2_s . In some cases, $T > T_0$, yielding a negative value of R^2_s . In this study the principal measures of interest are R^2_T for the models that include only the kindergarten screening measures, and R^2_s for the Model IV regressions, which include school-level variables. The Appendix Tables include the proportion of total variance explained for all models and the proportion of school-level variances explained for adjusted levels of achievement for Model IV.

The range of pupil-level variance explained by Model I is 7.47% to 20.47%. These R^2 are slightly larger than the R^2 observed using simple OLS regression which would be equivalent to the squares of the inter-correlations between the screening measures and outcome measures shown in Table 1 (which range from 5.48% to 18.84%).

Model V

Four Kindergarten Screening Measures in the Model

The fifth model allows for the examination of whether the relationships between achievement outcome measures and kindergarten screening measures change when all four kindergarten screening measures are included in the model. Appendix Table 19 presents the results of the regressions for each of the four outcome measures on all of the kindergarten screening measures, controlling for the other variables which have been introduced in prior models.

The estimated coefficients for the average within-school relationships of outcomes on kindergarten screening scores are lower with all four

kindergarten measures as covariates in the model than when only one kindergarten screening measure is included. The coefficients are lower because of the inter-correlations amongst the screening measures; the predictive power of the screening measures, in a sense, is shared across the four measures. The average within-school relationships of outcomes on kindergarten screening scores ranged from .09 to .30** for DAP, .34** to .55** for KLST, .18** to .23** for the Devereell and .26 to 1.10** for the Mann-Suiter. The estimated coefficients for all four outcome measures on the KLST and DEVTOT remained significant. The estimated coefficients for reading, vocabulary and language were significant on DAP while only mathematics and language were significant on the Mann-Suiter. This means that thirteen of the sixteen relationships between achievement and each kindergarten screening measure remained positive and statistically significant, even after controlling for all other variables and the other kindergarten screening measures.

The average within-school achievement of pupils who obtained scores on all four kindergarten screening measures at the cut-off point for risk status are all lower than for models including only one kindergarten screening measure. Table 7 presents the mean achievement scores for the district, for pupils who scored at the cut-off point for risk status on one kindergarten screening measure and for pupils who scored at the cut-off score on all four kindergarten screening measures.

Table 7
Average Achievement Scores for
Pupils Who Scored at the Cut-off Point for "Risk" Status

	Reading	Mathematics	Vocabulary	Language
Study Sample	41.45	41.67	41.52	43.89
K-Screen Cut-off Point Mean				
DAP	37.83	39.16	38.16	39.85
KLST	37.21	38.06	37.06	39.45
DEVTOT	37.72	38.17	37.98	39.36
MS	39.55	39.71	39.78	41.29
Four K-Screen at cut-off point	34.56	36.58	34.81	36.98

An examination of Table 7 indicates that pupils who obtained scores at the cut-off point for risk status on all four kindergarten screening measures were from five to seven months of schooling behind the average pupil after controlling for pupil characteristics, educational interventions and contextual effects of school mean ability.

With all four kindergarten screening measures in the model, no age effects reached a level of significance. The estimated coefficients of the gender effect for three outcomes were not significant, but the gender effect for language remained significant at 1.27**. Thus, by grade three, females achieve on average, more than one month higher than males, even after controlling for the kindergarten screening scores, pupil characteristics, educational interventions and school mean- ability.

The estimated effects of the educational interventions remained

negative; six of the eight coefficients are statistically significant. The effect of extended primary for mathematics was -2.21^{**} and for language -3.59^{**} . The range of the effects for receiving learning assistance was from -4.04^{**} to -4.86^{**} . An extended primary effect of -3.59 means that a pupil who obtained four kindergarten screening scores at the cut-off point for risk status and who attended 49 months of primary school, scored on average, 3.59 months of schooling below pupils with similar kindergarten screening scores but who had attended only 39 months of primary school. Similarly the pupils at risk on four kindergarten screening measures who received learning assistance achieved, on average, four and a half months of schooling below similar students who did not receive assistance. This finding again suggests that pupils who obtained scores at the cut-off for risk status, on average, who received special education interventions had lower achievement scores, even after controlling for their pupil characteristics, gender and physical problems, and school mean-ability than did their peers who obtained similar screening scores but who did not receive special education interventions.

The estimated coefficients for school mean-ability modelled on achievement were statistically significant across all models. The range of the estimated coefficients is $.30^{**}$ to $.42^{**}$. This suggests that pupils in schools with higher school mean-ability had higher achievement scores than similar pupils in schools with low school mean-ability. Pupils who scored at risk on the four measures and attended a low ability school would be about six to nine months behind their peers who attend a high ability school. The estimated variance of mean achievement for

pupils at the cut-off score for risk status on all four kindergarten screening measures, remained significant for reading (4.86*), mathematics (9.52**), and language (10.17**).

Model VI

Simplified Models Including Only Significant Variables

Appendix Table 20 presents the results of regressions of the four outcome measures modelled on the four kindergarten screening measures and other variables which had a significant effect in Model V. The final model estimates reported indicate there are only minor differences in the estimates when compared with Model V. There are no substantive changes which require presentation.

Summary

This chapter has presented the findings of six HLM regression models. The discussion of the findings and recommendations for further research are presented in Chapter 5.

Chapter 5

Summary and Conclusions

The final chapter presents an overview of the study, conclusions and discussion of the findings for the research questions addressed in the study. It also presents implications for the study and suggestions for future research. This study fills a gap in the literature on prediction-performance research by examining the relationships between kindergarten screening measures and achievement outcome measures in a hierarchical model. The analysis includes control for important pupil characteristics and the effects of educational interventions during the time of the study.

Overview of the Study

This study examines the relationships between kindergarten screening measures and grade three achievement in reading, mathematics, vocabulary and language for two entire age cohorts enrolled in 30 schools in one school district. The analysis employs a two-level hierarchical linear model to estimate the average within-school relationship between kindergarten screening measures and achievement. It also determines whether significant differences exist between the 30 elementary schools in the relationships between the screening measures and the achievement measures, and examines the extent to which educational interventions mediate the relationships.

The intention of the screening process in this school district was to identify children with handicaps or developmental delays. The presence of handicaps or developmental delays were considered to indicate the need for special education programming for the pupil's optimal educational

progress. The district implemented kindergarten screening in the belief the findings would result in: earlier identification of pupils "at risk" of experiencing difficulty learning; and provision of remedial programming which would alleviate or eliminate the difficulties.

I expected that educational interventions provided for pupils designated "at risk" would improve their achievement. The effect of improving achievement of pupils designated "at risk" would lower the correlation between the screening score and subsequent achievement. This would "mask" the initial prediction of risk. I also expected that schools might allocate resources differentially, and thus, some schools might be more effective than others in bolstering the achievement of at risk students. These achievement gains would have the effect of lowering the relationship between the screening measure and achievement in some schools which would result in the relationships varying between schools.

Principal Findings of the Study

1. Positive relationships were found for all screening measures with all outcome measures.

The findings of this study support that positive relationships exist between these kindergarten screening measures and grade three achievement. The relationships between screening scores and achievement scores varied across screening measures. In other words, these particular kindergarten screening measures were moderate predictors of subsequent achievement in particular skills during grade three. The best predictors for reading, vocabulary and language, language related achievement measures, were the KLST and DEVTOT, also language related measures. The strongest predictors for mathematics were visual-motor ability and expressive language.

The strong relationship between language ability (KLST) and language related achievement scores is not surprising. The assumption that proficiency in oral language underlies academic achievement is logical. Most primary curriculum materials rely heavily on skill in oral language. The recent interest in "whole language" approach to reading attests to the fact that many educators believe oral language is essential for academic success.

Recent research has questioned the strength of the relationships between oral language and achievement (Hammill & McNutt, 1980; Gray, Saski, McEntire & Larson, 1980). Hammill and McNutt (1980) conducted a review of literature which synthesized the results of 89 correlational studies. Studies were selected which illustrated the common belief that a child must have adequate oral language to learn to read. Hammill and McNutt concluded that oral expressive language is not related to reading performance although they suggested that some aspects of oral receptive language are minimally related.

The results of the present study controvert the findings of Hammill and McNutt (1980). There are several possible explanations for the contradictory findings including: different language tests may measure different aspects of language; tests administered to pupils at different ages may result in contradictions as acquisition of language skills is uneven in young children; language tests are correlated with measures of intelligence and there may be confounding effects when intelligence level is not included as a covariate.

The findings of this study suggest there is a strong positive relationship between performance on a measure of oral language in

kindergarten and achievement in grade three. This may illustrate that proficiency in oral language is important for enabling children to interact with the school environment. Oral language proficiency allows for opportunities to interact with the teacher and classmates which may facilitate further language development. Proficiency in oral language may reflect underlying abilities to perform with curriculum materials which rely heavily on language skills in the primary grades, and thus the language measure would be an appropriate measure for identifying students at risk of experiencing difficulty learning academic skills in school.

The strong relationship between achievement in mathematics and visual-motor ability is consistent with prior research in which visual-spatial abilities have correlated highly with performance in mathematics. The strong relationship between mathematics and oral language may be a reflection of the heavy emphasis on language in the presentation of basic mathematical concepts in primary grades.

I hypothesized that the kindergarten screen/achievement relationships would vary across schools as I expected that schools would allocate resources differently, and thus, some schools would be more successful in bolstering the achievement of low ability pupils. The findings for only one of the 16 screening/outcome relationships lends support for this hypothesis. The relationships for language skills with the kindergarten measure of cognitive ability varied significantly across schools. The skills measured by the language test include spelling, grammar and punctuation; skills which are highly dependent upon school learning. This finding indicates that some schools were better at developing skills measured by the language test for pupils who obtained similar scores on the

test of cognitive ability (DAP) than other schools. In general, however, the findings do not support the hypothesis. The strength of the relationships varied across the achievement measures with each of the kindergarten screening measures.

When all four screening measures are included in the analysis as covariates, most of the estimated coefficients for all four achievement outcomes with the screening measures remain significant. The exceptions are mathematics with the test of cognitive ability and reading and vocabulary with the visual-motor test.

2. The kindergarten screen/achievement relationships declined minimally, but remained significant, after controlling for the effects of individual pupil characteristics of age, gender and physical problems.

The addition of variables for student characteristics of age, gender and physical problems resulted in the kindergarten screen/achievement relationships declining minimally across all measures. The kindergarten screening relationships remained statistically significant. The hypothesis that the average within-school relationships would be weaker after controlling for pupil characteristics was supported. The second hypothesis was supported only by the decline in the one relationship which was significant (DAP/Language).

a) Effects of age-at-entry.

Prior to controlling for the effects of educational interventions, the effects of age-at-entry were significant for mathematics and language with all four screening measures and for vocabulary with the tests of knowledge of letters and numbers and visual-motor abilities. After controlling for the interventions, the effects of age-at-entry were not significant for any

of the achievement outcome measures.

There were significant age effects for mathematics and language with all four screening measures and for vocabulary with the test of letters and numbers and the visual motor test. This finding is consistent with research which consistently supports that older students perform better in early grades than younger children (Davis, Trumble & Vincent, 1980). This finding may have implications for making a decision regarding entry to school or providing educational intervention. Gredler (1980) pointed out that older children arrive at school knowing more or having experienced more than younger children. In some ways, therefore, they are more "ready" for learning and may appear to learn more if achievement at a single point in time is the criterion. The difference in achievement between younger and older students apparent at the grade three level declines continuously until it is no longer distinguishable by grade eight (Davis, Trimble & Vincent, 1980).

b) Effects of gender.

Gender may be influential in the prediction of risk status for learning academic skills with boys appearing to be more vulnerable than girls. Before controlling for educational interventions, gender effects favoring girls were significant for reading and language with all four screening measures. The gender effects were particularly large for the relationships between language achievement - spelling, grammar and punctuation, and kindergarten screening measures. This finding is consistent with prior research which suggests that, generally speaking, girls perform better than boys of the same chronological age on readiness tests and on later achievement (Beattie, 1970).

When considering an individual pupil's score on a kindergarten screening measure, both the chronological age and gender of the child should be given consideration. It is possible that an "at risk" score for an older child may be more meaningful than for a younger child.

c) Effects of physical problems.

The effects of physical problems were not significant for any kindergarten screen/achievement models.

3. The effects of the educational interventions of extended primary and learning assistance were large, significant and negative for all achievement outcomes with all kindergarten screening measures.

One can imagine similar analyses for investigating the effects of educational interventions. In such an analysis, the kindergarten screening measures would serve as control variables to adjust for differences between pupils who did or did not receive an intervention. The variables describing interventions used in this study would be inadequate for that purpose. If the primary interest of a study was to examine the effects of different intervention approaches or to examine the benefits for subjects within categorical groups, one would need to collect information such as the location of the intervention, time actively participating or duration of remedial instruction. If the investigation of the effects of the interventions was the purpose of a study, one might estimate the likelihood of being placed in a remedial program, given the pupil's kindergarten screen score.

In this study, the primary interest is the kindergarten screen/achievement relationship and the extent to which educational interventions mediate the relationship. For this reason, the educational interventions

were used only as control variables and are adequate for that purpose.

The hypothesis that educational interventions would mediate the kindergarten screen/achievement relationships was not supported; in fact, the effects were in the opposite direction. I hypothesized that the relationship would be stronger after controlling for the effects of learning assistance and retention because I expected that the interventions would be provided to students obtaining "at risk" scores on screening measures and would improve their subsequent achievement.

There are at least three possible explanations for the large, significant, negative effects. One is that the interventions were ineffective; that is, pupils who attended the special programs progressed at a slower rate than pupils with comparable ability in the regular program or the interventions were ineffective in developing the skills measured by the CTBS. Additionally, one could interpret that the interventions were detrimental to pupil progress and the effects were to help "fulfill the prophecy" made by the screening measures. Common educational practices such as decreasing the pace of instruction, grouping by ability, or lowered teacher expectations may have contributed to lower achievement scores for pupils participating in interventions (Slavin, 1987; Peterson, 1989).

The remaining two explanations concern model specification. One possibility is that pupils were assigned to special programs on the basis of low screening scores, but progress thereafter depended mainly on other factors, such as family socioeconomic status, that were not included in the model. The other possibility is that the assignment of pupils was not based solely on the screening information. Other factors, such as pupil behavior, attention or work completion, may have played a key role in these

decisions. Also, opportunities to receive assistance may have been provided later in the primary grades for some pupils with the decision based on actual achievement rather than on ability. The model did not include variables which represented these factors.

Scatterplots of the achievement data against the screening data suggest that the latter explanation is plausible; there were many pupils with low screening scores who did not receive intervention, and several with high screening scores who did (Appendices J & K).

Based on this study it would be inappropriate to suggest that the kindergarten screening was ineffective. The purpose of the screening was to identify students for whom educational interventions would be required for optimal progress. One hundred and twenty pupils who participated in screening were in special classes and their achievement in grade three was not measured by CTBS testing. Inclusion of this group in the data might have improved the correlations somewhat as these children were significantly at risk and required intensive interventions to make educational progress, their kindergarten screening and subsequent achievement scores probably would have been low.

The hierarchical model used in this study provides a means to study variation between schools in the effects and application of educational interventions. Most of the effects of interventions did not vary significantly across schools. The exceptions are for Learning Assistance on the relationships between reading and measures of cognitive ability, language and letters and numbers, and relationships between reading and the visual motor test; and Extended Primary Schooling on the relationship between language achievement with kindergarten oral language and reading

with the test of letters and numbers. These findings may indicate that interventions are requested differentially in response to screening information or to other variables, the interventions are differentially effective, or a selection process contributed to the effects particularly where the achievement measure is reading. A more in-depth study of the individual interventions might clarify the between-school variation in these effects.

4. The adjusted achievement levels of pupils "at risk" varied significantly among schools.

The adjusted achievement levels of pupils who obtained scores on one kindergarten screening measure at the cut-off point for risk status were lower than the district mean performance but not lower than the average expected performance for the grade based on the test norms. However, the adjusted achievement levels for pupils who obtained scores on four kindergarten screening measures at the cut-off point were considerably below expectancy for the grade placement.

The mean adjusted achievement levels of pupils who obtained scores on one kindergarten screening measure at the cut-off point for risk status varied significantly across schools for language with all four kindergarten screening measures, for mathematics with kindergarten tests of cognitive ability, expressive language and letters and numbers and for reading with the test of letters and numbers. This study did not attempt to explain why the achievement of pupils designated "at risk" varied significantly across schools; however, the model could be extended to include variables describing school policy and practice that might explain these differences. For example, the inclusion of variables such as the

performance on tests during the primary years, heterogeneity of classes, or teacher observations of behaviors, might illustrate policies and practices which differ between schools. Qualitative methods could then be employed for intensive study of particularly effective or ineffective schools.

5. Pupils who scored "at risk" on kindergarten screening measures performed better on all four achievement measures in schools with high school mean-ability than similar pupils in schools with low school mean-ability.

Although it was not the primary purpose of this study, school-level variables were added to the model in an attempt to explain the between-school differences in the average achievement of pupil's designated at risk. Preliminary analysis determined that school size, geographic location and heterogeneity did not contribute significantly to either the achievement differences (intercept) or the kindergarten screen/achievement relationships (slopes).

School mean-ability was found to have a positive effect on the within-school average achievement across all models. This indicated that pupil's who scored "at risk" on kindergarten screening measures performed better on all four achievement measures in schools with high school mean-ability than similar pupils in schools with low school mean-ability.

Because this study did not include variables for school or home processes which may contribute to differences in achievement, school mean-ability may act as a proxy for many other variables. This study did not attempt to identify variables which may be represented by school mean-ability but the amount of variance explained by school mean-ability suggest a strong argument for contextual effects.

6. The kindergarten screen/achievement relationships were not improved by controlling for the effects of interventions.

A major interest of this study was the application of a statistical procedure which would simultaneously analyze the data at two levels. I hypothesized that a major contributor to low correlations between screening measures and achievement outcomes was failure on the part of researchers to control for important student characteristics, the effects of interventions and between school differences. Although this study found significant positive effects of student characteristics and significant achievement differences between schools, it failed to demonstrate that controlling for interventions would improve the kindergarten screen/achievement relationships.

One possible explanation is that the choice of kindergarten screening measures and their relationships with subsequent achievement was not based on theory. In other words, the specific skills and abilities measured by the kindergarten measures may not be the most important prerequisites for the development of the skills measured on the achievement measures.

Another possible explanation is that these particular measures were not intended to be interpreted individually. The intent of administering four measures at different time points across the school year was to identify and intervene for children scoring "at risk" after each measure. Referral for more intensive diagnostic assessment was deemed appropriate only after a child scored at risk on two or more screening measures. The composite model demonstrated that pupils who scored at risk on all four measures and were administered the grade three CTBS, obtained, on average, much lower achievement scores than students designated at risk on

individual measures. This supports an argument for on-going assessment. Pupils who scored low on measures of different abilities over time were more likely to have difficulty or achieve at lower levels than pupils who were low on any one measure.

A consideration in examining predictive validity is that different analysis may be appropriate for different purposes. Predictive validity which examines the relationship of screening measures to subsequent achievement across the entire distribution of abilities may be particularly valuable for developing theories or for understanding differences in performances for pupils with various levels of ability. The addition of controlling variables provides important information for differences in performance related to age and gender. By definition, predictive validity implies time between the predictive measure and the outcome measure; it is important to control for important variables which may influence the outcomes.

The purpose for screening always entails intervention, therefore, controls for the effects of interventions are essential to understand the kindergarten screen/achievement relationships. If the large, negative effects of interventions found in this study are generalizable to other studies which did not control for interventions, the previously reported low correlations between kindergarten screening and achievement may have been inflated rather than deflated.

A second consideration involves predictive utility. School districts which choose to implement a screening program may be more interested in data analysis which identifies the proportion of pupils correctly identified as "at risk" and the proportion of individuals "not at risk" who

are correctly excluded from further assessment or intervention. Test sensitivity is the proportion of pupils with special needs who are identified accurately, and specificity is the proportion of pupils not in need of special services whose scores are above the cut-off score on the screening measure. The proportions of the correct classifications are inversely related. That is, by adjusting cut-off scores, the proportion of pupils identified as "at risk" can be increased although, some of the pupils are not truly "at risk" of school difficulties. By raising the cut-off score, more pupils will require interventions, thus some pupils not in need of special services will receive them. An increase in one group results in a decrease in the other. Prediction-performance matrices and classificational analysis are used to illustrate numerical data and calculate sensitivity and specificity. Provision of intensive assessment and intervention have cost implications, therefore, this type of information could be valuable in the decision for selecting cut-off scores and planning budgets for services.

Appendix Table 22 presents the data used in this study reported in proportions such as those presented in prediction-performance matrices. Extension of the analysis to identify the proportions of young children and representation by sex in the group classified as "at risk", illustrates how these variables could be considered using this technique. The most interesting finding relates to the proportion of pupils who obtained scores within the "at risk" category who received learning assistance or attended extended primary.

The overall "hit rate" of these measures was as high as expected from the research of kindergarten prediction, 57 to 75 percent. The proportion

of pupils scoring at risk on screening measures who participated in an educational intervention ranged from 12 percent to 49 percent. In other words, half of the pupils who were identified "at risk" on a kindergarten screening measure were not provided with an educational intervention intended to have a remedial effect on their performance.

Males made up the highest proportion of the "at risk" group (51-71 percent across models) and young pupils were over-represented in several cases (50-64 percent). These figures lead to some interesting questions regarding how, when and why teachers respond to kindergarten screening measure results: What response was provided to a pupil achieving an "at risk" score on a screening measure? When did a response to an "at risk" score result in referral for an intervention? What factors determined that a pupil would participate in learning assistance or extended primary?

If the large, negative effects of educational intervention found in this study indicate that pupils identified at risk attain higher achievement scores if they do not participate in learning assistance or extended primary, the findings could have significant financial implications for school districts.

Appendix K presents sample graphic representations of the predictive utility of a kindergarten screening measure with an outcome measure. The cut-off score used for this study is indicated on the graph. By adjusting the cut-off score up or down, the proportion of correct and incorrect decisions can be illustrated. Or, in other words, the proportion of pupils identified as requiring remedial intervention services can be increased or decreased by adjusting the cut-off score of the screening measure. As all measurement contains error, it can be expected that some pupils will always

be misclassified. The decision faced by school districts selecting cut-off scores is, which "misclassified" group will be the largest. Will services be provided to a larger group which includes some pupils not truly in need, or will services be provided to a smaller group, denying some pupils services who are truly in need?

In light of the findings of this study, the more important decisions may be whether to provide interventions, and if so, what kind of interventions and for whom.

Limitations of the Study

The major limitation of this study is that the prediction measures and the outcome measures are limited to results of standardized instruments. The appropriate use of standardized tests with young children has recently been under close scrutiny by educators and other professionals. The findings of this study indicate positive relationships exist between these kindergarten screening measures and the achievement outcomes.

The NAEYC (1988) recommended that, "The purpose of testing must be to improve services for children and ensure that children benefit from their educational experiences" (p.44). Emphasis is placed on standardized tests as only one of multiple sources of assessment information that should be used when decisions are made about what is best for young children. Perhaps, the major problem with predictive validity studies has been failure to include, in addition to standardized test results, measures of the many possible sources which influence academic performance over time.

There also are technical limitations of the study related to the use of extant data. The kindergarten screening measures and achievement measures were limited to those instruments previously selected by the

district for administration. Academic outcomes were limited to the results of the standardized measures. It was not possible to examine policies and procedures which may have been influential in the outcomes of the study.

Implications of the Study

The findings of this study have implications for kindergarten screening programs, for instructional practices, for provision of special services and for predictive validity studies.

1. Norm-referenced standardized tests are appropriate for inclusion in a kindergarten screening program.

The positive relationships between kindergarten screening measures and achievement measures demonstrates that the use of standardized test measures can be valuable in a screening program. Many teachers do not like to use standardized test measures because the results may be misinterpreted or misused. Following a review of the literature on teacher judgement, Hoge and Coladarci (1989) reported that teacher judgement about student achievement is generally accurate but levels of accuracy vary across teachers. A combination of standardized test measures with reports of teacher judgment would likely improve the assessment process.

The results of the analysis suggest that the administration of a single measure or a one time assessment is inadequate for making decisions about an individual pupil. Students with comparable screening scores may have different needs not identified by the screening measures. Placement and programming should be determined only after comprehensive data is collected. A more appropriate program would result in a comprehensive profile of pupil abilities and skills would include:

- a) a battery of screening measures which have high predictive validity

for particular outcomes;

b) knowledge of the SES area or mean ability of the school attended;

c) consideration for differential effects on performance of age, gender and physical problems;

d) teacher observation and checklists;

e) on-going assessment: kindergarten measures administered across the school year and possibly readministered to pupils of particular chronological age; outcome measures administered on several occasions such as, the middle of grade one, and the end of grade one, which would allow for interventions at the time the need presents.

2. Educators should monitor the effects of remedial programs to determine what is appropriate, for whom, when and for how long.

The large negative effects of the interventions found in this study have implications for the provision of special services. If receiving remedial help or progressing more slowly through the curriculum have the result of lowering achievement below that of pupils who remain in the classroom without assistance, the continued provision of special services would be questionable. The financial expenditures required to provide services are defensible only if the outcomes are positive.

Alternatively, if the provision of special services had a positive effect which was not measured by standardized test measures, it would be desirable to have documentation that the benefits to the pupils are worth the costs.

3. The examination of policies and practices in schools which are most effective should guide the development of such practices in schools which are less effective.

The significant differences in mean adjusted achievement between schools for pupils who obtained scores at the cut-off point for risk status indicates some schools are more effective for pupils at risk than other schools. The determination of policies and practices which make some schools more effective could guide the development of similar practices in the less effective schools.

4. Prediction-performance research should incorporate analysis of student- and school-level variables to accurately reflect the hierarchical nature of education.

The findings of this study have implications for future predictive validity studies. Examination of the models illustrates that strong positive relationships were consistently found for achievement with kindergarten screening measures. However, controlling for important variables changed the significance of the effects of other variables. For example, age-at-entry and gender were found to be significant across several measures when only pupil characteristics were included in the analysis. When the effects of educational interventions were controlled, age and gender effects were significant in fewer models. Also, when including four kindergarten screening measures in the analysis, only the gender effect on language was significant. This illustrates how analyses which fail to consider these variables, or perhaps other variables, may report age or gender effects which are inaccurate.

The significant effects of school mean-ability illustrate the importance of considering both within-school relationships and between-school relationships. Analyses which use small subject samples or which fail to consider contextual effects of schools may fail to identify

measures which are good predictors or alternatively, measures which appear to be predictive but are weak predictors when the effects of other variables are considered.

Recommendations for Future Research

Suggestions for research derived from the findings of this study include:

1. Examination of the relationship between specific language skills and achievement is appropriate to identify specific language skills which may be predictive of achievement.

The relationships between the kindergarten screening measure of expressive language with the outcome measures was relatively strong in the present study. Researchers have reported contradictory findings. Research studies could be directed at identifying specific receptive and expressive language skills which have significant effects on different areas of achievement. For example, tests of receptive and expressive language skills for pupils of different chronological ages could be compared with growth rates in language-related areas of reading, vocabulary and written language skills. Hierarchical linear modelling would allow for the analysis of growth rates and the analysis of between pupil differences in the rates. Multivariate analysis of individual and combinations of skills with achievement measures could identify the best language predictors. Greater understanding of these relationships could have strong implications for trends such as teaching reading through the "whole language" approach which is premised on a relationship between language and reading which has yet to be unequivocally proven.

2. The qualitative effects of particular interventions should be examined

to ensure pupils identified as "at risk" are provided optimal opportunities to progress.

The research designs for examining educational interventions are numerous; the access to data to complete such research is more difficult. Experimental studies in which treatment and control subjects are matched for age, gender, handicaps and school enrollment would be valuable to gain greater understanding of the effects of interventions. Caution would be required as denial of a particular intervention for an identified "at risk" pupil would be unethical. However, autonomy granted teachers, schools or districts in the selection of particular approaches to intervention should allow for investigation of various interventions and their effects.

Study of matched groups of subjects for whom the school recommends intervention and the parent refuses to accept the intervention would also be of interest. However, it would be important to investigate variables reflecting family processes, ethnic background, SES, which might have an effect on the decision and which might vary between parents who accept the intervention and those who refuse.

3. An examination of child variables which result in retention and the decision making processes of teachers and parents which lead to retention could be conducted through interviews and questionnaires.

Ten percent of the pupils in the achieved sample attended extended primary, although half or more of those pupils who remained in the primary grades an extra year scored above the cut-off score for "at risk" status. It is clear that the decision to extend the pupil's primary schooling was based on information other than screening information in many cases. Remaining an extra year in school is costly both to the pupil in earnings

at the completion of school and to districts which fund the additional year of school. Research on retention consistently reports negative effects on the child's social-emotional status and minimal benefits to achievement status.

Examination of social-emotional or behavioral factors of the pupils which contribute to the decision to provide or withhold interventions could be investigated through observational study or pupil-teacher-parent questionnaires.

4. The examination of home and school processes which lead to pupils performing better in some schools than pupils of similar ability in other schools could be undertaken.

The large adjusted average achievement differences between schools for children who obtain a score "at risk" illustrate the importance of examining performance at both the within-pupil and the between-school levels. While it is not possible to model all possible variables, theory driven investigation into these processes or experimental investigation might identify the important manipulable variables which improve performance. The availability of statistical programs which analyze the data at two or more levels simultaneously allows for structuring research which includes particular school-level variables or variables which represent district practices and policies. Alternatively, the achievement of pupils can be investigated over time by repeated measurement leading to a "growth rate" rather than a single point in time measure. Differences in growth rates between pupils can be investigated, and differences between schools in the average growth rates of pupils or for pupils of different ability or SES levels could be examined.

5. As school mean-ability may have acted as proxy for other important variables, it would be desirable to conduct research to identify factors highly associated with school mean-ability which add predictive power to achievement performance.

The contextual effects of school mean-ability identified in this study could be investigated in a variety of ways. One approach would be to identify if procedures or practices which lead to improved academic performance are different in schools with different levels of school mean-ability. Another approach would be to determine if the improved performance requires generally higher ability for the majority of class members or if some critical number of pupils of high ability could have the same effect (Willms, 1986). The question could be reversed regarding pupils with low ability. Hierarchical linear modelling is an appropriate method for this type of investigation. This investigation could have implications for grouping pupils in classrooms, for maintaining neighborhood boundaries for enrollment and for mainstreaming pupils with special needs.

REFERENCES

- Affleck, J.O., Madge, S. Adams, A. & Lowenbraun, S. (1988). Integrated classrooms vs resource model: Academic viability and effectiveness. *Exceptional Children*, 54, 339-348.
- Adelman, H.S. & Feshbach, S. (1971). Predicting reading failure: Beyond the readiness model. *Exceptional Children*, 37, 349-354.
- Aiken, L.R. (1972). Language factors in learning mathematics. *Review of Educational Research*, 42, 359-385.
- Aiken, M.A. & Longford, N.T. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society, A*, 149, 1-26.
- Alexander, K.L., Fennessey, J. McDill, E.L. & D'Amico, R.J. (1979). School SES influences -- composition or context? *Sociology of Education*, 52, 222-237.
- Algozzine, B., Mercer, C.D. & Countermine, T. (1975). The effects of labels and behavior on teacher expectations. *Exceptional Children*, 44, 131-132.
- Algozzine, B., Mercer, C.D. & Countermine, T. (1977). Labeling exceptional children: An analyses of expectations. *Exceptional Children*, 44, 131-132.
- American Psychological Association (1974). *Standards for Educational and Psychological Tests*. Washington, DC: APA.
- Ames, L.B. (1963). *Is Your Child in the Wrong Grade?* New York: Harper & Row.
- Anastasi, A. (1976). *Psychological Testing* (4th Ed.). New York: MacMillan Publishing Co., Inc.
- Anderson, C.S. (1982). The search for school climate: A review of the research. *Review of Educational Research*, 52(3), 368-420.
- Anderson-Inman, L. (1986). Bridging the gap: Student-centered strategies for promoting transfer of learning. *Exceptional Children*, 52, 562-572.
- Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.), *Educational Measurement* (2nd ed.). Washington, DC: American Council on Education.

- Askov, W., Otto, W. & Smith, R. (1972). Assessment of the de Hirsch predictive index tests of reading failure. In R.C. Auckerman (Ed.), *Some Persistent Questions on Beginning Reading*. (pp. 33-427). Newark, DE: International Reading Association.
- Badian, N. (1986). Improving the prediction of reading for the individual child: A four-year follow-up. *Journal of Learning Disabilities*, 19(5), 262-269.
- Badian, N. & Serwer, B. (1975). The identification of high risk children: A retrospective look at selection criteria. *Journal of Learning Disabilities*, 8(5), 283-287.
- Baenen, N.R. (1988). *Perspectives after five years -- has retention passed or failed?* Paper presented to the American Educational Research Association in New Orleans.
- Bak, J.J., Cooper, E.M., Dobroth, K.M. & Siperstein, G.N. (1987). Special class placements as labels: Effects on children's attitudes toward learning handicapped peers. *Exceptional Children*, 54, 151-155.
- Barnes, K.E. (1982). *Preschool Screening: The Measurement and Prediction of Children At-risk*. Springfield, IL: Charles C. Thomas Publishing.
- Barr, R. & Dreeben, R. (1977). Instruction in classrooms. In L.S. Shulman (Ed.), *Review of Research in Education*, Vol 5, Itasca, IL: Peacock.
- Barr, R. & Dreeben, R. (1983). *How Schools Work*. Chicago, IL: University of Chicago Press.
- Beattie, C. (1970). Entrance age to kindergarten and first grade: Its effect on cognitive and affective development of students. (ERIC NO ED 133 050)
- Becker, W.C. & Gersten, R. (1982). A follow-up of follow-through: The later effects of the direct instruction model on children in fifth and sixth grades. *American Educational Research Journal*, 19, 75-92.
- Beers, C.S. & Beers, J.W. (1980). Early identification of learning disabilities: Facts and fallacies. *The Elementary School Journal*, 81, 2.
- Beery, K. (1989). *The VMI Developmental Test of Visual Motor Integration*. Toronto: Modern Curriculum Press.

- Beery, K. & Buktenica, N. (1982). *The Developmental Test of Visual Motor Integration*. Cleveland, OH: Modern Curriculum Press.
- Bender, L. (1938). A Visual Motor Gestalt Test and Its Clinical Use. Research Monograph Number 3. New York, NY: American Orthopsychiatric Association.
- Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C.W. Harris (Ed.), *Problems in Measuring Change*. Madison, WI: University of Wisconsin Press.
- Beyer, F.S. & Smey-Richman, B. (1988). *Addressing the "at-risk" challenge in the nonurban setting*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Bidwell, C.E. & Kasarda, J.D. (1980). Conceptualizing and measuring the effects of school and schooling. *American Journal of Education*, 88, 401-430.
- Bilka, L. (1972). An evaluation of the predictive value of certain readiness measures. In R.C. Auckerman (Ed.), *Some Persistent Questions on Beginning Reading*. (pp. 43-49). Newark, DE: International Reading Association.
- Black, T. (1971). Investigation of intelligence as a causal factor in reading problems. *Journal of Learning Disabilities*, 4(3), 22-25.
- Boehnlein, M. (1987). Reading intervention for high-risk first-graders. *Educational Leadership*, 44(6), 32-37.
- Book, R.M. (1974). Predicting reading failure: A screening battery for kindergarten children. *Journal of Learning Disabilities*, 7(1), 43-47.
- Book, R.M. (1980). Identification of educationally at-risk children during the kindergarten year: A four-year follow-up study of groups test performance. *Psychology in the Schools*, 17, 153-158.
- Borg, W.R. & Gall, M.D. (1983). *Fourth Edition Educational Research: An Introduction*. New York: Longman.
- Bracken, B.A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment*, 4, 313-326.

- Brandon, P.R., Newton, B.J. & Hammond, O. (1987). Children's mathematics achievement in Hawaii: Sex differences favoring girls. *American Educational Research Journal*, 24(3), 437-461.
- Bremer, N. (1980). Do reading tests predict success in reading? *Elementary School Journal*, 59, 222-229.
- Bricker, D.D. (1986). *Early Education of At-risk and Handicapped Infants, Toddlers, and Preschool children*. Glenview, IL: Scott, Foresman, & Co.
- Brookover, W.B., Sweitzer, J.H., Schneider, J.M., Beady, C.H., Flood, P.K. & Wisenbaker, J.M. (1978). Elementary school social climate and school achievement. *American Educational Research Journal*, 15, 301-318.
- Bryk, A.S. & Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 101, 147-158.
- Burton, L., Drake, P., Ekins, J., Graham, L., Taplin, M. & Weiner, G. (1986). *Girls into mathematics*. Centre for mathematics education. The Open University in association with The London Education Authority. Cambridge, MA: Cambridge University Press.
- Busch, R. (1980). Predicting first grade reading achievement. *Learning Disability Quarterly*, 3, 38-47.
- Butler, S.R. (1979). Predictive antecedents of reading disability in the early years of schooling. *Journal of Special Education*, 3, 263-274.
- Buttram, J., Covert, R. & Hayes, M. (1976). Prediction of school readiness and early grade achievement by classroom teachers. *Educational and Psychological Measurement*, 36, 543-546.
- Casto, G. & Mastropieri, M.A. (1986). The efficacy of early intervention programs: A meta-analysis. *Exceptional Children*, 52, 417-424.
- Chall, J.S. (1967). *Learning to Read: The Great Debate*. Toronto: McGraw Hill Book Co.
- Clifford, P. & Heath, A. (1984). Selection does make a difference. *Oxford Review of Education*, 10(1) 85-97.
- Coleman, J.M. (1983). Handicapped labels and instructional segregation: Influences on children's self-concepts versus

the perceptions of others. *Learning Disabilities Quarterly*, 6, 3-11.

Colligan, R.C. & O'Connell, E.J. (1974). Should psychometric screening be made an adjunct to the pediatric preschool examination? *Clinical Pediatrics*, 13, 29-34.

Cowen, E.L., Weissberg, R.P. & Gisare, J. (1984). Differentiating attributes of children referred to a school mental health program. *Journal of Abnormal Child Psychology*, 12, 397-409.

Crary, M.A. (1984). Phonological characteristics of developmental verbal apraxia. *Seminars in Speech and Language*, 5(2), 71-82.

Davis, B.G., Trimble, C.S. & Vincent, D.R. (1980). Does age of entrance affect school achievement? *The Elementary School Journal*, 80, 134-143.

de Hirsch, K., Jansky, J. & Langford, W. (1966). *Predicting Reading Failure*. New York, NY: Harper & Row.

Denham, C. & Lieberman, A. (Eds.) (1980). *Time to Learn*. Washington, DC: National Institute of Education.

Deno, S.L. (1986). Formative evaluation of individual student programs: A new role for school psychologists. *School Psychology Review*, 15(3), 358-374.

Deverell, A. (1974). The Deverell classification test for use with school beginners (letters and numbers). In *Teaching Children to Read and Write*. Toronto Ont.: Holt, Rinehart & Winston.

Diamond, G.H. (1983). The birthdate effect -- A maturational effect? *Journal of Learning Disabilities*, 16(3), 161-164.

Di Pasquale, G.W., Moule, A.D. & Flewelling, R.W. (1980). The birthdate effect. *Journal of Learning Disabilities*, 13(5), 4-8.

Donofrio, A.F. (1977). Grade repetition: Therapy of choice. *Journal of Learning Disabilities*, 10(6), 349-351.

Dreeben, R. & Gamoran, A. (1986). Race, instruction, and learning. *American Sociological Review*, 51, 660-669.

Duffy, J., Ritter, D. & Fedner, M. (1976). Developmental test of visual motor integration and the Goodenough draw-a-man

test as predictors of academic success. *Perceptual and Motor Skills*, 43, 543-546.

Dunleavy, R.A., Hansen, J.L., Szasz, C.W. & Baade, L.E. (1981). Early kindergarten identification of academically not-ready children by use of human figure drawing developmental score. *Psychology in the Schools*, 18(1), 35-38.

Dunn, J.A. (1967). Note on the relation of Harris' draw a woman to WISC I.Q.'S. *Perceptual and Motor Skills*, 24, 316.

Durkin, D. (1974). *Teaching Young Children To Read*. Boston, MA: Allyn and Bacon, Inc.

Durkin, D. (1987). Testing in the kindergarten. *The Reading Teacher*, 40, 766-770.

Durrell, D.D. (1958). First-grade reading success study: A summary. *Journal of Education*, 140, 2-6.

Dykstra, R. (1967). The use of reading readiness tests for diagnosis and prediction: A critique. In T.C. Barrett (Ed.), *The Evaluation of Children's Reading Achievement*. Newark DE: International Reading Association.

Eaves, L.C., Kendall, D.C. & Crichton, J.A. (1974). The early identification of learning disabilities: A follow-up study. *Journal of Learning Disabilities*, 7(10), 42-48.

Eccles, J. & Jacobs, J.E. (1986). Social forces shape math attitudes and performance. Signs: *Journal of Women in Culture and Society*, 11(2), 367-380.

Evans, R. (1976). The prediction of educational handicap - a longitudinal study. *Educational Research*, 19, 57-68.

Fennema, E. & Peterson, P. (1985). Autonomous learning behavior: explanation of gender-related differences in mathematics. In L.C. Wilkinson & C.B. Marett (Ed.), *Gender Influences In Classroom Interaction*. (pp. 17-35). Orlando, FL: Academic Press.

Ferinden, W., Jacobson, S. & Linden, N. (1970). Early identification of learning disabilities. *Journal of Learning Disabilities*, 3(11), 589-593.

Feshbach, S., Adelman, H. & Fuller, W. (1974). Early identification of children with high risk of reading failure. *Journal of Learning Disabilities*, 7(10), 49-54.

- Fletcher, J.M. & Satz, P. (1982). Kindergarten prediction of reading achievement. A seven-year longitudinal follow-up. *Educational and Psychological Measurement*, 42, 681-685.
- Foster, G., Schmidt, C. & Sabatino D. (1976). Teacher expectancies and the label "learning disabilities". *Journal of Learning Disabilities*, 9(2), 58-61.
- Foster, G., Ysseldyke, J.E. & Reese, J.H. (1975). I wouldn't have seen it if I hadn't believed it. *Exceptional Children*, 41, 469-473.
- Fox, L.H. & Cohn, S.J. (1980). Sex differences in the development of precocious mathematical talent. In L.H.Fox, L.Brody, & D.Tobin, (Eds.), *Women and the Mathematical Mystique*. Baltimore, MD: The Johns Hopkins University Press. 94-112.
- Frisch, G.R. & Handler, L. (1974). A Neuropsychological investigation of functional disorders of speech articulation. *Journal of Speech and Hearing Research*, 17, 432-445.
- Galante, M.B., Flye, M.E., Stephens, R.N. & Stephens, L.S. (1972). Cumulative minor deficits: A longitudinal study of the relation of physical factors to school achievement. *Journal of Learning Disabilities*, 5, 75-80.
- Gallagher, J.J. (1984). Learning disabilities and the near future. *Journal of Learning Disabilities*, 54(4), 339-348.
- Gallagher, J.J. & Bradley, R.H. (1972). Early identification of developmental difficulties. In I. Gordon (Ed.), *Early Childhood Education*. Chicago, IL: University of Chicago Press.
- Gamoran, A. (in press). Schooling and achievement: Additive versus interactive multilevel models.
- Garmoran, A. (1984). *Teaching, grouping and learning: A study of the consequences of educational stratification*. Unpublished Ph.D. dissertation, Dept. of Education. University of Chicago.
- Gamoran, A. (1988). Resource allocation and the effects of schooling: A sociological perspective. (pp.207-232). In D.H. Monk & J. Underwood (Eds.). *Microlevel School Finance: Issues and Implications for Policy*. Ninth Annual Yearbook of the American Educational Finance Association, Cambridge, MA: Ballinger.

- Gamoran, A. & Dreeben, R. (1986). Coupling and control in educational organizations. *Administrative Science Quarterly*, 31, 612-632.
- Gauthier, S. & Madison, C. (1973). *Kindergarten Language Screening Test*. Tigard, OR: CC Publications.
- Germain, R.b. & Merlo, M. (1985). Best practices in assisting in promotion and retention decisions. In A. Thomas & J. Grimes (Eds.). *Best Practices In School Psychology*. Kent, OH: National Association of School Psychologists.
- Gettinger, M. (1984). Achievement as a function of time spent in learning and time needed for learning. *American Educational Research Journal*, 21, 617-628.
- Ghiselli, E., Campbell, J. & Zedeck, S. (1981). *Measurement Theory for the Behavioural Sciences*. San Francisco, CA: W.H. Freeman & Co.
- Glass, G. & Hopkins, K. (1984). *Statistical Methods in Education and Psychology*. Englewood Cliffs, NJ: Prentice Hall, Inc.
- Glavin, J., Quay, H., Annesley, F. & Werry, J. (1971). An experimental resource room for behavior problem children. *Exceptional Children*, 38, 131-138.
- Glazzard, M. (1979). The long-range effectiveness of three kindergarten predictors for first-, second-, third- and fourth-grade achievement. *Journal of Learning Disabilities*, 12(10), 55-60.
- Glazzard, P.H. (1982). Long-range kindergarten prediction of reading achievement in first through sixth grade. *Learning Disability Quarterly*, 5, 85-88.
- Goetz, E. (1971). Hearing growth and reading. In M. Douglass (Ed.), *Thirty-fifth Yearbook of the Claremont Reading Conference*, 120-126.
- Goldman, R., Fristoe, M. & Woodcock, R. (1970). *Test of auditory discrimination*. Circle Pines, MN: American Guidance Service, Inc.
- Goldman, R.K. & Velasco, Jr., M. (1980). Toward the development of a rational scale in the use of human-figure drawings as a kindergarten screening measure. *Perceptual and Motor Skills*, 50, 571-577.

- Goldstein, H. (1986). Multilevel mixed linear model analyses using iterative generalized least squares. *Biometrika*, 73(1), 43-56.
- Good, T.L. & Slavings, R.L. (1988). Male and female question-asking behavior in elementary and secondary mathematics and language arts classes. *Journal of Research in Childhood Education*, 3(1), 5-23.
- Goodenough, T. (1926). *Measurement of Intelligence by Drawings*. Yonkers-on-Hudson, NY: World Book.
- Goodwin, W.L. & Driscoll, L.A. (1980). *Handbook for Measurement and Evaluation in Early Childhood Education*. San Francisco, CA: Josey-Bass Publishers.
- Gordon, N. & McKinlay, I. (Eds.). (1980). *Helping Clumsy Children*. Churchill Livingstone, Edinburgh London and New York.
- Gortmaker, S.L. & Sappenfield, W. (1984). Chronic childhood disorders: Prevalence and impact. *Pediatric Clinics of North America*, 31, 3-18.
- Gottfredson, G.D. (1988). *You get what you measure - you get what you don't: Higher standards, higher test scores, more retention in grade*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.
- Gray, R.A., Saski, J., McEntire, M.E. & Larsen, S.C. (1980). Is proficiency in oral language a predictor of academic success? *The Elementary School Journal*, 80, 261-268.
- Gredler, G.R. (1978). A look at some important factors in assessing readiness for school. *Journal of Learning Disabilities*, 11(5), 25-31.
- Gredler, G.R. (1980). Cumulative retention rate as an index of academic progress: A third look. *Journal of Learning Disabilities*, 13, 15-18.
- Gredler, G.R. (1980). The birthdate effect: Fact or artifact? *Journal of Learning Disabilities*, 13, 9-12.
- Gredler, G.R. (1984). Transition classes: A viable alternative for the at-risk child? *Psychology in the Schools*, 21, 463-470.
- Greenwood, C.R., Dinwiddie, G., Terry, B., Wade, L., Stanley, S., Thibadeau, S. & Delquadri, J. (1984). Teacher-versus

peer-mediated instruction: An ecobehavioral analysis of achievement outcomes. *Journal of Applied Behavior Analysis*, 17, 521-538.

- Griffin, M. & Eberly, D.W. (1971). *Visual problems and reading*. Washington, DC: National Reading Center Foundation.
- Grimley, A.M. & McKinlay, I.A. (1977). *The Clumsy Child*. The Association of Pediatric Chartered Psychotherapists, Sussex.
- Groff, P. (1977). Oral Language and reading. *Reading World*, 17, 71-78.
- Groff, P. (1978). Children's oral language and their written composition. *The Elementary School Journal*, 78, 181-191.
- Gronlund, N.E. (1975). *Measurement and Evaluation in Teaching*. New York: MacMillan Publishing Co.
- Gubbay, S.S. (1975). *The Clumsy Child*. Philadelphia, PA: Saunders.
- Gulliford, R. (1976). The early identification of educationally at risk children. In K. Wedell & B.C. Raybould (Eds.), *The early identification of 'at risk' children*. Educational Review, Occasional Publication #6, University of Birmingham.
- Guralnick, M.J. & Brickey D.D. (1987). The effectiveness of early intervention for children with cognitive and general developmental delays. In M.J. Guralnick & F.C. Bennett (Eds.), *The Effectivness of Early Intervention for At-Risk and Handicapped Children*, (pp. 115-173). Orlando, FL: Academic Press.
- Guskin, S.L., Barte, N.R., & Macmillan, D.L. (1975). Perspective of the labeled child. In N. Hobbs (Ed.), *Issues in the Classification of Children*, 2. San Francisco, CA: Jossey-Bass.
- Hall, R.V. (1963). Does entrance age affect achievement: *Elementary School Journal*, 64, 391-396.
- Hallinan, M.T., & Sorenson, A.B. (1987). Ability grouping and sex differences in mathematics achievement. *Sociology of Education*, 60, 63-72.
- Hammill, D.D. & McNutt, G. (1980). Language abilities and

- reading: A review of the literature on their relationship. *The Elementary School Journal*, 80, 269-277.
- Hammond, J.M. (1980). Nutrition and hyperactivity. *American Education*, 4, 34.
- Hanna, G. & Kuendiger, E. (1986). *Differences in mathematical achievement levels in attitude for girls and boys in twenty countries*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Harber, J.R. (1981). Assessing the quality of decision making in special education. *Journal of Special Education*, 15, 77-90.
- Haring, N.G. & Ridgway, R.W. (1967). Early identification of children with learning disabilities. *Exceptional Children*, 33, 387-395.
- Harris, D. (1963). *Children's Drawings as Measures of Intellectual Maturity*. New York, NY: Harcourt, Brace & World, Inc.
- Hartlage, L.C. & Lucas, D.G. (1973). Group screening for reading disability in first grade children. *Journal of Learning Disabilities*, 6(5), 48-52.
- Hauser, R. (1970). Context and Consequence: A Cautionary Tale. *American Journal of Sociology*, 75, 645-663.
- Hayden, A.H. (1974). Perspectives of early childhood education in special education. In N.G. Haring (Ed.), *Behavior of Exceptional Children: An Introduction To Special Education*. Columbus, OH: Charles E. Merrill.
- Haynes, M.C. & Jenkins, J.R. (1986). Reading instruction in special education resource rooms. *American Educational Research Journal*, 23, 161-190.
- Hedges, W.D. (1977). *At What Age Should Children Enter First Grade. A Comprehensive Review of the Research*. Published for Florida Educational Research & Development Council. Ann Arbor, MI: University Microfilms International.
- Henderson, V., Mieszkowski, P. & Sauvageau, Y. (1978). Peer group effects and educational production functions. *Journal of Public Economics*, 10, 97-106.
- Hieronymus, A.N., King, E.M., Bourdon, J.W., Gossling, D., Grywinski, N.J., & Moss, G.L. (1976). *Canadian Tests of*

Basic Skills: Manuals for Administrators, Supervisors, and Counsellors. Canada: Thomas Nelson & Sons.

- Hinton, G.G. & Knights, R.M. (1971). Children with learning problems: Academic history, academic prediction and adjustment three years after assessment. *Exceptional Children*, 37, 513-519.
- Hoge, R. & Coladarci, J. (1989). Teacher-based judgements of academic achievement: A research of literature. *Review of Educational Research*, 59(3), 297-313.
- Holmes, C.T. & Matthews, K.M. (1984). The effects of nonpromotion on elementary and junior high pupils: A meta-analysis. *Review of Educational Research*, 54, 225-236.
- Hoover, H.D. (1984). The most appropriate scores for measuring educational development in the elementary schools: GE's. *Educational Measurement: Issues and Practice*, 3(4), 8-14.
- Hoover, H.D. (1984). Rejoinder to Burket. *Educational Measurement: Issues and Practices*, 3(4), 16-18.
- Hopkins, K.D. & Glass, G.V. (1978). *Basic Statistics for the Behavioral Sciences*. Englewood Cliffs, NJ: Prentice-Hall.
- Horst, D.P. (1975). *A practical guide to measuring project impact on student achievement*. Number 1 in a Series of Monographs on Evaluation in Education. U.S. Dept. of Health, Education and Welfare. Washington, DC: U.S. Gov't Printing Office.
- Horst, D.P. (1986). *What's bad about grade-equivalent scores*. ESEA Title 1 Evaluation and Reporting System, Technical Paper No. 1. Mountain View, CA: RMC Research Corporation.
- Hunter, E.J. & Johnson, L.C. (1971). Developmental and psychological differences between readers and nonreaders. *Journal of Learning Disabilities*, 4(10), 572-577.
- Hyde, J. (1981). How large are cognitive gender differences? *American Psychologist*, 36, 292-301.
- Ilg, F.L. & Ames, T.B. (1964). *School Readiness. Behavior Tests used at the Gessell Institute*. New York, NY: Harper & Row.

- Jackson, G.B. (1975). The research evidence on the effects of grade retention. *Review of Educational Research*, 45, 613-635.
- Jacob, S., Snider, K.P. & Wilson, J.F. (1988). Validity of the DIAL-R for identifying children with special education needs and predicting early reading achievement. *Journal of Psychoeducational Assessment*, 6, 289-297.
- Jansky, J. & de Hirsch, K. (1972). *Preventing Reading Failure*. New York, NY: Harper & Row.
- Jansky, J.J. (1973). *Early prediction of reading problems*. The Orton Society Reprint Series, XXIII, Reprint No. 55, Towson, MD: The Orton Society, Inc.
- Jansky, J.J. (1978). A critical review of some developmental and predictive precursors of reading disabilities. In A.L. Benton & D. Pearl (Eds.), *Dyslexia an Appraisal of Current Knowledge*. New York, NY: Oxford University Press.
- Jarman, R.R. (1980). Theoretical and methodological issues in the early identification of cognitive-developmental disabilities: Descriptions of a longitudinal study. In R.F. Jarman & J.P. Das (Eds.), *Issues in Developmental Disabilities*. Ann Arbor, MI: University Microfilms International.
- Jenkins, E. & Lohr, F.E. (1964). Severe articulation disorders and motor ability. *Journal of Speech and Hearing Disorders*, 29(3), 286-292.
- Johnson, S. (1987). Early-developed sex differences in science and mathematics in the United Kingdom. *Journal of Early Adolescence*, 7(1), 1-3.
- Johnstone, P., Allington, R. & Afflerbach, P. (1985). The confluence of classroom and remedial instruction. *Elementary School Journal*, 85, 465-477.
- Judy, J. (1986). Early screening is essential for educational accountability: Response to Salzer and to Shepard and Smith. *Educational Leadership*, 11, 87-88.
- Karweit, N. (1988). *Effective elementary programs and practices for at risk students*. Paper prepared for annual meeting of American Educational Research Association, New Orleans.

- Kaufman, A.S. & Kaufman, N.L. (1972). Tests built from Piaget's and Gesell's tasks as predictors of first-grade achievement. *Child Development*, 43, 521-535.
- Keogh, B. (1977). Early I.D.: selective perception or perceptive selection? *Academic Therapy*, 12, 267-274.
- Keogh, B. & Becker, L.D. (1973). Early detection of learning problems: Questions, cautions and guidelines. *Exceptional Children*, 40, 5-11.
- Keogh, B.K. & Daley, J.E. (1983). Early identification: One component of comprehensive services for at-risk children. *Topics in Early Childhood Special Education*, 3(3), 7-16.
- Keogh, B.K., Tchir, M.A. & Wendeguth-Behn, A. (1974). Teachers' perceptions of educationally high risk children. *Journal of Learning Disabilities*, 7(6), 43-50.
- King, E.M. (1982). *Canadian Test of Basic Skills (Teacher's Guide)*. Canada: Nelson Canada Limited.
- Kirk, W.D. (1966). A tentative screening procedure for selecting bright and slow children in kindergarten. *Exceptional Children*, 33, 235-241.
- Klein, A.E. (1977). The validity of the Screening Test of Academic Readiness in predicting achievement in first and second grades. *Educational and Psychological Measurement*, 37, 493-499.
- Koppitz, E.M. (1975). *The Bender Gestalt Test for Young Children*. New York, NY: Greene & Stratton.
- Kornberg, M.S. & Kaplan, G. (1980). Risk factors and preventive intervention in child therapy, A review. *Journal of Prevention*, 1, 71-133.
- Larter, S. (1982). *The physically handicapped and health impaired children: Do they prosper in regular Toronto elementary school?* Toronto, Ontario: Toronto Board of Education Research Department.
- LaTorre, R.A., Hawkhead, F., Kawahira, R. & Bilow, L. (1982). Kindergarten screening pilot project in Vancouver schools 1979-80: A two-year follow-up of the McCarthy Screening Test. *B.C. Journal of Special Education*, 6, 23-41.
- Lazar, I. & Darlington, R. (1978). *Lasting Effects After Preschool*, Washington, DC: Office of Human Development Services, U.S. Department of Health, Education and Welfare.

- Lazar, I., Hubbell, R., Murray, H., Rosche, M. & Royce, J. (1978). Preliminary findings of the developmental continuity longitudinal study. In K. Allen, V. Holm & R. Schiefelbusch (Eds.), *Early Intervention -- A Team Approach*. Baltimore, MD: University Park Press.
- Leigh, J.E. (1983). Early labeling of children: Concerns and alternatives. *Topics In Early Childhood Special Education*, 3(3) 1-6.
- Leinhardt, G. (1980). Transition rooms: Promoting maturation or reducing education? *Journal of Educational Psychology*, 72, 55-61.
- Lerner, J., Mardell-Czudnowski, C. & Goldenberg, D. (1981). *Special Education for the Early Childhood Years*. Englewood Cliffs, NJ: Prentice-Hall.
- Lesiak, W.J. (1978). Developmental tasks for kindergarten readiness. *Archives of Behavioral Science Monograph* (No.52).
- Levi, G., Musatti, L., Letizia Priedda, M. Sechi, E. (1984). Cognitive and linguistic strategies in children with reading disabilities in an oral storytelling test. *Journal of Learning Disabilities*, 17, 406-410.
- Lewis, A. (1980). The early identification of children with learning difficulties. *Journal of Learning Disabilities*, 13(2), 51-57.
- Lichtenstein, R. (1981). Comparative validity of two preschool screening tests: Correlational and classificational approaches. *Journal of Learning Disabilities*, 14, 68-72.
- Lichtenstein, R. & Ireton, N. (1984). *Preschool Screening: Identifying Young Children with Developmental and Educational Problems*. Orlando, Fl: Grune & Stratton.
- Lidz, C.S. (1983). Issues in Assessing Preschool Children. In K.D. Paget & B.A. Bracken (Eds.). *The Psychoeducational Assessment of Preschool Children*, (pp. 17-27). New York, NY: Grune & Stratton.
- Lindquist, G.T. (1982). Preschool screening as a means of predicting later reading achievement. *Journal of Learning Disabilities*, 15(6), 331-332.
- Lindsay, G.A. & Wedell, K. (1982). The early identification of

- educationally at risk children revisited. *Journal of Learning Disabilities*, 15, 212-216.
- Linn, R.J. & Slinde, J.A. (1977). The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research*, 47, 121-50.
- Lloyd, D.N. (1978). Prediction of school failure from third-grade data. *Educational and Psychological Measurement*, 38, 1193-1200.
- Maccoby, E.E. & Jacklin, C.N. (1974). *The Psychology of Sex Differences*. Stanford, CA: The Stanford University Press.
- MacMillan, L., Jones, R. & Aloia, G. (1974). The mentally retarded label: a theoretical analysis and review of research. *American Journal of Mental Deficiency*, 79, 241-261.
- Madden, N. & Slavin, R. (1983). Mainstreaming students with mild academic handicaps: Academic and social outcomes. *Review of Educational Research*, 43, 519-569.
- Madden, N. & Slavin, R. (1987). *Effective Pull-out Programs for Students At Risk*. Baltimore, MD: Centre for Research on Elementary and Middle Schools, The Johns Hopkins University.
- Maitland, S., Nadeau, J.B.E. & Nadeau, G. (1974). Early school screening practices. *Journal of Learning Disabilities*, 7(10), 55-59.
- Mann, P., Suiter, P. & McClung, R. (1987). *Teacher's Handbook of Diagnostic Screening*. Boston, MA: Allyn and Bacon, Inc.
- Mardell, C. & Goldenberg, D. (1975a). For prekindergarten screening disabilities: DIAL. *Journal of Learning Disabilities*, 8(3), 140-47.
- Mardell, C. & Goldenberg, D. (1975b). *Developmental Indicators for the Assessment of Learning*. Highland Park, IL: DIAL, Inc.
- Martin, D.J. & Hoover, H.D. (1987). Sex differences in educational achievement: A longitudinal study. *Journal of Early Adolescence*, 7(1), 65-83.
- Mayron, L.W. (1978). Ecological factors in learning disabilities. *Journal of Learning Disabilities*, 11(8), 40.

- McAfee, J.K. (1981). *Toward a theory of promotion: Does retaining students really work?* Paper presented at the meeting of the American Educational Research Association, Los Angeles, CA.
- McCann, R. & Austin, S. (1988). *At-risk youth: Definitions, dimensions, and relationships.* Paper presented for the Annual Meeting of the American Educational Research Association.
- McLoughlin, J.A., Hall, M., Isaacs, B., Petroski, J., Karibo, J. & Lindsey, B. (1983). The relationship of allergies and allergy treatment to school performance. *Annals of Allergy*, 51, 506.
- McNutt, G. & Friend, M. (1985). Status of the resource room model in local education agencies: A descriptive study. *Learning Disability Quarterly*, 8, 101-108.
- Medway, F.J. & Rose, J.S. (1986). Grade retention. In T. Kratochwill (Ed.), *Advances in School Psychology*, 1, 141-175.
- Meehl, P.E. & Rosen, A. (1955). Antecedent probability and the efficacy of psychometric signs, patterns or cutting scores. *Psychological Bulletin*, 52, 194-216.
- Meisels, S.J. (1984). *Prediction, prevention and developmental screening in the EPSDT program.* In H.W. Stevenson & A.E. Siegel (Eds.). *Child Development Research and Social Policy*. Chicago, IL: University of Chicago Press.
- Meisels, S.J. (1985). *Developmental Screening in Early Childhood: A Guide (Rev.Ed.)*. Washington DC: National Association for the Education of Young Children.
- Meisels, S.J. (1986). Testing four- and five-year-olds. Response to Salzer and to Shepard and Smith. *Educational Leadership*, 44(3) 90-92.
- Meisels, S.J. (1987). Uses and abuses of developmental screening and school readiness testing. *Young Children*, 42(2), 68-73.
- Meisels, S.J. (1989). High-stakes testing. *Educational Leadership*, 46(7), 16-22.
- Mercer, C.D. (1979). *Children and Adolescents with Learning Disabilities*. Columbus, OH: Charles E. Merrill.

- Mercer, C.D., Algozzine, B. & Trifiletti, J. (1988). Early identification -- an analysis of the research. *Learning Disability Quarterly*, 11, 176-188.
- Meyer, J.W. (1977). The effects of education as an institution. *American Journal of Sociology*, 83(1), 55-77.
- Meyer, J.W. (1980). Levels of the educational system and schooling effects. In C. Bidwell & D. Windham, (Eds.). *The Analyses of Educational Productivity*, Vol II. Cambridge, MA: Ballinger.
- Meyers, C.E., Attwell, A.A. & Orpet, R.E. (1968). Prediction of fifth grade achievement from kindergarten test and rating data. *Educational and Psychological Measurement*, 28, 457-463.
- Miller, L.J. (1988). Differentiating children with school-related problems after four years using the Miller assessment for preschoolers. *Psychology in the Schools*, 25, 10-15.
- Miller, L. & Sprong, T. (1986). Psychometric and qualitative comparison of four preschool screening instruments. *Journal of Learning Disabilities*, 19, 480-484.
- Miller, L.J. & Schouten, P.G. (1988). Age-related effects on the predictive validity of the Miller assessment for preschoolers. *Journal of Psychoeducational Assessment*, 6, 99-106.
- Miller, W.D. & Norris, R.C. (1967). Entrance age and school success. *Journal of School Psychology*, 6, 47-59.
- Morency, A. & Wepman, J. (1973). Early perceptual ability and later school achievement. *Elementary School Journal*, 73, 323-327.
- Murnane, R.J. (1981). Interpreting the evidence on school effectiveness. *Teachers College Record*, 83, 19-35.
- Naglieri, J.A. (1988). *Draw-a-Person: A Quantitative Scoring System*. Toronto: Harcourt-Brace, Jovanovich, Inc.
- Naglieri, J.A. & Maxwell, S. (1981). Inter-rater reliability and concurrent validity of the Goodenough-Harris draw-a-person and McCarthy. *Perceptual and Motor Skills*, 53, 343-348.
- National Association for the Education of Young Children (NAEYC). Position Statement on Standardized Testing of

- Young Children 3 Through 8 years of Age. (1988). *Young Children*, 88, 42-47.
- Norton, L. (1979). The identification of "at risk" kindergarten children. *Special Education in Canada*, 53(4), 20-22.
- O'Connor, A. (1980). Early screening - some problems and practices. *British Columbia Journal of Special Education*, 4, 271-283.
- Paget, K.D. & Bracken, B.A. (Eds.). (1983). *The Psycho-educational Assessment of Preschool Children*. New York: Grune & Stratton.
- Paget, K.D. & Nagle, R.J. (1986). A conceptual model of preschool assessment. *School Psychology Review*, 15, 154-165.
- Palardy, J. (1969). What teachers believe, what children achieve. *Elementary School Journal*, 69, 370-374.
- Papalia, D.E. & Olds, S.W. (1975). *A Child's World: Infancy Through Adolescence*. New York, NY: McGraw-Hill Book Co.
- Park R. (1978). Performance on geometric figure copying test as predictors of types of errors in decoding. *Reading Research Quarterly*, 14, 100-118.
- Pattison, P. & Grieve, N. (1984). Do spatial skills contribute to sex differences in different types of mathematical problems? *Journal of Educational Psychology*, 76, 678-689.
- Pearson, L. & Lindsay, G. (1986). *Special Needs in the Primary School: Identification and Intervention*. Windsor, Berkshire, England: NFER-Nelson.
- Pedhazur, E. (1982). *Multiple Regression in Behavioral Research*. New York, NY: Holt, Rinehart & Winston.
- Perrin, J.M. (1986). Chronically ill children: An overview. *Topics in Early Childhood Special Education*, 5, 1-11.
- Peterson, J.M. (1989). Remediation is no remedy. *Educational Leadership*, 46(6), 24-25.
- Peterson, S.E., De Gracie, J.S. & Ayabe, C.R. (1987). A longitudinal study of the effects of retention/promotion on academic achievement. *American Educational Research Journal*, 24, 107-118.

- Phye, G.D. & Halderman, M.A. (1980). Cumulative retention rate as an index of academic progress: A second look. *Journal of Learning Disabilities*, 13(5) 13-14.
- Pikulski, J.J. (1972). A comparison of figure drawings and WISC I.Q's among disabled readers. *Journal of Learning Disabilities*, 5(3), 41-44.
- Pomplun, M. (1988). Retention: The earlier the better? *Journal of Educational Research*, 81, 281-287.
- Pope, J., Lehrer, B. & Stevens, J. (1980). A multiphasic reading screening procedure. *Journal of Learning Disabilities*, 13(2), 47-51.
- Posno, R. (1982). Early identification/intervention: educational foppery or promise. *Special Education in Canada*, 56, 12-17.
- Potter, M. (1987). Children and chronic illness. In A. Thomas & J. Grimes (Eds.). *Children's Needs: Psychological Perspectives*. Washington, DC: National Association of School Psychologists.
- Pullis, M. & Smith, D.C. (1981). Social cognitive development of learning disabled children: Implications of Piaget's theory for research and intervention. *Topics in Learning and Learning Disabilities*, 1, 43-55.
- Randel, M.A., Fry, M.A. & Ralls, E.M. (1977). Two readiness measures as predictors of first- and third-grade reading achievement. *Psychology in the Schools*, 14(1), 37-40.
- Rapoport, H.G. & Flint, S.H. (1976). Is there a relationship between allergy and learning disabilities? *The Journal of School Health*, XLVI, 139.
- Raudenbush, S.W. (1988). Educational applications of hierarchical linear models: A review. *Journal of Educational Statistics*, 13, 85-116.
- Raudenbush, S.W. (1989). Educational applications of hierarchical linear models: A review. *Journal of Educational Statistics*, 13(2), 85-116.
- Raudenbush, S.W. & Bryk, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.
- Raudenbush, S.W. & Bryk, A.S. (1988). Methodological advances in analysing the effects of schools and classrooms on student learning. In E.Z. Rothkopf (Ed.). *Research in*

Education, (pp.423-475). Washington, DC: American Educational Research Association.

- Rawls, D.J., Rawls, J.R. & Harrison, C.W. (1971). An investigation of six-to eleven- year-old children with allergic disorders. *Journal of Consulting and Clinical Psychology*, 36, 260-264.
- Reynolds, L., Egan, R. & Lerner, J. (1983). Efficacy of early intervention on preacademic deficits: A review of the literature. *Topics In Early Childhood Education*, 10, 47-55.
- Richardson-Koehler, V. (1988). *Teachers' beliefs about at-risk students*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.
- Ritter, D., Duffey, J. & Fischman, R. (1974). Comparisons of the intellectual estimates of the draw-a-man test, Peabody picture vocabulary test and Stanford-Binet (L-M) for kindergarten children. *Psychology in the Schools*, 11, 412-415.
- Roger, C., Smith, M.D. & Coleman, J.M. (1978). Social comparison in the classroom: The relationship between academic achievement and self-concept. *Journal of Educational Psychology*, 70, 50-57.
- Rosenberg, J.B. & Weller, G.M. (1973). Minor physical anomalies and academic performance in young school-children. *Developmental Medicine and Child Neurology*, 15, 131-135.
- Rosenshine, B.V. & Berliner, D.C. (1978). Academic engaged time. *British Journal of Teacher Education*, 4, 3-16.
- Rosenthal, R. & Jacobson, L. (1968). *Pygmalion in the Classroom*. New York, NY: Halt, Rinehart and Winston.
- Rourke, B. & Orr, R.R. (1977). Prediction of the reading and spelling performances of normal and retarded readers: four-year follow-up. *Journal of Abnormal Child Psychology*, 5, 9-20.
- Rubin, R., Balow, B., Dorle, J. & Rosen, M. (1978). Preschool prediction of low achievement in basic school skills. *Journal of Learning Disabilities*, 11(10), 62-78.
- Rutter, M. (1983). School effects on pupil progress: Research findings and policy implications. *Child Development*, 54(1), 1-29.

- Rutter, M., Maughan, B., Mortimore, P., Ouston, J. & Smith, A. (1979). *Fifteen Thousand Hours*. London, England: Open Books.
- Sabers, D., Cushing, K., & Sabers, D. (1987). Sex differences in reading and mathematics achievement for middle school students. *Journal of Early Adolescence*, 7(1), 117-128.
- Salvia, J., Clark, G., & Yssledyke, J. (1973). Teacher retention of stereotypes of exceptionality. *Exceptional Children*, 39, 651-652.
- Salvia, J. & Yssledyke, J. (1985). *Assessment in Special and Remedial Education*. Boston, MA: Houghton Mifflin.
- Samuels, S.J. (1973). Success and failure in learning to read: A critique of the research. *Reading Research Quarterly*, 8, 200-239.
- Sandoval, J. & Hughes, P.G. (1981). *Success in nonpromoted first-grade children*. Final Report. Davies, CA: University of California. (ERIC Document Reproduction Service No. ED 212-371.)
- Sattler, J. (1982). *Assessment of Children's Intelligence and Special Abilities*. (2nd Ed.). Boston, MA: Allyn & Bacon, Inc.
- Sattler, J. (1988). *Assessment of Children*. San Diego, CA: Jerome M. Sattler, Publisher.
- Satz, P. & Fletcher, J.M. (1979). Early screening tests: some uses and abuses. *Journal of Learning Disabilities*, 12(1), 65-69.
- Satz, P. & Friel, J. (1974). Some predictive antecedents of specific reading disability: A preliminary two-year follow-up. *Journal of Learning Disabilities*, 7(7), 48-55.
- Satz, P. & Friel, J. (1978). Predictive validity of an abbreviated screening battery. *Journal of Learning Disabilities*, 11(6), 20-24.
- Satz, P., Friel, J. & Rudegeair, F. (1976). Some predictive antecedents of specific reading disability: A two-, three- and four year follow-up. In J.t. Guthrie (Ed.). *Aspects of Reading Acquisition*. Baltimore, MD: Johns Hopkins Press.

- Satz, P., Taylor, H.G., Friel, J. & Fletcher, J.M. (1979). Some developmental and predictive precursors of reading disabilities: A six year follow-up. In A. Benton & D. Pearl (Eds.), *Dyslexia: A Critical Appraisal of Current Theory*. Oxford, England: Oxford University Press, (pp. 313-318).
- Schmidt, S. & Perino, J. (1985). Kindergarten screening results as predictors of academic achievement, potential, and placement in second grade. *Psychology in the Schools*, 22, 146-151.
- Schweinhart, L.J. & Weikart, D.P. (1986) Early childhood development programs: A public investment opportunity. *Educational Leadership*, 44(3), 4-12.
- Scott, L.H. (1981). Measuring intelligence with the Goodenough-Harris drawing test. *Psychological Bulletin*, 89, 483-505.
- Shapiro, E.S. (1987). *Behavioral Assessment in School Psychology*. Hillsdale, NJ: Lawrence Erlbaum Assoc.
- Shapiro, E.S. (1988). Preventing academic failure. *School Psychology Review*, 17(4), 601-613.
- Shepard, L.A. & Smith, M.L. (1986). Synthesis of research on school readiness and kindergarten retention. *Educational Leadership*, 44(3), 78-86.
- Shepard, L.A. & Smith, M.L. (1987). Effects of kindergarten retention at the end of first grade. *Psychology In the Schools*, 24, 346-357.
- Shuard, H. (1986). The relative attainment of girls and boys in mathematics in the primary years. In L. Burton (Ed.), *Girls Into Mathematics Can Go*. London: Holt, Rinehart, and Winston.
- Silver, A.A. (1978). Prevention. In A.L. Benton & D. Pearl (Eds.), *Dyslexia An Appraisal of Current Knowledge*. New York, NY: Oxford University Press.
- Simner, M.L. (1982). Printing errors in kindergarten and the prediction of academic performance. *Journal of Learning Disabilities*, 15(3), 155-159.
- Simner, M.L. (1983). The warning signs of school failure: An updated profile of the at-risk kindergarten child. *Topics in Early Childhood Special Education*, 3(3), 17-27.

- Simner, M.L. (1985). School readiness and the draw-a-man test: An empirically derived alternative to Harris scoring system. *Journal of Learning Disabilities*, 18(2), 72-81.
- Slavin, R.E. (1988). Synthesis of research on grouping in elementary and secondary schools. *Education Leadership*, 46(1), 67-77.
- Smith, H.W. & Kennedy, W. (1967). Effects of three educational programs on mentally retarded children. *Perceptual and Motor Skills*, 24, 174.
- Sorensen, A.B. & Hallinan, M.T. (1977). A reconceptualization of school effects. *Sociology of Education*, 50, 273-289.
- Stehbens, J., Kisker, C.T. & Wilson, B. (1983). School behavior and attendance during the first year of treatment of childhood cancer. *Psychology in the Schools*, 20, 223-228.
- Steinbauer, E. & Heller, M.S. (1978). The Boehm test of basic concepts as a predictor of academic achievement in grade 2. *Psychology in the Schools*, 15, 357-360.
- Stevens, L.M. (1987). Assessment and intervention with early identified, educationally at-risk children: Some methodological considerations. *European Journal of Special Needs Education*, 2(1), 1-12.
- Stevenson, H.W., Parker, T., Wilkinson, A., Hegion, A. & Fish, E. (1976). Longitudinal study of individual differences in cognitive development and scholastic achievement. *Journal of Educational Psychology*, 68, 377-400.
- Stockard, J. (1980). Sex inequities in the experience of students. In J. Stockard et al. (Eds.), *Sex Equity in Education*. New York, NY: Academic Press.
- Stone, B., Swanson, P. & Cundick, B. Special education screening system: Group achievement test. *Exceptional Children*, 55, 71-75.
- Summers, A.A. & Wolfe, B.L. (1977). Do schools make a difference? *The American Economic Review*, 67, 639-652.
- Tabachnick, B. & Fidell, L. (1983). *Using Multivariate Statistics*. New York, NY: Harper & Row Publishers.
- Telegdy, G. (1975). The effectiveness of four readiness tests as predictors of first grade academic achievement. *Psychology in the Schools*, 12, 4-11.

- Thorndike, R.L. & Hagen, E. (1971). *Measurement and Evaluation in Psychology and Education* (4th Edition). New York: Wiley.
- Thurlow, M.L., O'Sullivan, P.J. & Ysseldyke, J.E. (1986). Early screening for special education: How accurate? *Educational Leadership*, 44(3), 93-95.
- Thurlow, M.L. Ysseldyke, J.E., Graden, J.L. & Algozzine, B. (1983). What's 'special' about the special education resource room for learning disabled students? *Learning Disability Quarterly*, 6(3), 283-288.
- Tizard, B., Blatchford, P., Burke, J., Farquhar, G., & Plewis, I. (1988). *Young Children At School in the Inner City*. Hove, Sussex: Lawrence Erlbaum.
- Tjossem, T.D. (Ed.) (1976). *Intervention Strategies for High Risk Infants and Young Children*. Baltimore, MD: University Park Press.
- Tramill, J.L., Edwards, R.P. & Tramill, J.K. (1980). Comparison of the Goodenough-Harris drawing test and the WISC-R for children experiencing academic difficulties. *Perceptual and Motor Skills*, 50, 543-546.
- Walden, R. & Walkerdene, V. (1985). *Girls and mathematics: From primary to secondary schooling*. Bedford Way Papers, 24. Institute of Education, University of London: Turnaround Distribution Ltd.
- Walsh, D.J. (1988). *The two-year route to first grade: The use of testing to validate decisions based on class and age*. Paper presented at the Annual Meeting of the American Educational Research Association in New Orleans.
- Wanczycki, W. (1983). The kindergarten early identification component of Bill 82. *Special Education in Canada*, 57(1), 25-32.
- Wedell, K. & Raybould (Eds.) (1976). *The early identification of educationally 'at risk' children*. Educational Review Occasional Publications Number six. Edgbaston, Birmingham: University of Birmingham.
- Weintraub, S. (1966). What research says to the reading teacher. Sex differences in reading achievement. *The Reading Teacher*, 20, 155-165.
- Weintraub, S. (1967). What research says to the reading

teacher. Readiness measures for predicting reading achievement. *The Reading Teacher*, 20, 551-558.

- Wells, M.G & Peterson, G.V. (1978). Kindergarten behavior rating as a predictor of first-grade achievement. *Journal of Learning Disabilities*, 11(6), 17-20.
- Wendt, R.N. (1978). Kindergarten entrance assessment: Is it worth the effort? *Psychology in the Schools*, 15, 56-62.
- White, K.R. (1986). Efficacy of early identification. *The Journal of Special Education*, 19, 401-415.
- White, M. (1979). A first-grade intervention program for children at risk for reading failure. *Journal of Learning Disabilities*, 12(4), 27-32.
- Wick, J.W. (1973). *Educational Measurement. Where Are We Going and How Will We Know When We Get There?* Columbus, OH: Chas. E. Merrill Pub. Co.
- Wiley, D.E. & Harnischfeger, A. (1974). Explosion of a myth: Quantity of schooling and exposure to instruction, major educational vehicles. *Educational Researcher*, 3, 7-12.
- Willms, J.D. (1984). School effectiveness within the public and private sectors: An evaluation. *Evaluation Review*, 8, 113-135.
- Willms, J.D. (1985). The balance thesis: Contextual effects of ability on pupils' O-grade examination results. *Oxford Review of Education*, 11(1), 33-41.
- Willms, J.D. (1986). Social class segregation and its relationship to pupils' examination results in Scotland. *American Sociological Review*, 51, 224-241.
- Willms, J.D. (1987). Differences between Scottish education authorities in their examination attainment. *Oxford Review of Education*, 13(2), 211-232.
- Willms, J.D. & Chen, M. (1989). The effects of ability grouping on the ethnic achievement gap in Israeli elementary schools. *American Journal of Education*, ???, ???-???
- Willms, J.D. & Jacobsen, S. (1990). Growth in mathematics skills during the intermediate years: Sex differences and school effects. *Perspectives on Research in Mathematics Education*, 157-174.

- Willms, J.D. & Kerr, P.D. (1987). Changes in sex differences in Scottish examination results since 1976. *Journal of Early Adolescence*, 7(1), 85-105.
- Willms, J.D. & Raudenbush, S.W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement*, 26(3), 209-232.
- Wilson, B.J. & Reichmuth, M. (1985). Early screening programs: When is predictive accuracy sufficient? *Learning Disabilities Quarterly*, 8, 182-188.
- Winkler, D.R. (1975). Educational achievement and school peer group composition. *Journal of Human Resources*, 10, 189-204.
- Wolfenden, G.A. (1980). The effects of kindergarten intervention on the need for continuing individual assistance. *British Columbia Journal of Special Education*, 4, 355-363.
- Wood, C., Powell, S. & Knight, R.C. (1984). Predicting school readiness: The validity of developmental age. *Journal of Learning Disabilities*, 17(1), 8-11.
- Yoss, K.A. & Darley, F.L. (1974). Developmental apraxia of speech in children with defective articulation. *Journal of Speech and Hearing Research*, 17, 399-416.
- Ysseldyke, J.E., Thurlow, M.L., O'Sullivan, P.O. & Bursaw, R.A. (1986). Current screening and diagnostic practices in a state offering free preschool screening since 1977: Implications for the field. *Journal of Psychoeducational Assessment*, 4, 191-201.
- Zeitlin, S. (1976). *Kindergarten Screening: Early Identification of Potential High Risk Learners*. Springfield, IL: Charles C. Thomas.
- Zuzovsky, R. & Aitkin, M. (in press) A Two-level model analyses of achievement in the second IEA science study in Israel.

Appendix A

Descriptors of Physical Problems

Visual problems:

- Wears glasses
- Identified as visually impaired

Hearing problems:

- History of ear infections
- Conductive hearing loss
- Identified hearing impaired

Allergies

Chronic illness

Physically handicapped

Other

Three sources were used to obtain indicators of physical problems for the subject sample. The school-based learning assistance teachers completed a data checklist with the names of all subjects and the above categories of physical problems. The district-level resource teachers responsible for categorical services completed a similar checklist. Lastly, the school records were reviewed to locate references to the above listed problems.

Although approximately twenty percent of the sample were reported to have physical problems, the individual variables lacked explanatory power in exploratory analyses and thus, were combined to act as a control variable.

Appendix B

Number of Pupils Identified with Physical Problems
and Number Receiving Interventions
(Based on Achieved Sample - 459 Females, 498 Males)

<u>Physical Problems</u>	<u>Female</u>	<u>Male</u>	<u>Total</u>
Visually Impaired	1	0	1
Wears Glasses	46	32	78
History of Ear Infections	17	16	33
Conductive Hearing Loss	6	11	17
Hearing Impaired	8	5	13
Allergies	41	61	102
Chronic Illness	10	7	17
Physical Handicap	3	1	4
One or more physical problems	97	103	200
 <u>Interventions</u>			
Learning Assistance	23	68	91
Extended Primary	29	60	89
One or Both Interventions	42	99	141

Appendix C

Technical Information - Draw-A-Person

Authors: Goodenough's Draw-A-Man Test

Scoring system -Harris, D.B. (A modified scoring system was applied.

Date: 1963

Purpose: The Draw-a-person test is a screening instrument for use as a nonverbal measure of cognitive ability.

Age Range: 5 years through 15 years.

Administration: Can be administered to groups by classroom teacher.

Description: The child is provided a pencil and blank paper and is requested to draw a complete person. No additional instructions are provided although encouragement to complete the task may be given.

The score is calculated and points are given for inclusion, elaboration, and proportionality of body parts, not for realism or esthetic quality. The total raw score is compared to norms for the point scale or converted to standard scores ($M=100$, $SD=15$).

Reliability: Split-half reliability $r=.89$

Test-retest reliability $r=.74$ (Scott, 1981)

Interrater reliability .70 to .90 (Sattler, 1982)

Validity: With Stanford-Binet IQ Scores .36 to .55

With Stanford-Binet MA Scores .26 to .92 (Ritter, Duffy & Fischman, 1974).

Appendix D

Technical Information - Mann-Suiter Visual Motor Screen

Authors: P. Mann, P. Suiter and R. McClung

Date: (based on norms from Ilg & Ames, 1964)

Purpose: To measure the degree to which visual perception and motor behavior are integrated.

Age Range: Three years to nine years.

Description: Four designs are presented to the student. The student is given three chances to copy each design, but only the best effort is counted.

(Reliability and validity information was not available for the Mann-Suiter but form-copying tests have satisfactory reliability and validity. Statistics are reported for similar tests.)

Reliability: Test-retest of a similar form-copying test (Beery, 1982) range from .63-.92.

Validity: Correlations between some form-copying tests and readiness tests in kindergarten range from .50 to .70 (Berry, 1989; Sattler, 1988).

Normative Data (Ilg and Ames, 1964)

Design

1. A child 3 years. of age should be able to make a single circle.
2. A rectangle shape is normative for children after age 4.
3. A triangle is normative for girls after age 5 1/2 and for boys after age 6.
4. The diamond is normative for children after age 6.

Appendix E

Technical Information - Kindergarten Language Screening Test

Test: Kindergarten Language Screening Test (KLST)

Authors: Gauthier, S. and Madison, C.L

Date: 1978

Purpose: The KLST is designed to compare kindergarten pupil's language abilities to a level appropriate for their age and grade.

Age Range: 48 to 83 months.

Administration: Classroom teacher administers to individual pupils.

Description: The KLST is designed to measure both receptive and expressive language. The items on the test include: giving full name and age; name the primary colors; count to thirteen; identify major body parts; follow a sequential command and demonstrate understanding of prepositional concepts; repeat sentences of varying length and complexity and a spontaneous language sample elicited through the use of serial photographs.

A cut-off point for the total test is used to indicate likelihood of later school problems and need for further diagnostic testing.

Reliability: Test-retest correlation .87

Kuder-Richardson Reliability Coefficient .86

Validity: Construct validity - correlation .70 between KLST and Boehm Test of Basic Concepts.

Correlation with subtests of the ITPA = .36, .37, .40.

Predictive Validity: 30 children who obtained low KLST scores were administered the Northwestern Syntax Screening Test and Boehm Test of Basic Concepts. Twenty-three (82%) of the 30 low scoring students were functioning below grade level, had repeated a grade or been placed in Special Education.

Appendix F

Technical Information - Deverell Test of Letters and Numbers

Test: Deverell Test of Letters and Numbers

Date: Derived from a study in Saskatoon in the late 1960's.

Book published 1974.

Purpose: The purpose of this test is to evaluate the child's visual perception of symbolic material. The child must have learned to perceive visually the distinctive features which differentiate one symbol from another. Uppercase and lower case letters and the numerals 1-12 are included in this test.

Age Range: This is a readiness test appropriate for assessing a child's knowledge of letters regardless of age. The tables below were derived from a population of "school beginners". Children who had attended kindergarten, scored significantly than those who had not attended kindergarten.

Administration: Classroom teacher administers.

Description: A child is presented with the visual symbols for numbers, capital letters then small letters. The examiner points to a letter and the child is asked to name the number or letter.

Reliability and Validity information was not available for this test but, the literature review identified several studies which reported a task of identifying letters and numbers to be a good predictor of achievement (Chall, 1967; Jansky & de Hirsch, 1972; Adelman & Feshbach, 1971).

Percent of Correct Response in Descending Order: Ability to Name Numbers Shown

Test item	1	10	3	6	2	4	12	7	5	0	8	9
Percent	95	93	92	92	92	92	92	91	91	91	87	83

Average number of numbers known, 11; average percent 91.

Percent of Correct Response in Descending Order: Ability to Name
Capital Letters Shown

Test item	X	O	A	B	C	Z	S	H	K	J	P	D	F
Percent	93	92	85	83	80	80	78	76	74	73	72	72	71

Test item	M	W	Q	E	Y	T	V	G	N	U
Percent	68	67	65	65	63	62	60	60	59	56

Average number of capital letters known, 19; average percent, 72.

Percent of Correct Response in Descending Order: Ability to Name
Small Letters Shown

Test item	o	x	s	z	m	c	j	p	k	y	w	r	n
Percent	91	89	78	76	73	72	68	68	66	65	65	63	62

Test item	e	i	v	b	a	f	h	d	l	u	t	g	q
Percent	61	60	60	57	56	54	50	49	45	45	45	38	26

Average number of small letters known, 16; average percent, 61 letters.

Appendix G

Technical Information - Canadian Tests of Basic Skills (CTBS)

Test: Canadian Test of Basic Skills (CTBS)

Editors: Hieronymus, A.N. and King, E.M.

Date: 1976

Purpose: The CTBS was designed to serve several purposes including: to determine the developmental level of each pupil in order to adapt materials and instructional procedures to individual needs and abilities; to diagnose strengths and weaknesses of individual students and of groups; to provide information useful in making administrative decisions, to assess the effects of alternate methods of instruction conditions, experimentation and innovation.

Age Range: Grade 3 to Grade 8 (Age 9 to 14).

Administration: Administered by classroom teachers to groups of students.

Description: The CTBS are concerned only with generalized intellectual skills and abilities and do not provide separate measures of achievement in content subjects. The skills measured by the CTBS are classified into five major areas: vocabulary, reading, language, work-study and mathematics.

Reliability: Split-Halves Reliability Analysis (Tables provided in the Technical Manual) Pearson product-moment correlations .72 to .98.

Validity: The CTBS were not designed as aptitude tests nor as predictors of future academic success. "All the commonly used principles in the validation of test content have been applied in the preparation of individual test items" (p.7). The test was constructed to "correspond to the widely accepted goals of instruction in schools across the nation" (p.41). The validity of the CTBS is dependent on how closely the items on the test match the objectives of instruction within the district.

Appendix H
Characteristics of Schools in the Study

School	School Size	Rural=0 Urban=1	School Mean-Ability	Standard Deviation Mean-Ability
1	7	1	96.57	20.00
2	10	0	97.19	11.06
3	10	0	99.00	12.27
4	22	0	99.03	14.77
5	17	0	100.44	14.49
6	20	0	101.53	10.49
7	39	1	101.79	11.31
8	45	0	103.53	15.90
9	11	1	103.91	7.46
10	8	1	104.29	6.79
11	26	1	104.59	8.42
12	23	0	104.64	11.66
13	65	0	105.01	12.04
14	20	1	105.32	11.43
15	47	1	105.35	12.92
16	33	0	106.10	16.73
17	38	0	106.53	12.90
18	23	1	106.94	13.94
19	55	0	107.18	11.91
20	53	0	107.38	13.95
21	45	1	107.42	11.01
22	8	1	107.57	9.55
23	52	0	108.10	13.19
24	48	0	108.33	12.83
25	12	0	108.83	11.82
26	68	1	108.93	10.55
27	39	0	110.10	13.66
28	38	0	110.27	12.39
29	55	0	116.67	9.37
30	15	1	117.27	10.38

Appendix I

Use of Grade Equivalent Scores

The choice of the metric to be used in analysis is important to ensure interpretability of findings and to ensure necessary statistical assumptions are met. The selection of grade-equivalent (GE) scores for this study was determined after reviewing the literature regarding the use of GE's and after exploratory analyses to ensure the CTBS GE's met the assumptions necessary for statistical analysis.

The CTBS test scores for each student can be represented as raw scores, percentiles, or grade-equivalent scores. (Other types of scores, such as stanines, standard scores, or normal curve equivalents can be obtained from simple one-to-one transformations of percentiles, and so will not be considered separately.) The appropriate metric depends on the use of the reported score. Percentile scores are useful for representing relative standing amongst peers, or for comparing an individual's standing across subject areas. Percentiles are not useful for comparing school performance, or for examining the relationship between predictor variables and performance, because they have unequal intervals: they spread out raw scores in the middle of the distribution and squeeze them together at the extremes.

Raw scores and grade-equivalent scores (GE scores), which Hoover (1984) refers to as developmental scores, are appropriate for making comparisons between schools, or for examining relationships with other measures. Scale values are closer to being equal interval. The problem

with raw scores, however, is that they are less easily interpreted; their interpretation requires translation to a metric that shows the level at which various items are mastered. Grade equivalent scores do not have this problem.

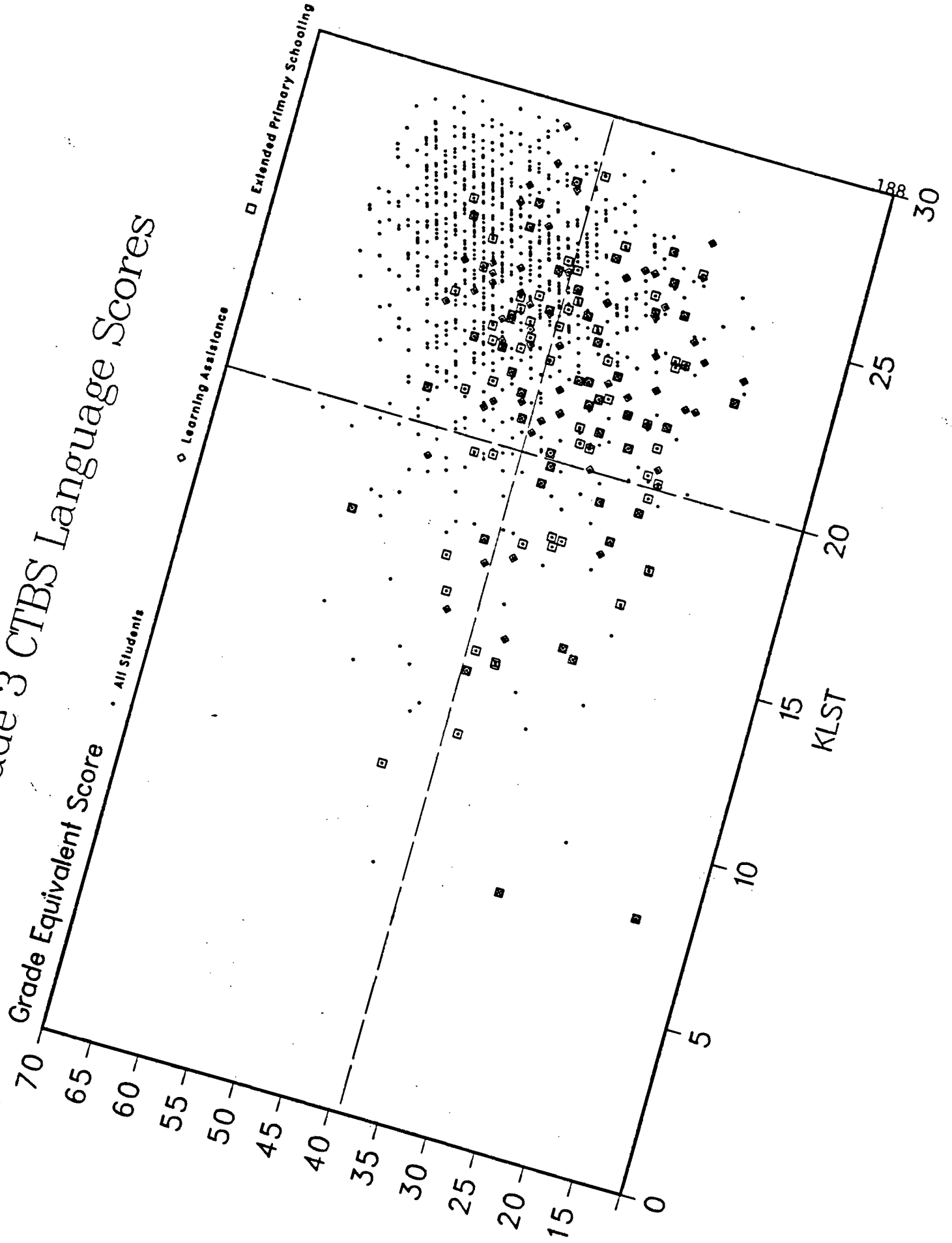
Some researchers have eschewed GE scores because they do not have equal intervals at the extreme values on a test (e.g., Angoff, 1971; Linn & Slinde, 1977; Horst, 1986). This is a problem particularly when making diagnostic assessments based on an individual's scores. However, unless a school had a large percentage of pupils who scored at the extremes, average scores for a school would be about the same if based on either raw scores or GE's. The same would be true of correlations between outcomes and predictor variables. For example, Hoover (1984) showed that the correlation between school mean GE scores and developmental scores derived from forcing raw scores onto a normal distribution within each grade, were above .995 for three separate measures at three separate grade levels.

In this study, therefore, I am confident that the relationships between kindergarten screening scores and outcome scores would not differ substantially if raw scores had been used. Similarly, if I had estimated the predicted raw score for an "at risk" child, and then transformed it to a GE score, I expect the results would be virtually the same. The problem with using GE scores (or for that matter raw scores or percentiles) for this purpose is that there are few items which cover skills mastered by these pupils. Estimates therefore are more unstable than for estimates of pupils nearer the middle of the distribution. Although less stable, they are not necessarily biased. If the estimation

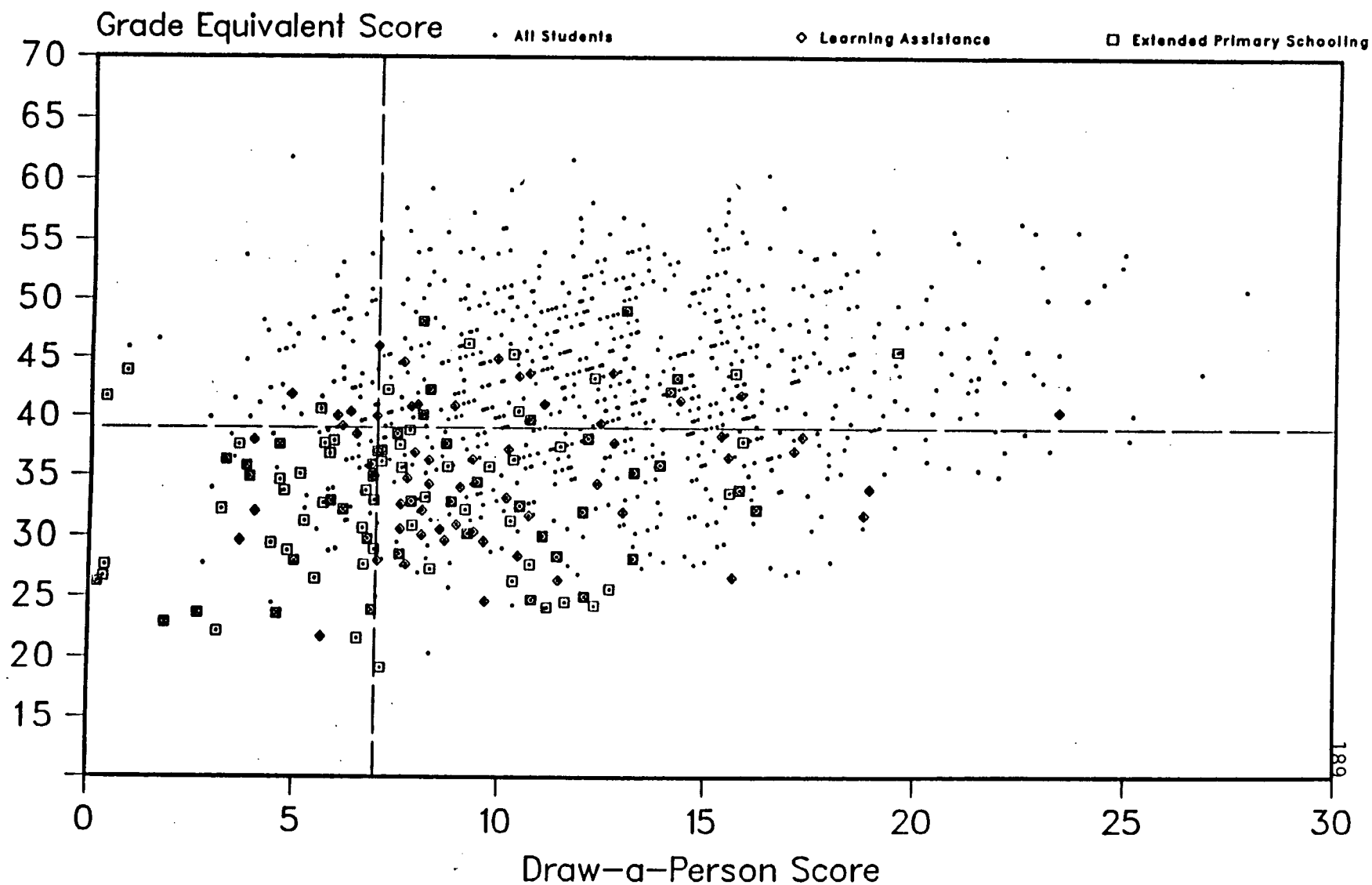
of the effects of intervention are the main purpose of a study, a researcher would achieve more accurate estimates by using the level of test appropriate for this subpopulation.

Appendix J: Data Plots Reflecting Interventions

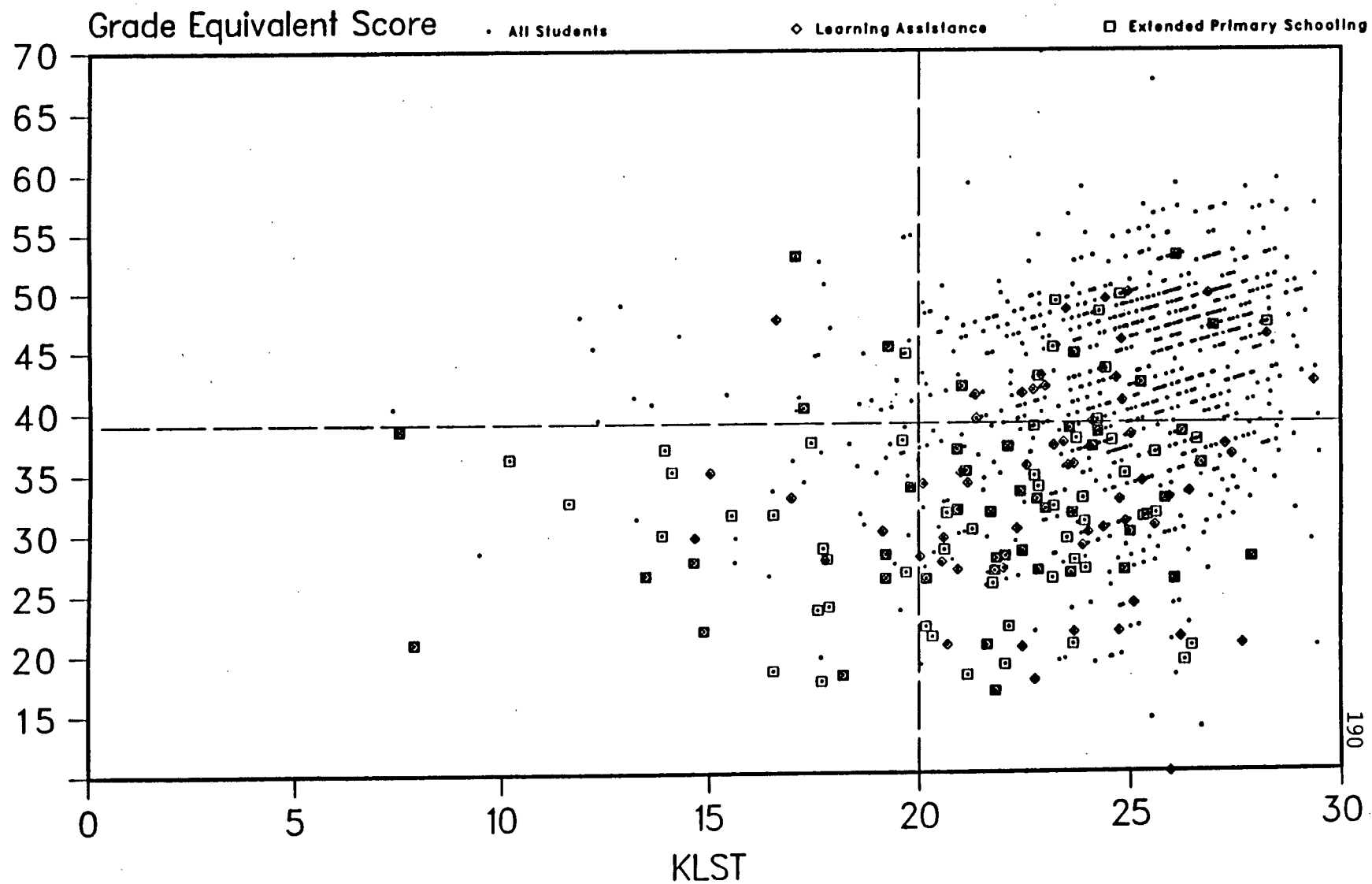
KLST vs. Grade 3 CTBS Language Scores



Draw-a-Person Scores vs. Grade 3 CTBS Mathematics Scores

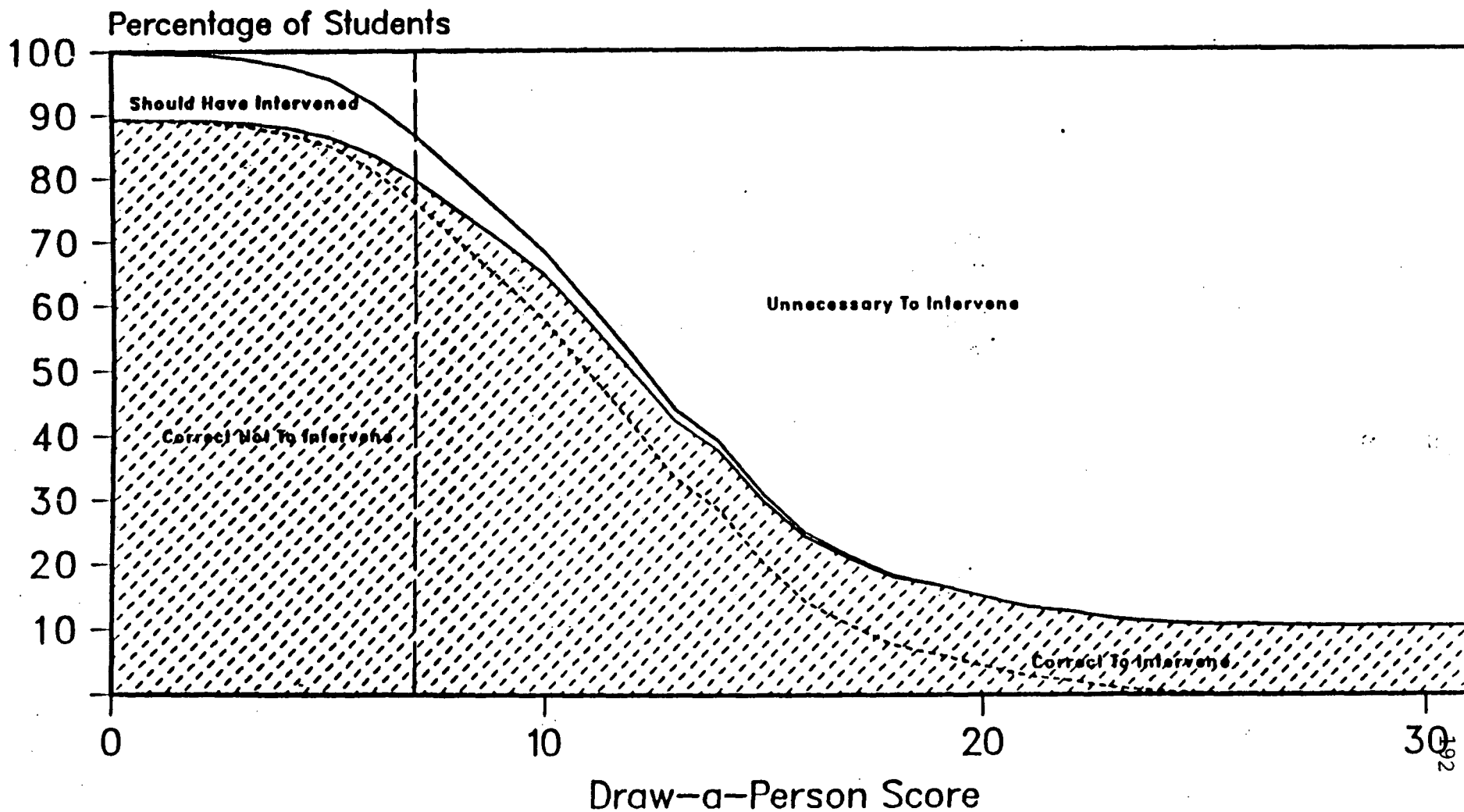


KLST vs. Grade 3 CTBS Reading Scores

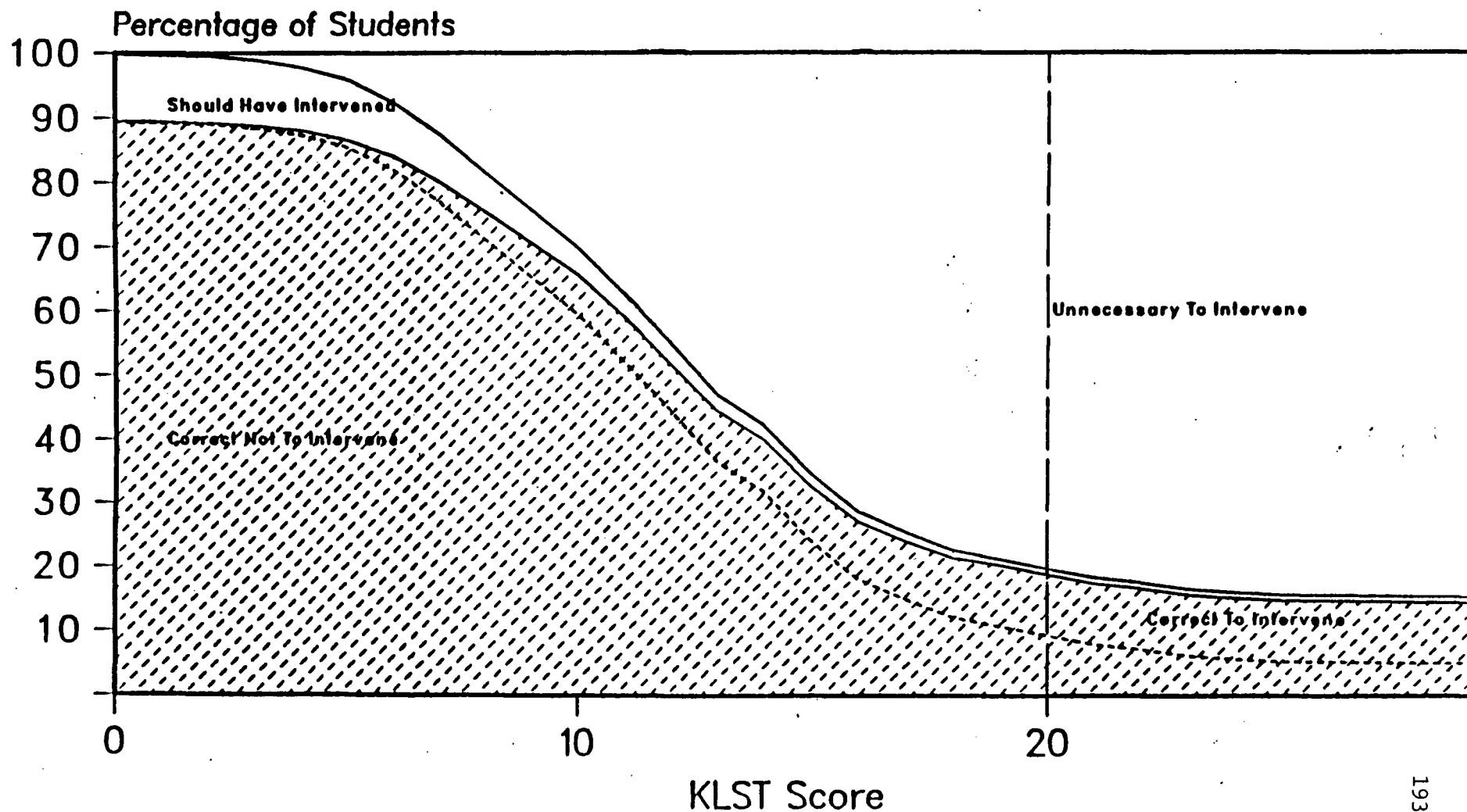


Appendix K: Graphic Representation of Predictive Utility

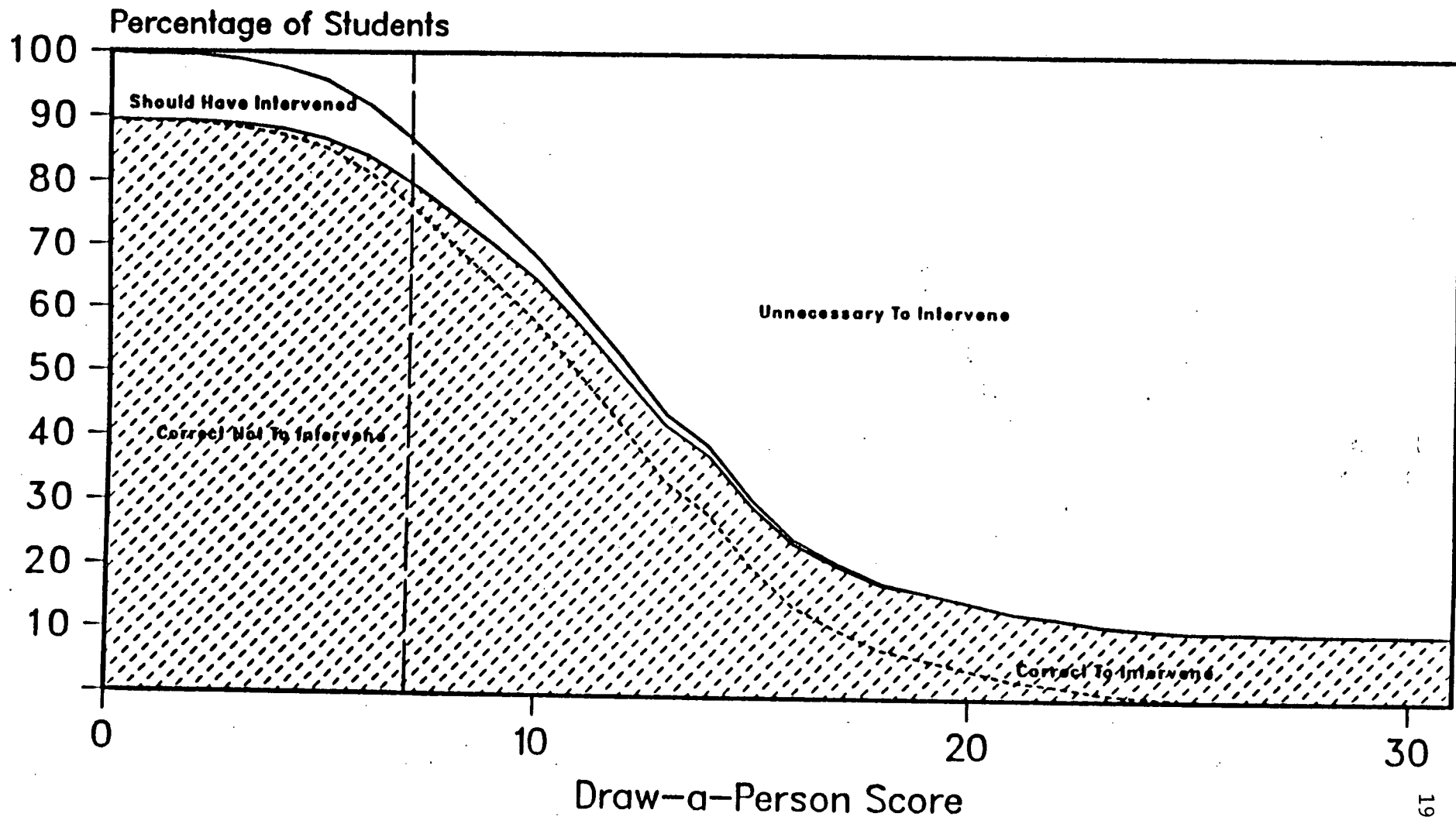
Percentages of Students Correctly Identified
As At-Risk in Reading at Grade 3 Level
For Varying Cut-off Scores on the Draw-a-Person



Percentages of Students Correctly Identified
As At-Risk in Mathematics at Grade 3 Level
For Varying Cut-off Scores on the KLST



Percentages of Students Correctly Identified
As At Risk in Mathematics at Grade 3 Level
For Varying Cut-off Scores on the Draw-a-Person



Appendix Table 1

Summary of Selected Prediction Performance Studies

<u>Study</u>	<u>Sample</u>	<u>Time</u>	<u>Screening Measures</u>	<u>Outcome Measures</u>	<u>Analysis</u>	<u>Correlation</u>
Badian (1986)	208 males, white lower mid SES	K - Gr.3	Holbrook Screening Battery WPPSI, Language Sample Naming, Visual Motor, DAP	Stanford Ach. Test	Regression ANOVA	R=.65
Badian & Serwer (1975)	300 K children 60 at risk 37 m 25 f	K - Gr.1	DAP, Primary Mental Geometric form copy, A test of letter naming MRT	WISC or WPPSI Metropolitan Ach. Test	ANOVA N/S	
Book (1974)	725 K All SES, races & religions	K - Gr.2	SIT Bender-Gestalt	Scott Foresman Reading	Correlation	gr.1 r=.99 gr.2 r=.99
Book (1980)	193 K children 105 m 88 f	K - Gr.3	Kind. Evaluation of Learning Potential(KELP), Bender- SIT	Stanford Ach. Tests Gestalt,	Correlation 2-way ANOVA	gr.1 r=.76 gr.2 r=.71 gr.3 r=.72 gr.4 r=.62
Book (1980b)	472 K, 193 gr.4 105 m, 88 f	K - Gr.4	KELP, Bender-Gestalt, SIT	Stanford Ach. Tests	Correlation 2-way ANOVA	gr.1 r=.76 gr.2 r=.71 gr.3 r=.72 gr.4 r=.62
Busch (1980)	1052	B 1- End 1	Cognitive Abilities Test Dev. Test of Visual Motor Integration (VMI), Pre-Reading Procedures, Stanford Early School Achievement, MET Readiness, Boehm Concepts Behavior Rating	Gates-MacGinitie Reading	Correlation Regression	r's=.01-.68 R=.76

<u>Study</u>	<u>Sample</u>	<u>Time</u>	<u>Screening Measures</u>	<u>Outcome Measures</u>	<u>Analysis</u>	<u>Correlation</u>
Butler (1979)	392 K children 204 m 188 f Mean age 5yr.8mo. Representative SES	K - 2	Ayres-Sensory Motor Schonell Integration, Finger ID, Doren Fine Motor Coord. Frostig-Visual Perception Stanford Auditory Tests, PPVT	Reading Test t- tests Diag. Reading Test of Word Recognition, Diagnostic Test	Correlation Regression	r=.24 - .57 R=.81
Buttram, Covert, & Hayes (1976)	121 K children private school	K - K	Hayes Early ID Listening Response	MRT	Correlation	gr.1 r=.62 gr.2 r=.81
Duffy, Ritter, & Fedner (1976)	82 caucasian 35 m 47 f SES lower-up.mid.	K - 2	Visual Motor Integration Draw-a-Man (Goodenough)	Stanford Ach. Test	Correlation	r=.20 - .46
Dunleavy, Hansen, Szasz & Baade (1981)	141 kindergartner 62 m 79 f SES low, mid, high	K	Human Figure Drawing	MRT, Met. Ach. Test	ANOVA	
Ferinden, Jacobsen & Linden (1970)	67 K children	K - 1	WRAT, Evanston Early ID Scale, Bender Gestalt, MRT	Pre- post same	Correlation	Pre. r=.28 Post r=.76
Eaves, Kendall, & Crichton (1972)	228 children 25 (random sel) 25 (matched)	K	DAP, Bender, Name printing, pencil use, Wepman, words in a story, word recognition	Beery VMI, WISC, ITPA, Kephart Motor Survey MRT	t- tests Correlation Regression reported)	r= .40 - .66 (means
Feshbach, Adelman, & Fuller (1974)	888 K (children IQ <90 eliminated)	K - 1	WPPSI, Otis Lennon Group IQ, deHirsch Predictive Index, Student Rating Scale, Prim.A,	Cooperative Primary Gates Mac Ginitie Reading Comprehension	Correlation Regression	R= .57 w/Gates R= .63 w/ de H i r s c h Prediction 74 % 196

<u>Study</u>	<u>Sample</u>	<u>Time</u>	<u>Screening Measures</u>	<u>Outcome Measures</u>	<u>Analysis</u>	<u>Correlation</u>
Fletcher & Satz (1982)	195 white males	K - 6	PPVT, VMI, Recognition Alphabet Recitation	MRT (Four groups Discriminant Severely disabled = 53 Mildly disabled = 28 Average = 79 Superior = 35	Function Analysis	77% accuracy
Galante, Flye & Stephens (1972)	114 K mid to upper class	K - 6	Birth & developmental history, Medical examination	Dominance test, speech evaluation, group IQ, Calif.Ach. Tests, Stanford Ach. Tests, WISC	Three groups	
Glazzard (1982)	107 K 50 m 57 f	K - 6	Teacher rating, Gates MacGinitie, Readiness Skills each grade level	Gates- MacGinitie Reading Pairwise	Regression ANCOVA	gr.1 R= .77 R= .75 R= .57
Goldman & Velasco (1980)	123 K 69 m 51 f	K	Draw-a-person	Koppitz Scale of Emotional Indicators	Inter-rater Rel. Correlation	r=.35 -.64
Haring & Ridgway (1967)	1200 K	K	Stanford Binet, Teacher Observation	ITPA, Detroit TLA, WRAT, Aud. Discrim., Beery VMI, Perc.Motor Survey	Correlation Prin. Components	r=.49 -.71 1st.factor 20%
Hartlage & Lucas (1972)	44 K	K - 1	Test-visual seq., auditory seq., visual w/WRAT & auditory space w/rating	Wrat, Bender- Gestalt, Draw-a-person	Correlation Regression	R=.78 R=.77
Hartlage & Lucas (1973)	1132 K	B1 - E1	MRT	WRAT Teacher ranks	t-tests ANOVA	
Jacob, Snider & Wilson (1988)	463 K 51% m 94% caucasian	K - 1	Clymer-Barrett Readiness Test	Stanford Ach. Test Reading Matrix	Regression specificity sensitivity	R=.65 .97 .43

<u>Study</u>	<u>Sample</u>	<u>Time</u>	<u>Screening Measures</u>	<u>Outcome Measures</u>	<u>Analysis</u>	<u>Correlation</u>
LaTorre, Hawkhead, Kawahua & Bilow (1982)	796 K Teacher's predictions	K - 1	McCarthy Screening Test Ginn 720 series level	Evaluative survey	ANOVA Disc. Function average Matrix	77.1% poor 67.9% Overall hit 77%
Lewis (1980)	86 75 44 m 37 m	K - 2	English Picture Vocab. Test Test, Croydon checklist	Young's Group Reading Test	Reported as % Use of screening did not improve hits beyond chance.	
Lindquist (1982)	351	K -1,2,3	Denver Developmental Screening Test	Gates-MacGinitie Reading Test	Correlation	r=.009 - .32 gr.1 r=.46 gr.2 r=.29 gr.3 r=.27
Meyers, Attwell & Orpet (1968)	57 K 25 m 32 f	K - 5	13 Individually admin. tests (motor, perception, Ravens matrices & Binet digits)	Calif. Ach. Test Calif. Mental Maturity	Correlation Regression	all sig R=.63 -.76
Miller (1988)	338	Pre.K - 3	Miller Assessment for Preschoolers	Retention Teacher observation-poor Received special services	Correlation t-tests	
Miller & Schouten (1988)	338 184 m 154 f	Pre.K - 3 309 normal 29 at-risk	Miller Assessment for Preschoolers	Woodcock- Johnson Ach. Reading, Math & Lang.	Correlation Regression	R=.18 -.59
Pope, Lehrer & Stevens (1980)	46 105 35 m 42 m 11 f 63 f	K - 5 high & low	Kind. Reading Screening Battery, WRAT, Teacher checklist, SIT	Woodcock Reading Mastery	ANOVA Correlation	Differed Sig r=.50
Randel, Fry & Ralls (1977)	153 K	K - 1 & 3 mid class	DAP, counting, Gen. Info., MRT, Prim. Mental Abilities	Stanford Ach. Test	Regressions	R=.34 -.57

<u>Study</u>	<u>Sample</u>	<u>Time</u>	<u>Screening Measures</u>	<u>Outcome Measures</u>	<u>Analysis</u>	<u>Correlation</u>
Rourke & Orr (1977)	23 normal readers 19 retarded "	182 -586	MAT, WISC, PPVT, WRAT, underlining test	Met. Ach Test, WRAT, PPVT, underlining test	Regression Disc.	R=.82 73.7%
Rubin, Balow, Dorle, & Rosen (1978)	732	K - 1	MRT	Stanford Ach. Test	Correlation	r=.50 -.70
Satz & Friel (1974)	497	K - 1	22 Variables -PPVT, VMI, white,male alphabet recitation,	10 item scale of reading no readiness-advanced reader	Disc. Regression	84.4% R=.81
Satz & Friel (1978)	28 104 13 m 54 m 15 f 50 f	K - 2 Black/ White	Modified Screening, finger localization, VMI, PPVT alphabet recitation, etc.	IOTA, classroom reading level materials	Matrix	Accurate 90% severe 69% mild
Schmidt & Perino (1985)	378 K 201 m 177 f	K - 2	Vane-Test of Language Vane Kindergarten Test Draw-a-man, perc.motor	Met. Ach Test Otis- Lennon Ability Test	Regressions	R=.48 -.50
Simner (1982)	166 K 79 m 87 f	K - K	41 reversible letters and numbers	Teacher's rating rank ordering	Correlation Matrix	r=.53 -.63
Silver, Hagin Beecher (1978)	2319 K 51.4% m 48.6% f	K -1,2,3,4	SEARCH -ten component test	WRAT oral reading	Matrix	10% false & neg.
Simner (1985)	118 61 m 57 f	K - 1	Draw-a-man	Criterion. Ref. Read & & Math, WRAT, Simner Printing, Developmental Tasks for K Readiness	Correlation	rs=.52 -.57
Stevenson, et al. (1976)	255 K 133 m 122 f SES Ave-Hi Mid	K -1,2,3	11 Cognitive measures 14 Psychometric tasks Teacher ratings -13 vars Ach Test (Gr.3)	WRAT (Gr.2), MET Readiness (Gr.1), Gray Oral Reading, Stanford	t-tests Correlation ANCOVA Regression	R=.56 -.77

<u>Study</u>	<u>Sample</u>	<u>Time</u>	<u>Screening Measures</u>	<u>Outcome Measures</u>	<u>Analysis</u>	<u>Correlation</u>
Wells & Peterson (1978)	111 K	K - 1	KELP -Kindergarten Teacher Checklist for Potential Learning Problems	Devereux Behavior Rating Iowa Test of Basic Skills	Correlation t-tests Regression	r=-.33--.43 27% variance

Note: WPPSI - Wechsler Preschool and Primary Scale of Intelligence
WISC (R) - Wechsler Intelligence Scale for Children (Revised)
DAP - Draw-a-Person
KELP - Kindergarten Evaluation of Learning Problems
MRT - Metropolitan Readiness Test
SIT - Slosson Intelligence Test
VMI - Beery-Developmental Test of Visual Motor Integration
WRAT - Wide Range Achievement Test

Appendix Table 2
HLM Results for Attrition

Fixed Effects	Effects (SE)
Average within school equation:	
Intercept	-.08 (0.08)
Slope (Study)	.09 (0.07)
Estimates of Parameter Variance	Estimate (x ²) df 29
Intercept	.09** (68.41)
Slope (Study)	.07** (56.56)

Appendix Table 3

HLM Models Explaining Variation in Draw-A-Person/Grade 3 Reading Relationships								
Fixed Effects	Model I		Model II		Model III		Model IV	
	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	38.31**	(0.66)	38.46**	(0.66)	40.04**	(0.59)	37.83**	(0.71)
K-Screen Slope	.57**	(0.07)	.51**	(0.07)	.39**	(0.07)	.38**	(0.08)
Age on Entry			.07	(0.08)	.06	(0.08)	.08	(0.08)
Gender			1.26*	(0.54)	.87	(0.51)	.88	(0.52)
Physical Problems			.70	(0.68)	1.00	(0.65)	1.04	(0.64)
Extended Primary Schooling					-3.90**	(1.23)	-4.21**	(1.11)
Learning Assistance					-5.53**	(1.29)	-4.84**	(1.30)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.36**	(0.08)
On K-Screen/Ach Relationship: Attrition							-.16	(0.25)
Estimates of Parameter Variance	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	7.35**		7.20**		4.33**		2.04	
K-Screen/Ach Relationship	.04**		.04*		.05		.05	
Extended Primary Schooling					12.92		5.97	
Learning Assistance					13.49*		15.19*	
Model Statistics:								
Maximum likelihood estimate of σ^2	62.83		62.65		56.43		56.76	
R ² : Percent of Total Pupil Level Variance Explained:	15.15		15.40		23.79		23.35	
R ² : Percent of Residual Parameter Variance Explained								
On Adjusted Levels of Achievement							52.88	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 4

HLM Models Explaining Variation in Draw-A-Person/Grade 3 Mathematic Relationships

	Model I		Model II		Model III		Model IV	
Fixed Effects	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	39.48**	(0.65)	39.58**	(0.66)	41.12**	(0.65)	39.16**	(0.68)
K-Screen Slope	.35**	(0.06)	.32**	(0.06)	.21**	(0.06)	.19**	(0.06)
Age on Entry			.18*	(0.07)	.16*	(0.07)	.18*	(0.07)
Gender			-.06	(0.47)	-.42	(0.45)	-.44	(0.45)
Physical Problems			-.09	(0.59)	.20	(0.56)	.12	(0.59)
Extended Primary Schooling					-4.52**	(0.93)	-4.42**	(0.90)
Learning Assistance					-5.13**	(0.86)	-4.80**	(0.86)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.37**	(0.07)
On K-Screen/Ach Relationship: Attrition							-.06	(0.17)
Estimates of Parameter Variance	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	8.22**		8.40**		7.58**		3.76**	
K-Screen/Ach Relationship	.03		.03		.03		.03	
Extended Primary Schooling					3.20		1.93	
Learning Assistance					.66		.72	
Model Statistics:								
Maximum likelihood estimate of σ^2	48.38		48.28		43.58		43.56	
R ² : Percent of Total Variance Explained:	11.67		11.85		20.43		41.18	
R ² : Percent of Residual Parameter Variance Explained								
On Adjusted Levels of Achievement							50.40	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 5

HLM Models Explaining Variation in Draw-A-Person/Grade 3 Vocabulary Relationships

	Model I		Model II		Model III		Model IV	
Fixed Effects	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	38.62**	(0.59)	38.72**	(0.60)	40.15**	(0.56)	38.16**	(0.67)
K-Screen Slope	.53**	(0.07)	.50**	(0.07)	.40**	(0.07)	.38**	(0.07)
Age on Entry			.14	(0.08)	.13*	(0.07)	.15	(0.07)
Gender			.15	(0.52)	-.22	(0.50)	-.16	(0.50)
Physical Problems			.06	(0.64)	.42	(0.63)	.32	(0.66)
Extended Primary Schooling					-3.38**	(1.07)	-3.47**	(0.97)
Learning Assistance					-5.22**	(1.07)	-4.33**	(1.08)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.33**	(0.06)
On K-Screen/Ach Relationship: Attrition							-.16	(0.21)
Estimates of Parameter Variance	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	5.13**		5.16**		3.67		2.63	
K-Screen/Ach Relationship	.05*		.04*		.04*		.05	
Extended Primary Schooling					5.33		1.40	
Learning Assistance					5.49		7.27	
Model Statistics:								
Maximum likelihood estimate of σ^2	57.84		57.84		53.76		53.55	
R ² : Percent of Total Variance Explained:	13.40		13.39		19.51		19.82	
R ² : Percent of Residual Parameter Variance Explained								
On Adjusted Levels of Achievement							28.33	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 6

HLM Models Explaining Variation in Draw-A-Person/Grade 3 Language Relationships								
Fixed Effects	Model I		Model II		Model III		Model IV	
	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	40.17**	(0.79)	40.65**	(0.76)	42.55**	(0.73)	39.85**	(0.85)
K-Screen Slope	.61**	(0.08)	.52**	(0.07)	.37**	(0.07)	.37**	(0.07)
Age on Entry			.17*	(0.07)	.15*	(0.07)	.16*	(0.07)
Gender			2.05**	(0.50)	1.74**	(0.47)	1.68**	(0.47)
Physical Problems			-.65	(0.64)	-.34	(0.60)	-.46	(0.66)
Extended Primary Schooling					-6.26**	(1.10)	-6.18**	(1.06)
Learning Assistance					-5.77**	(1.16)	-5.35**	(1.15)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.46**	(0.10)
On K-Screen/Ach Relationship: Attrition							.01	(0.23)
Estimates of Parameter Variance	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	13.36**		12.46**		10.41**		6.57**	
K-Screen/Ach Relationship	.07**		.05**		.04		.04	
Extended Primary Schooling					8.47		7.04	
Learning Assistance					10.55		10.40	
Model Statistics:								
Maximum likelihood estimate of σ^2	55.14		54.21		46.77		46.55	
R ² : Percent of Total Variance Explained:	20.47		21.80		32.54		32.86	
R ² : Percent of Residual Parameter Variance Explained								
On Adjusted Levels of Achievement							36.88	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 7

HLM Models Explaining Variation in KLST/Grade 3 Reading Relationships								
Fixed Effects	Model I		Model II		Model III		Model IV	
	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	37.52**	(0.67)	37.68**	(0.67)	39.33**	(0.61)	37.21**	(0.77)
K-Screen Slope	.86**	(0.08)	.80**	(0.10)	.61**	(0.09)	.60**	(0.09)
Age on Entry			.11	(0.08)	.08	(0.07)	.10	(0.07)
Gender			1.35*	(0.53)	1.02	(0.51)	1.01	(0.51)
Physical Problems			.66	(0.67)	.97	(0.65)	1.00	(0.66)
Extended Primary Schooling					-3.46**	(1.18)	-3.70**	(1.11)
Learning Assistance					-5.41**	(1.19)	-4.78**	(1.19)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.34**	(0.08)
On K-Screen/Ach Relationship: Attrition							-.10	(0.27)
<hr/>								
Estimates of Parameter Variance	Estimate		Estimate		Estimate		Estimate	
<hr/>								
Residual Parameter Variance:								
Achievement	6.85**		6.84**		3.76		2.87	
K-Screen/Ach Relationship	.01		.02		.01		.02	
Extended Primary Schooling					9.15		5.70	
Learning Assistance					9.26		9.83	
<hr/>								
Model Statistics:								
Maximum likelihood estimate of σ^2	62.06		61.71		56.84		56.94	
R ² : Percent of Total Variance Explained:	16.18		16.67		23.24		23.10	
R ² : Percent of Residual Parameter Variance Explained								
On Adjusted Levels of Achievement							23.67	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 8

HLM Models Explaining Variation in KLST/Grade 3 Mathematics Relationships								
Fixed Effects	Model I		Model II		Model III		Model IV	
	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	38.54**	(0.68)	38.61**	(0.68)	40.14**	(0.69)	38.06**	(0.74)
K-Screen Slope	.66**	(0.09)	.64**	(0.09)	.48**	(0.09)	.46**	(0.09)
Age on Entry			.17*	(0.07)	.14*	(0.07)	.15*	(0.07)
Gender			-.26	(0.45)	-.60	(0.44)	-.63	(0.43)
Physical Problems			-.22	(0.58)	.09	(0.56)	.08	(0.56)
Extended Primary Schooling					-3.59**	(0.94)	-3.70**	(0.90)
Learning Assistance					-4.92**	(0.85)	-4.63**	(0.86)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.34**	(0.07)
On K-Screen/Ach Relationship: Attrition							-.10	(0.23)
Estimates of Parameter Variance								
	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	8.72**		8.35**		8.18*		5.36*	
K-Screen/Ach Relationship	.09*		.08*		.06		.08	
Extended Primary Schooling					3.34		2.31	
Learning Assistance					.87		.98	
Model Statistics:								
Maximum likelihood estimate of σ^2	46.38		46.24		42.54		42.35	
R ² : Percent of Total								
Variance Explained:	15.30		15.56		22.33		22.65	
R ² : Percent of Residual Parameter								
Variance Explained								
On Adjusted Levels of Achievement							34.47	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 9

HLM Models Explaining Variation in KLST/Grade 3 Vocabulary Relationships

Fixed Effects	Model I		Model II		Model III		Model IV	
	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	37.49**	(0.59)	37.58**	(0.59)	38.95**	(0.58)	37.06**	(0.69)
K-Screen Slope	.90**	(0.08)	.87**	(0.09)	.72**	(0.08)	.70**	(0.09)
Age on Entry			.16*	(0.07)	.14	(0.07)	.16	(0.07)
Gender			.15	(0.50)	-.19	(0.49)	-.16	(0.49)
Physical Problems			.04	(0.63)	.31	(0.62)	.24	(0.64)
Extended Primary Schooling					-2.61*	(0.97)	-2.66*	(0.95)
Learning Assistance					-5.08**	(0.98)	-4.45**	(1.00)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.31**	(0.07)
On K-Screen/Ach Relationship: Attrition							.10	(0.25)
Estimates of Parameter Variance	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	4.71**		4.43**		3.06		1.94	
K-Screen/Ach Relationship	.02		.02		.01		.02	
Extended Primary Schooling					1.33		1.04	
Learning Assistance					2.62		4.35	
Model Statistics:								
Maximum likelihood estimate of σ^2	56.02		55.95		52.88		52.64	
R ² : Percent of Total								
Variance Explained:	16.11		16.22		20.81		21.17	
R ² : Percent of Residual Parameter								
Variance Explained								
On Adjusted Levels of Achievement							36.60	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 10

HLM Models Explaining Variation in KLST/Grade 3 Language Relationships								
	Model I		Model II		Model III		Model IV	
Fixed Effects	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	39.77**	(0.82)	40.22**	(0.82)	42.12**	(0.74)	39.45**	(0.86)
K-Screen Slope	.86**	(0.10)	.76**	(0.09)	.54**	(0.08)	.54**	(0.09)
Age on Entry			.21**	(0.07)	.18**	(0.07)	.18*	(0.07)
Gender			2.23**	(0.49)	1.95**	(0.46)	1.84**	(0.46)
Physical Problems			-.80	(0.63)	-.48	(0.60)	-.68	(0.69)
Extended Primary Schooling					-5.99**	(1.21)	-5.84**	(1.21)
Learning Assistance					-5.92**	(1.08)	-5.32**	(1.00)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.46**	(0.09)
On K-Screen/Ach Relationship: Attrition							.27	(0.27)
<hr/>								
Estimates of Parameter Variance	Estimate		Estimate		Estimate		Estimate	
<hr/>								
Residual Parameter Variance:								
Achievement	14.16**		13.86**		9.98**		6.40**	
K-Screen/Ach Relationship	.08		.07		.02		.06	
Extended Primary Schooling					13.86*		14.04*	
Learning Assistance					6.75		3.85	
<hr/>								
Model Statistics:								
Maximum likelihood estimate of σ^2	55.46		53.97		47.04		46.62	
R ² : Percent of Total Variance Explained:	19.99		22.14		32.15		32.75	
R ² : Percent of Residual Parameter Variance Explained								
On Adjusted Levels of Achievement							35.87	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 11

HLM Models Explaining Variation in Devere11/Grade 3 Reading Relationships								
	Model I		Model II		Model III		Model IV	
Fixed Effects	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	38.23**	(0.62)	38.30**	(0.61)	39.89**	(0.67)	37.72**	(0.78)
K-Screen Slope	.47**	(0.06)	.44**	(0.06)	.33**	(0.07)	.33*	(0.06)
Age on Entry			.16	(0.08)	.13	(0.08)	.15	(0.07)
Gender			1.79**	(0.52)	1.29*	(0.51)	1.27*	(0.50)
Physical Problems			.82	(0.67)	1.17	(0.65)	1.17	(0.68)
Extended Primary Schooling					-2.89**	(1.29)	-3.11*	(1.21)
Learning Assistance					-6.50**	(1.23)	-5.80**	(1.25)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.33**	(0.08)
On K-Screen/Ach Relationship: Attrition							-.17	(0.17)
Estimates of Parameter Variance	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	5.48**		5.02**		6.05		3.48*	
K-Screen/Ach Relationship	.02		.02		.04		.03	
Extended Primary Schooling					13.08		7.98*	
Learning Assistance					10.74		11.84*	
Model Statistics:								
Maximum likelihood estimate of σ^2	63.47		62.69		57.02		56.97	
R ² : Percent of Total Variance Explained:	14.27		15.33		22.99		23.05	
R ² : Percent of Residual Parameter Variance Explained								
On Adjusted Levels of Achievement							42.48	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 12

HLM Models Explaining Variation in Devere11/Grade 3 Mathematics Relationships

	Model I		Model II		Model III		Model IV	
Fixed Effects	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	38.85**	(0.49)	38.92**	(0.48)	40.46**	(0.46)	38.17**	(0.69)
K-Screen Slope	.40**	(0.05)	.38**	(0.05)	.26**	(0.05)	.27**	(0.06)
Age on Entry			.21**	(0.07)	.19*	(0.07)	.19**	(0.07)
Gender			.09	(0.45)	-.28	(0.44)	-.37	(0.44)
Physical Problems			-.08	(0.58)	.20	(0.56)	.26	(0.57)
Extended Primary Schooling					-3.30**	(0.98)	-3.37**	(0.92)
Learning Assistance					-5.22**	(1.01)	-5.02**	(0.90)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.37**	(0.06)
On K-Screen/Ach Relationship: Attrition							.00	(0.15)
Estimates of Parameter Variance	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	2.79		2.52		1.06		3.58*	
K-Screen/Ach Relationship	.02		.02		.01		.03	
Extended Primary Schooling					3.60		1.78	
Learning Assistance					5.67		2.22	
Model Statistics:								
Maximum likelihood estimate of σ^2	47.11		46.85		43.42		42.87	
R ² : Percent of Total Variance Explained:	13.97		14.45		20.71		21.71	
R ² : Percent of Residual Parameter Variance Explained								
On Adjusted Levels of Achievement							-237.73	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 13

HLM Models Explaining Variation in Devereil/Grade 3 Vocabulary Relationships								
Fixed Effects	Model I		Model II		Model III		Model IV	
	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	38.39**	(0.43)	38.46**	(0.43)	39.80**	(0.50)	37.98**	(0.64)
K-Screen Slope	.46**	(0.06)	.44**	(0.06)	.34**	(0.06)	.33**	(0.05)
Age on Entry			.23**	(0.08)	.21*	(0.07)	.23**	(0.07)
Gender			.67	(0.05)	.24	(0.49)	.26	(0.49)
Physical Problems			.24	(0.63)	.57	(0.62)	.44	(0.67)
Extended Primary Schooling					-2.22	(1.08)	-2.25*	(1.00)
Learning Assistance					-5.52**	(1.07)	-4.69**	(1.09)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.30**	(0.06)
On K-Screen/Ach Relationship: Attrition							-.30	(0.15)
<hr/>								
Estimates of Parameter Variance	Estimate		Estimate		Estimate		Estimate	
<hr/>								
Residual Parameter Variance:								
Achievement	.71		.51		.94		.80	
K-Screen/Ach Relationship	.03		.03		.02		.00	
Extended Primary Schooling					3.55		.80	
Learning Assistance					4.58		6.78	
<hr/>								
Model Statistics:								
Maximum likelihood estimate of σ^2	58.32		57.92		54.51		54.31	
R ² : Percent of Total Variance Explained:	12.67		13.28		18.38		18.67	
R ² : Percent of Residual Parameter Variance Explained								
On Adjusted Levels of Achievement							14.89	
<hr/>								
Note: * Significant at the .05 level.								
**Significant at the .01 level.								

Appendix Table 14

HLM Models Explaining Variation in Devere11/Grade 3 Language Relationships								
	Model I		Model II		Model III		Model IV	
Fixed Effects	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	39.93**	(0.67)	40.22**	(0.69)	42.16**	(0.81)	39.36**	(0.88)
K-Screen Slope	.55**	(0.06)	.51**	(0.06)	.36**	(0.07)	.36**	(0.06)
Age on Entry			.25**	(0.07)	.21**	(0.46)	.21**	(0.06)
Gender			2.56**	(0.48)	2.18**	(0.46)	2.09**	(0.46)
Physical Problems			-.56	(0.63)	-.24	(0.59)	-.41	(0.65)
Extended Primary Schooling					-5.17**	(1.06)	-4.92**	(1.04)
Learning Assistance					-6.42**	(1.07)	-5.81**	(1.04)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.45**	(0.09)
On K-Screen/Ach Relationship: Attrition							.17	(0.18)
Estimates of Parameter Variance	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	8.13**		8.81**		13.12		7.77**	
K-Screen/Ach Relationship	.03*		.03*		.04		.05*	
Extended Primary Schooling					5.54		4.42	
Learning Assistance					7.25		5.96	
Model Statistics:								
Maximum likelihood estimate of σ^2	54.63		52.54		46.36		46.15	
R ² : Percent of Total Variance Explained:	21.18		24.21		33.12		33.42	
R ² : Percent of Residual Parameter Variance Explained								
On Adjusted Levels of Achievement							40.78	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 15

HLM Models Explaining Variation in MS/Grade 3 Reading Relationships								
	Model I		Model II		Model III		Model IV	
Fixed Effects	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	40.64**	(0.50)	40.57**	(0.49)	41.80**	(0.44)	39.55**	(0.60)
K-Screen Slope	1.89**	(0.35)	1.70**	(0.35)	.98*	(0.35)	.95*	(0.34)
Age on Entry			.16	(0.08)	.13	(0.08)	.15	(0.08)
Gender			2.19**	(0.53)	1.61**	(0.51)	1.61**	(0.51)
Physical Problems			.76	(0.69)	1.09	(0.66)	1.17	(0.67)
Extended Primary Schooling					-4.66**	(1.20)	-4.95**	(1.09)
Learning Assistance					-5.78**	(1.24)	-5.02**	(1.24)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.35**	(0.08)
On K-Screen/Ach Relationship: Attrition							-2.47	(1.59)
Estimates of Parameter Variance								
	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	4.55**		4.06**		2.73**		.64	
K-Screen/Ach Relationship	.16		.15		.21		.14	
Extended Primary Schooling					10.41		4.50	
Learning Assistance					10.46		11.03*	
Model Statistics:								
Maximum likelihood estimate of σ^2	67.32		66.14		59.36		59.45	
R ² : Percent of Total Variance Explained:	9.07		10.67		19.83		19.71	
R ² : Percent of Residual Parameter Variance Explained								
On Adjusted Levels of Achievement							76.56	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 16

HLM Models Explaining Variation in MS/Grade 3 Mathematics Relationships								
	Model I		Model II		Model III		Model IV	
Fixed Effects	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	40.67**	(0.44)	40.68**	(0.44)	41.78**	(0.41)	39.71**	(0.53)
K-Screen Slope	2.29**	(0.30)	2.16**	(0.30)	1.56**	(0.31)	1.53**	(0.31)
Age on Entry			.19*	(0.07)	.16*	(0.07)	.16*	(0.07)
Gender			.35	(0.45)	-.13	(0.44)	-.11	(0.44)
Physical Problems			-.17	(0.59)	.09	(0.57)	.07	(0.60)
Extended Primary Schooling					-4.41**	(0.98)	-4.40**	(0.90)
Learning Assistance					-5.12**	(0.89)	-4.57**	(0.86)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.33**	(0.07)
On K-Screen/Ach Relationship: Attrition							-1.50	(1.42)
Estimates of Parameter Variance								
	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	3.61**		3.56**		2.62**		.69	
K-Screen/Ach Relationship	.13		.14		.38		.32	
Extended Primary Schooling					5.36		2.17	
Learning Assistance					1.49		1.64	
Model Statistics:								
Maximum likelihood estimate of σ^2	47.98		47.74		43.18		43.24	
R ² : Percent of Total Variance Explained:	12.38		12.81		21.15		21.03	
R ² : Percent of Residual Parameter Variance Explained								
On Adjusted Levels of Achievement							73.66	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 17

HLM Models Explaining Variation in MS/Grade 3 Vocabulary Relationships								
	Model I		Model II		Model III		Model IV	
Fixed Effects	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	40.78**	(0.41)	40.78**	(0.41)	41.85**	(0.39)	39.78**	(0.53)
K-Screen Slope	1.92**	(0.33)	1.73**	(0.34)	1.16**	(0.35)	1.04**	(0.33)
Age on Entry			.22**	(0.08)	.20*	(0.07)	.22**	(0.07)
Gender			1.09*	(0.51)	.52	(0.50)	.58	(0.49)
Physical Problems			.12	(0.66)	.47	(0.64)	.45	(0.67)
Extended Primary Schooling					-4.16**	(1.04)	-4.16**	(1.05)
Learning Assistance					-5.64	(1.05)	-4.67**	(1.05)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.32**	(0.06)
On K-Screen/Ach Relationship: Attrition							3.65*	(1.50)
Estimates of Parameter Variance								
	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	2.41**		2.26**		1.62		.25	
K-Screen/Ach Relationship	.16		.16		.45		.10	
Extended Primary Schooling					3.52		.50	
Learning Assistance					4.39		4.69	
Model Statistics:								
Maximum likelihood estimate of σ^2	61.79		61.24		56.24		55.98	
R ² : Percent of Total Variance Explained:	7.47		8.31		15.80		16.17	
R ² : Percent of Residual Parameter Variance Explained								
On Adjusted Levels of Achievement							84.57	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 18

HLM Models Explaining Variation in MS/Grade 3 Language Relationships								
	Model I		Model II		Model III		Model IV	
Fixed Effects	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	42.62**	(0.56)	42.71**	(0.56)	44.07**	(0.55)	41.29**	(0.70)
K-Screen Slope	2.57**	(0.33)	2.29**	(0.32)	1.52**	(0.31)	1.56**	(0.31)
Age on Entry			.24**	(0.08)	.20*	(0.07)	.20*	(0.07)
Gender			2.94**	(0.49)	2.41**	(0.46)	2.32**	(0.46)
Physical Problems			-.67	(0.65)	-.38	(0.61)	-.53	(0.69)
Extended Primary Schooling					-6.74**	(1.13)	-6.65**	(1.13)
Learning Assistance					-6.01**	(1.10)	-5.48**	(1.05)
Effects of Between-School Variables:								
On Achievement: School Mean Ability							.46**	(0.09)
On K-Screen/Ach Relationship: Attrition							1.05	(1.47)
<hr/>								
Estimates of Parameter Variance	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	6.63**		6.52**		6.23**		2.61**	
K-Screen/Ach Relationship	.13		.10		.08		.10	
Extended Primary Schooling					9.96		10.32*	
Learning Assistance					7.14		5.93	
<hr/>								
Model Statistics:								
Maximum likelihood estimate of σ^2	58.93		56.41		48.43		48.06	
R ² : Percent of Total Variance Explained:	14.99		18.62		30.13		32.75	
R ² : Percent of Residual Parameter Variance Explained								
On Adjusted Levels of Achievement							58.11	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 19

HLM Results for Grade Three Achievement on Kindergarten Screening Measures

Fixed Effects	Model V							
	Reading		Mathematics		Vocabulary		Language	
	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	34.56**	(0.89)	36.58**	(0.88)	34.81**	(0.81)	36.98**	(0.96)
DAP	.30**	(0.07)	.09	(0.05)	.28**	(0.06)	.24**	(0.06)
KLST	.44**	(0.09)	.34**	(0.09)	.55**	(0.09)	.35**	(0.09)
DEVTOT	.23**	(0.06)	.18**	(0.05)	.19**	(0.05)	.27**	(0.06)
MS	.26	(0.34)	1.10**	(0.32)	.35	(0.33)	.98**	(0.31)
Effects of Between Students Covariates:								
Age on Entry	.01	(0.07)	.10	(0.07)	.08	(0.07)	.09	(0.07)
Gender	.46	(0.51)	-.86	(0.43)	-.71	(0.49)	1.27*	(0.46)
Physical Problems	1.05	(0.62)	.20	(0.53)	.25	(0.58)	-.46	(0.57)
Extended Primary Schooling	-1.59	(1.29)	-2.21*	(0.99)	-.79	(1.05)	-3.59**	(1.14)
Learning Assistance	-4.51**	(1.24)	-4.40**	(0.90)	-4.04**	(1.06)	-4.86**	(1.02)
Effects of Between-School Variables:								
On Achievement:								
School Mean Ability	.34**	(0.80)	.33**	(0.07)	.30**	(0.07)	.42**	(0.08)
On K-Screen: Attrition								
DAP	-.07	(0.31)	-.02	(0.23)	-.03	(0.26)	-.18	(0.26)
KLST	.36	(0.38)	.11	(0.32)	.52	(0.35)	.11	(0.38)
DEVTOT	-.25	(0.22)	-.10	(0.19)	-.39	(0.20)	.13	(0.22)
MS	-2.41	(1.69)	-1.47	(1.57)	-3.15	(1.65)	.96	(1.55)
Estimates of Parameter Variance								
	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	4.86*		9.52**		4.37		10.17**	
DAP	.04		.01		.02		.02	
KLST	.04		.07		.03		.05	
DEVTOT	.01		.01		.00		.04	
MS	.21		.47		.25		.16	
Extended Primary Schooling	13.99*		5.39		3.79		9.59	
Learning Assistance	13.15*		2.57		7.65*		5.66	
Model Statistics:								
Maximum likelihood estimate of σ^2	52.88		40.00		49.24		42.94	
R ² : Percent of Total Pupil Level Variance Explained:	28.59		26.95		26.26		38.06	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 20

HLM Models Explaining Variation in Draw-A-Person/Grade 3 Reading Relationships

Fixed Effects	Reading		Model VI Mathematics		Vocabulary		Language	
	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Average within-school equation:								
Intercept	34.58**	(0.88)	36.91**	(0.82)	34.71**	(0.72)	36.97**	(0.96)
DAP	.32**	(0.06)	.29**	(0.06)	.25*	(0.06)		
KLST	.46**	(0.09)	.34**	(0.09)	.57**	(0.08)	.36**	(0.09)
DEVTOT	.23**	(0.06)	.18**	(0.05)	.20**	(0.05)	.27**	(0.06)
MS	1.28**	(0.30)	.97**	(0.30)				
Effects of Between Students Covariates:								
Extended Primary Schooling	-1.67	(1.28)	-2.42*	(0.97)	-3.61**	(1.14)		
Learning Assistance	-4.95**	(1.20)	-4.35**	(0.87)	-4.21**	(0.99)	-4.87**	(1.01)
Effects of Between-School Variables:								
On Achievement:								
School Mean Ability	.35**	(0.08)	.34**	(0.07)	.30**	(0.06)	.39**	(0.08)
Random Effects:								
	Estimate		Estimate		Estimate		Estimate	
Residual Parameter Variance:								
Achievement	5.13		8.28**		2.91		10.41**	
DAP	.03				.02		.02	
KLST	.03		.06		.01		.05	
DEVTOT	.02		.01		.01		.04	
MS			.32				.19	
Extended Primary Schooling	13.11*		4.34				10.12	
Learning Assistance	10.91*		2.02		5.80		5.75	
Model Statistics:								
Maximum likelihood estimate of σ^2	52.92		40.37		50.03		42.82	
R ² : Percent of Total Pupil Level Variance Explained:	28.53		26.28		25.84		38.23	

Note: * Significant at the .05 level.

**Significant at the .01 level.

Appendix Table 21

Means and Standard Deviations of Outcome Measures For Four Samples

Outcome Measure	Mean	Std.Dev	Number
District			
Read3	40.39	8.981	2193
Math3	40.55	7.692	2175
Vocab3	40.58	8.355	2180
Lang3	42.69	8.717	2188
Cohort1			
Read3	41.44	8.650	708
Math3	41.29	7.575	705
Cohort2			
Read3	40.84	8.856	745
Math3	41.19	7.485	740
Achieved Sample			
Read3	41.45	8.605	957
Math3	41.67	7.400	957
Vocab3	41.52	8.172	957
Lang3	43.89	8.326	957

