1983

Evaluation of the predictive ability of four screening measures used in pre-kindergarten screening: A follow-up study

Monica Marie Reichmuth
University of Northern Iowa

EVALUATION OF THE PREDICTIVE ABILITY OF FOUR SCREENING MEASURES USED IN PREKINDERGARTEN SCREENING: A FOLLOW-UP STUDY

An Abstract of a Thesis

Submitted

In Partial Fulfillment

of the Requirements for the Degree

Specialist in Education

Monica Marie Reichmuth

University of Northern Iowa

July 1983

This is to certify that

Monica Reichmuth

✓ satisfactorily completed the comprehensive oral examination

___ did not satisfactorily complete the comprehensive oral examination

for the Specialist in Education degree with a major in Educational Psychology: School Psychology at the University of Northern Iowa at Cedar Falls on July 14, 1983.

Examining Committee

Barry J. Wilson

Chairperson

Gaile S. Cannella

Member

Donna Raschke

Member

Lawrence L. Kavich

Member

Lawrence L. Kavich

Transmitted by:

Lawrence L. Kavich, Head
Department of Educational
Psychology & Foundations

This Study by: Monica Marie Reichmuth

Entitled: EVALUATION OF THE PREDICTIVE ABILITY OF FOUR SCREENING MEASURES USED IN PREKINDERGARTEN SCREENING: A FOLLOW-UP STUDY

has been approved as meeting the thesis requirement for the Degree of Specialist in Education

Barry J. Wilson

7/14/83 — Date

Chairman, Thesis Committee

Gaile S. Cannella

7/14/83 — Date

Member, Thesis Committee

Donna Raschke

7/14/83 — Date

Member, Thesis Committee

Lawrence L. Kavich

7/14/83 — Date

Member, Thesis Committee

Date

Dean of the Graduate College

ABSTRACT

The purpose of this study was to evaluate the expectancies of a set of screening measures (Peabody Picture Vocabulary Test, Developmental Test of Visual-Motor Integration, Peotone Language Scale, and the Motor Activity Scale) which have been used to identify prekindergarten children as either "ready" or "at-risk" to begin kindergarten. These screening measures were selected on the basis of their ability to measure readiness skills which were considered related to success in kindergarten (i.e., language, vocabulary, visual-motor perception, auditory and visual memory, and fine and gross motor skills).

Subjects included 113 white middle class children from four parochial schools in a large-midwestern city. The study was somewhat limited in its generalizability due to the small homogeneous sample of students; neither minority groups nor a significant range of socioeconomic groups were represented. Also, the sample did not include public school students. Randomization was not possible since it was necessary to use intact school classes.

Data from the four screening tests were gathered in the Spring of 1980 and the two criterion measures (Gates Mac-Ginitie and Behavior Rating Profile - Teacher Rating Scale) were administered three years later.

The Pearson product-moment correlations and multiple regression analysis were used to determine the relationship between the screening tests and the criterion variables. In addition prediction-performance matrices were compiled to assist in analyzing the "hit-rate" for false positives, false negatives, valid positives, and valid negatives among the second graders. In addition a separate analysis of the first graders' (i.e., those who were delayed entrance into kindergarten or who repeated one year) reading achievement and behavioral functioning was completed.

Results showed that while there were statistically significant correlations ($p < .05$) between each of the predictor variables and reading achievement, they tended to be small ($r = .24$ to .34). There were no significant correlations between the predictor variables and behavioral functioning.

The overall "hit-rate" was 86% for reading achievement and 88% for behavioral functioning with a high percentage of false positives (62.5%) and false negatives (70% and 75%). The first grade reading scores placed 38% of the children below the 50 percentile rank on the Gates while only one of the 13 was rated as having a behavior problem.

In summary, the study concluded that this particular set of screening instruments does not have sufficient predictive validity to warrant its continuation in a prekindergarten screening program. Multiple correlation ($R = .48$) showed approximately 23% of the

variance in reading achievement was accounted for by four independent variables which leaves 77% of the variance unexplained. It is recommended that a search be made to include more highly predictive measures and that follow-up evaluations be continued to assure the correct identification of children before they enter kindergarten.

EVALUATION OF THE PREDICTIVE ABILITY OF FOUR SCREENING MEASURES USED IN PREKINDERGARTEN SCREENING: A FOLLOW-UP STUDY

A Thesis

Submitted

In Partial Fulfillment

of the Requirements for the Degree

Specialist in Education

Monica Marie Reichmuth

July 1983

ACKNOWLEDGEMENTS

I would like to express gratitude to the members of my thesis committee, Dr. Gaile Cannella, Dr. Larry Kavich, Dr. Donna Raschke, and especially to the chairman of my committee, Dr. Barry Wilson, for their time, expert advice, and continual encouragement in preparation of this thesis.

Also, I would like to express my special appreciation to the members of the Sisters of Mercy Community, especially those Sisters living at Mercy Fontenelle in Omaha, who offered their continual prayers and loving support making the accomplishment of this thesis possible.

Finally, I want to thank the first and second grade teachers and the students who participated in this study. Without their assistance and the cooperation of their principals, this study would not have been completed.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

THE PROBLEM

Introduction

One of today's major social problems is the vast number of children who, as a result of some learning problems, are unable to realize their intellectual and educational potential. Children entering kindergarten frequently display a variety of handicaps which directly affect their capacity to learn. As a result, kindergarten screening of children has become a routine practice in many school districts today.

The U. S. Office of Special Education and Rehabilitative Services (1979) concurs with most special educators today in estimating that about 8%, which is between 3 and 4 million handicapped school-age children and youth, are eligible for special education funding. It is generally accepted that some kind of early identification is needed if these large numbers of children are to be helped effectively. Support for the importance of early identification has come from medical, psychological, and educational professionals as well as from parents (Keogh & Becker, 1973). However, the greatest impetus to the development of early screening programs has come as a result of federal legislation (PL 94-142, 1975) which has mandated early assessment and identification of children suspected of having an exceptional education need. Serious concerns about errors in the course of the identification process is

a major issue regarding kindergarten screening. There remains a great need for reliable, valid, and economical preschool screening measures to identify the psycho-educational problems as mandated by Public Law 94-142.

Many educational researchers have attempted to identify efficient measures of early identification of potential learning problems (Anderson, 1970; Cowgill, Friedland, & Shapiro, 1973; de Hirsch, Jansky, & Langford, 1966; Feshback, Adelman, & Fuller, 1974; Glazzard, 1977; Keogh & Smith, 1970; Koppitz, Sullivan, Blyth, & Shelton, 1959). The most frequently used procedures to identify prekindergartners who may have learning and/or behavior problems have been test batteries (e.g., Hainsworth Preschool Screening System or the de Hirsch Predictive Index), single instruments (e.g., Developmental Test of Visual-Motor Integration and Peabody Picture Vocabulary Test), and teacher-completed rating scales (e.g., Pupil Rating Scale or Croydon Checklist).

Teacher-completed rating scales are a more recent development. Investigators (Camp & Zimet, 1974; Cowgill et al., 1973; Hammill, 1971; Stevenson, Parker, & Wilkinson, 1976) have studied the relationship between teacher identification and future academic achievement in school. In addition, several long-range studies (Cowgill et al., 1973; Lilenfield, 1975; Meyers, Attwell, & Orpet, 1968; Ryan & French, 1974) have evaluated the results of early prediction over several years.

Most of the above researchers concluded that kindergarten screening programs on the whole are not without fault. A number of important issues are inherent in providing screening services for prekindergartners. The potential dangers of early identification are a concern to persons involved in the development and administration of early screening programs. For example, there is the possibility of identifying a problem when it may not exist. Since children do not mature at the same rate, readiness for school often is a matter of timing. Some children do not have developmental lags that may disappear by the time they are ready for formal schooling. Even for those children without developmental lags, test instruments when used with young children, are frequently inaccurate. Another concern is the possibility that early identification may impose limits on teacher expectancies. The term self-fulfilling prophecy has been used by Rosenthal and Jacobson (1968) to describe the effects of teacher expectancy on pupil performance. Many new instruments are evolving and as Wendt (1978) cautioned, they are adopted for use without consideration of philosophical issues *underlying assessment and are assumed to be appropriate without* establishing their predictive merits within each school system.

Researchers such as Hunt and Kirk (1974) and Lewis (1980) stress the need for screening programs to match the particular needs of certain schools. Practices vary so widely between schools that it soon becomes obvious that systematic screening programs

need to be devised by local interdisciplinary teams for their own schools. Consultation with local special services would enable the team to understand the unique requirements of each school and hence establish useful tests at this level. Similarly, the use of cut off scores should be determined by each school as the staff from each particular school is in the best position to know what entry skills are required to meet the demands of a relevant teacher or classroom. Each school should check the degree to which their predictions are accurate. Validation should be a continual process within each school setting.

In spite of the problems associated with preschool screening, it appears that screening programs have become an established educational practice across the country. Very few authorities question the need to identify children who are "at-risk." They do have serious concerns about how the results of the screening measures are used to categorize and place young children.

In summary, it can be said that a thorough understanding of developmental psychology and the kindergarten educational philosophy and programs in the local district is necessary. The various issues involved such as, test validity/reliability, false labeling, timing of assessment, and parental involvement must all be considered when electing to use a particular type of screening program. As for screening itself, it should be viewed only as a preliminary process to be followed by more intensive assessments

(Gallerani, O'Regan, & Reinherz, 1982) which might include the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) or the McCarthy Scales of Children's Ability (McCarthy).

Statement of the Problem

The purpose of this study was to evaluate the kindergarten screening program used in four elementary schools in a large metropolitan, mid-western community. In this evaluation an attempt was made to determine how well the screening tests were able to predict later reading achievement and behavioral functioning. From this information a prediction-performance comparison was compiled to evaluate the overall hit rate and to determine the percentage of false negatives, false positives, valid negatives, and valid positives of the children involved in the screening program.

Hypotheses

1. There is no significant relationship between prekindergarten measures and reading achievement three years later.

2. There is no significant relationship between prekindergarten measures and behavioral functioning three years later.

Additional Research Questions

1. Of the children identified as "at-risk," what percent lag in reading achievement (below the 50th percentile rank on the Gates-MacGinitie) and what percent are achieving adquately (at or above the 50th percentile rank according to the Gates-MacGinitie)?

2. Of the children identified as "ready," what percent are achieving adequately (at or above the 50th percentile rank according to the Gates-MacGinitie) and what percent are having reading difficulty (below the 50th percentile rank on the Gates)?

3. Of the children identified as "at-risk" what percent show a signfiicant number of behavioral problems as measured by the Teacher Rating Form of the Behavior Rating Profile (BRP) and what percent fall within the "normal range" or above as measured by the Teacher Rating Form of the BRP?

4. Of the children identified as "ready," what percent show a significant number of behavior problems and what percent fall within the normal range or above as measured by the Teacher Rating Form of the BRP?

5. How are the children who were delayed a year for kindergarten entry and who were in first grade at the time of the study doing in reading achievement?

6. How are the children who were in first grade at the time of the study doing in behavioral functioning?

Significance of the Study

This study was particularly relevant to the four stated schools who had adopted the screening program in the Spring of 1980 as no follow-up evaluations have been conducted to determine the effectiveness of the screening program. Further use of the screening instruments will depend upon the results of this study.

In addition, this study attempted to provide more evidence regarding the effectiveness of kindergarten screening for those entrusted with the design and implementation of screening programs in their school systems. Continued research and evaluation of screening programs is critical so that the most useful types of screening information can be gathered in a short period of time and misclassification of children can be prevented.

Limitations of the Study

This study is limited in its generalizability due to the relatively small sample of students; neither minority groups nor a significant range of socioeconomic groups are represented. Also, the sample does not include public school students. Randomization was not possible since it was necessary to use intact school classes.

While three of the screening instruments (i.e., Peabody Picture Vocabulary Test, PPVT; Developmental Test of Visual Motor, VMI; and Preschool Language Scale, PLS) are considered to be reliable and valid, the Motor Activity Scale (MAS) has not been standardized or normed; thus cannot be considered either valid or reliable as a screening instrument.

No attempt was made to control or account for a variety of factors which may have contributed to academic success or failure over the three years (e.g., parental divorce, loss of a significant family member, teacher effectiveness, parents' knowledge of the

screening results, student motivation, health factors, and curriculum adjustments).

Definition of Terms

1. "At-Risk": In this study "at-risk" referred to those students who were judged as likely to have later school problems in later grades, even though they had not yet been exposed to a formal school program (Keogh, 1977).

2. Behavioral Adjustment: In this study behavioral adjustment referred to those behaviors which fell within the "normal range" or above as measured by the Behavior Rating Profile (Teacher Rating Form). Unacceptable behavior was that which fell three or more scaled scores below the mean as measured by the BRP.

3. False Negatives: False negatives reflected those individuals who were identified as "ready" but who performed poorly on the criterion measures (Mercer, Algozzine, & Trifiletti, 1979).

4. False Positives: False positives represented those children who were identified as "at-risk" but who performed well on the criterion measure (Mercer et al., 1979).

5. Hit Rate: In this study, hit rate referred to the total number of correctly identified children (i.e., overall hit rate - 86 out of the 100 or 86%).

6. Readiness: For purpose of this study, readiness to begin kindergarten involved the development of language skills of

listening and speaking, of motor development, of auditory and visual discrimination, of concept and cognitive thinking, and of the ability to attend to and concentrate.

7. Reading Achievement: Reading achievement referred to a student's performance on the Gates-MacGinitie. In this study, unsuccessful reading achievement was a student's performance below the 50th percentile rank while successful reading achievement referred to performance at or above the 50th percentile rank.

8. Screening: Screening was a technique used for acquiring information about a large number of children in a short amount of time. The purpose was to identify those children who might have the characteristics of "at-risk" learners.

9. Valid Negatives: Valid negatives reflected those children who were identified as "ready" but who performed well on the criterion measures (Mercer et al., 1979).

10. Valid Positives: Valid positives reflected those children who were identified as "at-risk" but who performed poorly on the criterion measures (Mercer et al., 1979).

Summary

There has been extensive research over the years to evaluate the effectiveness of kindergarten screening programs. This study evaluated the predictive validity of a screening program used in four elementary schools by looking at the correlations between

early screening measures and reading achievement and behavioral functioning three years later. The information gained from the study gave the related schools evidence needed in order to evaluate the effectiveness of their particular screening program. The study is limited in that the population was small (N = 113) and not entirely representative of the larger population.

In Chapter 2 related literature will be reviewed that deals with instruments used in prediction, relevant prediction studies, and issues that are related to kindergarten screening programs.

CHAPTER 2

REVIEW OF THE LITERATURE

The assessment of school readiness and prediction of grade school achievement have been educational issues that have concerned classroom teachers as well as educational researchers. Within the last decade, early identification has received considerable support from parents and professionals from a variety of disciplines. The emphasis on early identification has been given additional impetus by Public Law 94-142 which requires assessments of health, vision, hearing, language and motor functioning, and social and emotional functioning on all children prior to their entrance to school (Abeson & Zettel, 1977).

The need to identify children who might develop learning problems seems apparent as it allows remedial intervention to take place before the children experience failure. This assumption has resulted in an investigation of 4 and 5 year-old children's psychological development. However, as Lindsay and Wedell (1982) point out, ". . . as the focus shifts down the age range, the type of process or ability investigated becomes more remote from the target task (e.g., reading)" (p. 212).

This review of the literature includes a summary of (a) the various types of screening instruments commonly in use; (b) the advantages and disadvantages of kindergarten screening; (c) five

unfavorable studies; (d) five favorable studies; and (e) the problems associated with the reporting of early screening research studies.

Types of Early Screening Measures

Prekindergarten screening practices throughout the nation tend to be quite diverse with varying degrees of effectiveness. Some of the major types of prekindergarten screening in current use have been reviewed and are summarized here. There are a wide variety of early measures which, according to recent studies, tend to fall into the categories of (a) perceptual-motor tests; (b) school readiness tests; (c) language tests; and (d) teacher rating scales.

Perceptual-motor tests

There is extensive evidence supporting relationships between visual-motor ability and educational achievement in primary grades (Bryan, 1964; Koppitz, 1964; Smith & Keogh, 1962) as well as in the upper elementary school grades (Keogh & Smith, 1967). Results, however, are still unclear as to the predictive accuracy of these measures (Keogh & Smith, 1970). Some investigators suggest that visual-motor tasks are more accurate in identifying high potential rather than high risk children (Keogh, 1965; Koppitz, 1964). De Hirsch, Jansky, and Langford (1966) found that one of the most predictive measures of later reading disability was perceptual-motor integration. Similar findings were reported in a three-year

follow-up of disadvantaged first graders (Weiner & Wepman, 1971). These authors found that measures of perceptual-motor functioning during first grade were extremely sensitive predictors of school achievement (largely language) at the end of the third grade.

Perceptual handicaps are frequently cited as an index of minimal brain dysfunction. The Bender Gestalt test, according to Koppitz (1964) is useful for the identification of both immature and bright youngsters. However, Keogh and Smith (1961) reported that children who performed well on the Bender in kindergarten tended to be good school performers but that poor Bender performance was non-predictive. The Developmental Test of Visual Motor Integration (VMI) is cited frequently in the literature as a useful measure in predicting visual-motor integration achievement (Flynn & Flynn, 1978; Haring & Ridgway, 1967). In addition, the Copy Forms Tests (Ernhart, Spaner, & Jordon, 1977) is sometimes used as an early screening measure. It was developed as an instrument for research on minimal brain damage and is, therefore, primarily perceptual-motor (Graham, Ernhart, Thurston, & Craft, 1962). The Hayes Early Identification Listening Response Test was utilized in a study by Buttram, Covert, and Hayes (1976). It was developed specifically for kindergarten pupils and emphasized not only the areas of visual perceptual and and fine motor skills but listening comprehension as well. Lindquist (1982) found that the Denver Developmental Screening Test (DDST) had limited value as a screening tool. However, the

Fine-Motor-Adaptive section of the DDST achieved almost as high a correlation as the test did as a whole.

School readiness tests

Many studies included the use of readiness tests in their early screening measures. The Metropolitan Readiness Test is, perhaps, used with the most frequency. Mitchell (1962) and Robinson (1966) reported that the Metropolitan is a valid and reliable instrument although its administration can be time consuming. The Metropolitan yielded favorable percentages in correctly identifying later reading achievement. In comparing it to the Lee-Clark Readiness Test (LC) and the California Test of Mental Maturity (CTMM), Lessler and Bridges (1973) concluded that the Metropolitan is the best predictor of the three. Rubin, Barlow, Dorle, and Rosen (1978) used the Metropolitan to predict later school achievement using the Stanford Achievement Test as a criterion measure. Their findings indicated correlations ranging from .50 to .75. However, the authors concluded that the preschool test performance did not provide sufficient information on which to make decisions regarding classification of individual children into high-risk groups. According to Rubin et al. (1978), the results of the Metropolitan suggest that when utilizing the total percentage test score the only effective predictor for screening potential problems in reading at the first-grade level is a score which falls below the 30th percentile.

Such results are dissimilar from the findings of Landsman and Dillard (1967) who reported that the Metropolitan did not predict serious learning problems.

Other readiness tests frequently cited in the literature are: the Stanford Early School Achievement Test (SESAT), found by Ames, Becker, and Dalton (1977) to show a high level of reliability; the Screening Test of Academic Readiness (STAR) (Ahr, 1967) which overlaps to some extent with the more established readiness tests, but is more comprehensive and easier to administer (Rogolsky, 1968-69); and the Sheppard School Entry Screening Test (SSEST) which Butler, Marsh, Sheppard, and Sheppard (1982) found to be significantly correlated with reading achievement in grades 1, 2, and 3. There are other readiness tests which have been viewed less favorably. For example, the Slingerland Screening Tests have been criticized by Keogh and Becker (1973) for their lack of validation. Similarly, McCandless (1972) says of the Boehm Test of Basic Concepts that none of the criteria of formal test construction was satisfied.

The work of Landsman and Dillard (1967), Wolff and Stein (1967), and Silverberg, Iversen, and Silverberg (1968) raised some doubt as to the efficacy of reading readiness tests in predicting later reading achievement. Silverberg gave the Gates Reading Test to kindergarten children and compared these scores with reading test scores achieved at the end of first grade. He found that the

Letters and Numbers subtest of the Gates administered in kindergarten, predicted end of first grade reading achievement as well as or better than the complete Gates test.

Language tests

The third category of early screening measures frequently used in prekindergarten screening programs is language. Research by Ilg and Ames (1965), Jansky and de Hirsch (1973), and Chomsky (1976) give support to language ability as a predictor of later reading achievement. Haring and Ridgway (1967) also indicated that language factors are the major indices of learning abilities. Analysis of studies using language tests as single predictors indicates that educational intervention can greatly influence achievement. Lyons and Bangs (1972) used the Language and Learning Assessment for Training Test (LLAT) to predict reading and mathematics achievement with and without intervention. The overall hit rate of the LLAT declined when children received intervention (e.g., the predictive outcomes were influenced by educational programming).

Hinton and Knights (1971) found that the highest correlations occurred on the tests of language skills such as the WISC, WRAT, PPVT, and Halstead-Wepman Aphasia Screening Test which included a variety of language skills. These findings indicated that school progress is best predicted by scores on tests which include language abilities including verbal IQ.

The PPVT is used frequently in preschool screening programs. Ernhart et al. (1971) found that it has both good concurrent and predictive validity because it is associated with those demographic variables (SES, Father's Occupation) which are usually related to intelligence test scores. The Quick Test (Ammons & Ammons, 1962), like the PPVT is a vocabulary recognition test with increasingly difficult stimulus words. There are indications that it is less valid than the PPVT for young kindergarten children (Ernhart et al., 1971). Durrett and Henman (1972) reported that the concurrent validity coefficients between PPVT and the Stanford Binet for preschool children was .52 ($p < .01$).

Teacher rating scales

In addition to perceptual-motor, readiness, and language tests, teachers' ratings is emerging as one of the most promising methods in the early identification of children educationally at risk (Evans, 1976; Feshbach et al., 1974; Glazzard, 1977). The use of a teacher checklist permits a range of social, behavioral, cognitive, and motor factors to be readily combined.

Kottmeyer (1947) suggested that teachers are able to predict reading problems better than the Metropolitan Readiness Test. He also reported that as years of experience increase, so does accuracy of prediction. However, a study by Kapelis (1975) does not support this purported relationship. Haring and Ridgway

(1967), in a screening of 1,200 kindergarten children, stated that teacher perceptions are accurate predictors of future school-related problems. An analysis of errors within a study by Keogh and Smith (1970) indicated that only false positive mistakes were made.

In a report by Keogh, Welles, and Hall (1976), teachers' ratings were considered feasible, valid, and economic "first screens" for recognition of children at risk or of high potential in school programs. However, these investigators suggest that the teachers' ratings are frequently summarizing scores and, therefore, should be followed up by a more differentiated educational assessment.

Several long-range prediction studies (Glazzard, 1979; Lilenfield, 1975; Meyers, Attwell, & Orpet, 1968; Ryan & French, 1974) have evaluated the results of early prediction three to six years later. Lilenfield (1975) found that kindergarten teachers' judgment correctly predicted academic outcomes six years later. Lloyd (1978) analyzed background characteristics, school performance, and achievement test data of 788 third-grade boys and 774 third-grade girls who were known later to have become high school dropouts or graduates. A discriminant function yielded a 75% overall correct classification of both groups of students. Six or seven of every ten high school failures had been correctly classified by characteristics exhibited in the third grade.

The Kirk Teacher Rating Scale frequently was cited as one of the more predictive teacher rating measures. The results of a long-range study done by Glazzard (1982) indicated that the Kirk was a more predictive measure than reading readiness test scores. Another scale, the Schenectady Kindergarten Rating Scales (SKRS), was evaluated by Tobiessen, Duckworth, and Conrad (1971) to evaluate the predictive efficacy of the SKRS profiles. The results indicated that deviant scores on scales involving impulse control and scales involving language and perceptual-motor skills predicted similar problems in first grade with considerable accuracy. Profiles reflecting withdrawn or hostile kindergarten behavior were less accurate predictors of similar behavior in first grade.

The teachers in a study researched by Ann Lewis (1980) tended to be overly generous in their ratings toward quiet girls and to penalize noisy boys and children they considered to be from problem families. This tendency has been noted elsewhere (Keogh, 1977; Lindsay, 1978). Lewis concluded that her findings suggested the need for training in the use of teacher rating scales so as to prevent biased ratings.

It may be concluded from the studies reviewed above that there is consensus that teachers represent a useful first level screen in the identification of educationally high-risk children.

These findings have particular relevance for school psychologists involved in the assessment and diagnosis of children for possible special education programming. Keogh, Tchir, and Windeguth-Behn (1974) suggest that ". . . such data are often more useful in remedial planning than are psychometrically derived scores which serve to classify, not explain, school failure" (p. 49).

Advantages and Disadvantages of Early Screening

While there is general agreement for the need to identify educationally high-risk children, differences arise regarding techniques which provide accurate and valid early prediction of school achievement.

Proponents of early identification of children at risk support the practice and suggest that it has many advantages. Many educators believe that if teachers can identify young children who are at risk for subsequent academic achievement, appropriate instructional modifications can be implemented. They realize that early intervention with an instructional program designed to remediate deficit skill areas is a more efficient and economical use of a teacher's and a child's time than a program of remediation later. The efforts of theorists such as Frostig, Kephart, Kirk, and others have greatly increased understanding of the development of visual, auditory, and motor functioning in the young. They have proposed that the behavior

of young children can be changed more easily than that of older children and, therefore, early intervention can prevent the trauma of school failure, especially by concentrating on the assessment and remediation of specific disabilities.

Personality theorists, also, have been supportive of early identification practices and have indicated that personality traits are established during the preschool years (Hayden, 1974). Keogh and Becker (1973) point out that the importance of early identification comes from the disease model where recognition or diagnosis is followed by a prescription for treatment. Dietary treatment of phynylketonuria (PKU), for example, confirms the need for early diagnosis and treatment.

All of the arguments advanced above stress that the earlier the screening and subsequent intervention, the greater the likelihood of success. However, there are problems in regard to early identification. As Keogh and Becker (1973) cautioned, we are only hypothesizing when we attempt to identify preschool or kindergarten children whom we fear may become learning failures. They have raised questions concerning the relevancy of any identification procedures depending upon criteria from outside the actual school environment. Their questions serve as cautions in the development of identification instruments: "(a) How valid are the identifying or predictive measures?, (b) What are the implications of diagnostic

data for remediation or educational intervention?, and (c) Do benefits of early identification outweigh possible damaging or negative efforts of such recognition?" (p. 6).

The problem of misdiagnosis is seen by many as the most distinct disadvantage of early identification practices. Educators who classify children as at risk or not at risk based on the child's performance on preschool measures could be making inaccurate predictions. According to Rubin et al. (1978), "There is a wide discrepancy between current interest in predicting children at high risk for future failure and our capacity to select such children with a reasonable level of confidence" (p. 64).

Another problem with early identification is that young five-year-olds are developing and changing skills at a fast pace, so much so, that predicting school achievement in preschoolers is difficult at best. The problem of predicting later achievement from kindergarten data may lie in the child, not in the measuring instruments (Flynn & Flynn, 1978). Thus, the optimum time for screening is still an area of debate. Haring and Ridgway (1967) have indicated, a failure to progress satisfactorily in learning may be as much the fault of the learning environment as a function of the organism. In addition personal and environmental factors and the interaction between the two can also shape the level of attainment achieved by the child in school. Wedell (1980) noted

that individuals are often able to compensate because of unusual skills, motivation, and processing abilities.

Of all of the cautions put forth regarding the disadvantages of early identification, perhaps the biggest concern is that the instruments used in the screening programs are not always valid or reliable. Many have only a small amount of information available on which they can be evaluated. This is the case, not only with the new and experimental instruments but also with those which have been used for ten years or more. Even the best show predictive validities in the .6 to .7 range (Lindsay and Wedell, 1982). The usual method for evaluating predictive validity is by use of the correlation coefficient. But as Carver (1978) argues, statistical significance is not a sufficient criterion as there may still be a large number of individual children misclassified.

As mentioned above, many studies have been done to determine the effectiveness of screening programs and that the conclusions drawn from these studies are frequently contradictory. Some studies have indicated that early screening programs have been successful in identifying high risk children while other studies report less favorable results.

Unfavorable Studies

The first study reviewed here, whose results were not extremely favorable toward early screening, was a follow-up study

of 163 children who were given a screening test in kindergarten and then followed to the end of grade 2. Eaves, Kendall, and Crichton (1974) began their study of kindergarten children (N = 228) in 1969 in an attempt to identify early in their school career those children who might show signs of minimal brain dysfunction, or who might otherwise be expected to have difficulty in school. Their study was modeled after the original work of de Hirsch et al. (1966). The original sample was from Vancouver, a rather mobile city, which caused the attrition to be high (29%). The predictor variables included a Modified Predictive Index (MPI) based on de Hirsch et al. (1966); a Teacher's Check List (TCL) to indicate the child's readiness for school, and in the case of a subgroup of 50 intensively examined children, an extensive psychological and neurological examination. These 50 children were classified as having minimal brain dysfunction (MBD) or as being immature or normal. After one to two years of school, further data from TCL's, Stanford Achievement Tests (SAT), Cooperative Primary Tests, and information obtained from parents were available for statistical examination.

Statistical analysis was done in four ways. First, correlation coefficients for certain pairs of variables were computed. Second, a step-wise multiple regression technique was used to select a subtest from a group which would best predict a given variable.

Third, discriminant analysis was used to select a subtest of variables for their ability to discriminate between two or three groups. Using discriminant analysis allowed the authors to determine which kindergarten measures grouped together and best classified students correctly in terms of school achievement. Finally, analysis of variance was done to determine if the group means on the test given at the end of grades 1 and 2 were significantly different in MBD, immature and normal children.

The results of this study showed that if the kindergarten children are tested in June rather than October of their kindergarten year, a better prediction of how well they will perform on word analysis and listening tests can be made. Some predictive ability was lost when the composite score was used instead of the individual scores.

The MPI subtests given in June of the kindergarten year yielded the following multiple correlation coefficients: .86 with the grade 2 Word Analysis Test, .70 with Listening, and .55 with Reading. Those ready for grade 3 included only one of the MBD children, half of the immatures, and 86% of the normal children. Only 41% of the failures were ready for grade 3 as opposed to 91% of those who passed the tests.

The authors noted that by slightly raising the cutoff score for failing the MPI, an even larger number of children (nearly 60%)

were identified as not ready for grade 3. They did not need different cutoff scores for boys and girls separately. The authors suggested, as did de Hirsch (1966), that cutoff scores should be adapted to each school as socioeconomic factors and academic expectations are unique to each district.

Although Eaves et al. (1974) admitted that there are definite advantages which involve the future total well-being of the child, their general conclusion was that:

> We simply do not know enough about the psychological processes involved in learning to read and write to be able to explain our findings . . . in terms of specific relationships between earlier performance on motor, perceptual, and language tasks and later progress in reading skills (p. 637).

A second unfavorable study, done by Flynn and Flynn (1978) addressed the question: "Do the instruments studied predict the child's school adjustment after two years of school" (p. 65)? The authors felt the need to answer this question before launching into remediation procedures which are devised to avoid possible school difficulties as advocated by Satz and Friel (1974) and Haring and Ridgway (1967).

The screening instruments selected for this study were the (a) Slosson Intelligence Test (SIT); (b) Peabody Picture Vocabulary Test (PPVT); (c) Goodenough-Harris Drawing Test; (d) Developmental Test of Visual Motor Integration (VMI); and (e) Metropolitan Readiness Test (MRT). The dependent variable was the California Achievement Test (CAT), 1970 edition.

The 81 children administered both the screening measures and CAT were attending public school kindergarten in a university community of approximately 25,000 population. Of these children, 4.3% were black. The original sample consisted of 123 children in six kindergarten classes. The data were analyzed using stepwise regression, using the CAT subscores and total score. The intercorrelations indicated that only the MRT was significantly related to the CAT and subtest scores. However, even this relationship was not strong, as only 10% of the variance is accounted for in the dependent variable. The other multiple correlations between the predictor and criterion tests ranged from .27 to .30. The authors concluded that, on the basis of their results, screening test results for the most part have little relationship to the child's future school achievement. According to the authors these results may be partially explained by factors that may lie in the child rather than the instruments because children from five to six years of age are in a state of flux. They recommended that because of instability of the deficits within children at this age, perhaps the teacher would be better advised to use his/her own observations and interactions with the child to devise a treatment program.

A third study done by Rubin et al. (1978) also cautioned that educators who are currently classifying children as at risk based on the child's performance on preschool measures may be making

unwarranted assumptions regarding the validity of predictions for individual children.

The purpose of their study was, therefore, to determine the extent to which children identified on preschool readiness tests as "high risk" for development of learning difficulties do, in fact, demonstrate low achievement in basic school skills by the time they have completed several years of elementary school.

The 732 subjects in the Rubin et al. study were drawn from the 1,613 participants in the Educational Follow-Up Study (Rubin & Balow, 1977), a prospective longitudinal investigation of the educational and behavioral outcomes associated with perinatal and early childhood conditions and events.

The Metropolitan Readiness Tests (MRT) were administered to subjects at age 5 the summer prior to kindergarten entrance and again at age 6 prior to entering the first grade. The word meaning, spelling, and arithmetic computation sections of the Stanford Achievement Test (SAT) were individually administered during the summer of the calendar year in which subjects reached their ninth birthday. At the time of the SAT administration, 212 (29%) of the subjects had completed grade 2, and 520 (71%) had completed grade 3. Correlations between scores on the MRT and each of the three subtests of the SAT were first computed separately for the five-year and six-year MRT performances. The MRT scores from the

two administrations were then combined in a multiple regression prediction of the SAT subtest scores, the resulting multiple correlations being .71. For the second part of the analysis, subjects were classified into "Achieving" and "Low" groups on each of the MRT tests and on each of the three SAT subtests.

The results showed that scores on the pre-first MRT were more accurate achievement predictors, particularly of low performers, on all three subtests on the SAT than were scores on the prekindergarten MRT. One-half of the 92 subjects classified in the low groups on both the prekindergarten and pre-first grade MRT were later classified in low groups on the subtests of the SAT. The authors concluded that the findings of this study indicated that correlations in the .50s, .60s, and even as high as .70 between scores on predictor and outcome variables do not justify the assumption of consistency of performance over time for low-scoring children.

A fourth study yielding unfavorable results regarding the use of early screening measures is a study done by Lewis (1980). In an attempt to find a practical solution to a problem faced by teachers in an English Infant School (4 to 7 year olds), she found that a large number of children scored low on a standardized reading test when they completed their education at the Infants' School (age

7 plus). The teachers were looking for a method whereby these children could be identified and helped early in the Infants' School.

The total sample consisted of 162 children. Eighty-six of these children (44 boys, 42 girls) constituted the first year group and ranged in age from 5 years 1 month to 5 years 11 months when first screened. The second-year group was comprised of 75 children (37 boys, 38 girls) who ranged in age from 4 years 10 months to 5 years 8 months. This sample represented the total middle infant groups in two successive years at one school. Children who left or joined the groups during the two school years covered by this study were excluded from the final data.

The initial screening procedure consisted of the English Picture Vocabulary Test (EPVT), the English version of the PPVT. It was selected because Lewis felt that a measure of verbal ability would be an appropriate indicator of later reading attainment. A second predictor variable involved the Croydon Checklist, a teacher checklist developed in Croydon, England. This checklist was completed by the teachers for all children scoring below 90 on the EPVT and also on any other children about whom they felt concerned.

To assess the predictive ability of the screening procedure, the Young's Group Reading Test, a similar format as the Doren Diagnostic Reading Test, was administered to the whole of the first-year group.

Using stage one (EPVT) with stage two (Croydon) as the predictors and a cutoff of 84 on a standardized test (Young's) as the criterion, the probability of a positive diagnosis (those identified as "at risk") being correct was only .36. Therefore, the conclusion was that the use of a screening procedure did not improve hits beyond chance. The EPVT was a relatively poor predictor of reading attainment, giving 73% to 79% hits, despite the common linguistic bias that might have been expected to make a test of receptive vocabulary a good predictor of reading attainment.

The use of the teacher checklist did involve the teachers in the screening procedure. However, it was noted that the teachers in this study were inclined to be overly generous toward quiet girls who, on the basis of later specific ability and criterion-referenced tests, did need help. Also, those teachers tended to penalize noisy boys and children they considered to be from problem families. This tendency has been noted elsewhere (Keogh, 1977; Lindsay, 1978).

In this study a number of children, both positives and negatives, were misclassified, and the manipulation of cutoffs did little to improve the efficiency of the screening procedure. Changing the predictor (e.g., using the EPVT alone) or changing the criterion (Young's Group Reading Test score) did not eliminate

misclassifications. This suggested that individual children were moving from risk to not at risk and vice versa.

Lewis (1980) suggested that, predictions, particularly in relation to those who are failing, are likely to be weak, since they cannot take into account the range and interaction of compensatory factors available to the individual child. This suggests that a screening procedure should be seen in terms of identification initially and thereafter as a continuous process to incorporate the effects of change in the child, school, and family situations on the child's level of attainment.

The fifth unfavorable study reviewed was that done by Lindquist (1982) who also realized the importance of establishing the validity of measures for predicting later school achievement as the research that has been conducted thus far has resulted in conflicting findings. As a consequence, she conducted a study to determine if the Denver Developmental Screening Test (DDST) predicted the Gates-MacGinitie Reading Test scores of primary grade students.

The subjects consisted of primary grade students in the Weber School District in Utah. The sample was stratified to include approximately equal numbers of students from each of the first, second, and third grades. The sample was randomly selected from five of the elementary schools in the district. A total of 351 students were included in the study.

Results from the data showed a significant correlation between Gates-MacGinitie Reading Test scores and DDST total scores for all three grade levels. The relationship, however, was not strong enough to provide practical significance. The correlation was highest for first-grade students and lowest for those in third grade.

A total of 55 children were identified as being at risk for educational failure. Of these, 26 were found to be below the 25th percentile on the reading achievement tests. There were also 64 children that were not identified as at risk that were below the 25th percentile. Eighteen of the children identified at risk were achieving average scores on the reading achievement test. Of those students with scores above the 75th percentile, 11 children had been identified as at risk of educational failure and 78 had not.

Lindquist concluded that the DDST has very limited value as screening tool for identification of children with potential learning difficulties. In order to increase the predictive validity of the screening program, Lindquist suggested that better measures of language development than the items on the DDST should be included.

Favorable Studies

Among the more favorable studies reviewed is one done by Meyers, Attwell, and Orpet (1968). This study demonstrated the value of certain kindergarten examinations and test behavior ratings for anticipating fifth grade achievement, with emphasis upon efficient

selection of tests and ratings showing up best in stepwise regression procedures.

The 100 subjects used in their sample were white, mostly upper lower and lower middle class from the greater Los Angeles area. The subjects had originally been tested late in the kindergarten year and within three months of the sixth birthday. The fifth grade testing was done on 57 (25 boys and 32 girls) of the original sample. The children were close to ten and one half years of age at the time of the second testing. The achievement battery used was the California Achievement Test, and the intelligence test was the California Test of Mental Maturity. The criterion variables included achievement subtests, total achievement measures, and IQ scores. The predictor variables included 13 individually administered tests. There were three tests in each category of psychomotor, perceptual speed, linguistics, and figural reasoning categories, together with digit span (Binet style). In addition, the examiner completed a 9 point rating scale on 10 characteristics of test behavior on each child.

Results showed that of the 13 tests and 10 ratings across four and one half years, no single correlations exceeded .50 but stepwise multiple regression yielded Rs of .64 to .75 for the achievement subtests, .74 for total achievement tests, and .73 and .63 for the IQ's. The authors were impressed with the results of the behavior

ratings in anticipating later achievement. The best rating (attention) was as predictive as the most valid test. Thus, the authors concluded that the recommended minimal program for replicative or interventive effort can be reduced to one test, Picture Vocabulary, and to ratings of a few qualities to yield three scores: attention, energy, and performance rate.

A second study yielding favorable results regarding early identification of educationally high potential and high risk children was that of Keogh and Smith (1970). They followed the same children (N = 49) from kindergarten entrance through grade five of a regular school program. Subjects were enrolled in four schools in a Southern California public school district. The sample was predominantly white, middle class; no known mentally retarded were included.

Realizing that the predictive accuracy of visual-motor measures and teacher assessment for long-term prediction of high potential children had not been clear up to that point, Keogh and Smith's study considered the predictive accuracy of the Bender Gestalt and teachers' ratings. Kindergarten data included the group Bender Gestalt administered four times: at school entrance and at 2½ month intervals over the school year. Scores for the nine designs on the Bender were summed, range of possible total scores nine through 45.

In the spring of the year, kindergarten teachers rated subjects on a five point scale of reading readiness. Teachers were not provided with specific criteria on which to evaluate readiness; thus, the rating represented a global opinion by the teacher.

School achievement data, part of the district's regular testing program, included the Stanford Reading Test (SAT) at grade two and the California Achievement Test (CAT) at grades three through five.

Strength of relationship between kindergarten predictive measures and later school achievement was evaluated using the Pearson r. Teacher's ratings had consistently significant correlations with achievement measures. Relationships between the Bender at kindergarten and later school achievement were generally lower and for the most part nonsignificant, especially for girls. However, when teacher's ratings and kindergarten Bender scores were combined to predict third and fifth grade achievement, values of r were .67 and .69 between the predictive measures and reading scores for girls; comparable values of r were .61 and .71 for boys. Multiple Rs for the kindergarten measures and third and fifth grade arithmetic scores were .41 and .56 for girls, .71 and .76 for boys. Use of the multiple coefficient of correlation increased the strength of relationship between the predictive and criterion

measures, and the major contribution to the relationship came from teachers' ratings.

Results of the Keogh and Smith (1970) study suggest that performance on the Bender and teachers' ratings at kindergarten may be useful clues in initial screening of children who are possible school learning problems. Although the authors found the kindergarten teachers' ratings and later school achievement to have consistently high relationships, they also cautioned that more extensive follow-up and individual diagnosis should be done on those children identified as high risk. Their findings also suggest the need for more careful examination of the dimensions which teachers use to evaluate readiness.

A third favorable and early detection study which purported to identify the predictive antecedents of developmental dyslexia was completed by Satz and Friel (1974). The original sample consisted of 497 white male kindergarten pupils in the Alachua County, Florida, public school system and the University of Florida laboratory school. Data were collected on 86 school days between October and March. Subjects were tested individually. Four hundred and seventy-three subjects were available for retesting at the end of grade 1, 95.2% of the original sample. The mean age of the subjects included in the study was 66.3 months with 11.5% of the subjects from low socioeconomic status.

The predictor variables included the following: (a) Day of Testing; (b) Age; (c) Handedness; (d) Finger Tapping Difference; (e) Finger Tapping Total; (f) Peabody Picture Vocabulary Test, IQ Score; (g) Recognition-Discrimination Test; (h) Embedded Figures; (i) Verbal Fluency Test; (j) Developmental Test of Visual Motor Integration; (k) WISC Similarities Subtest; (l) Alphabet Recitation; (m) Right-Left Discrimination Test; (n) Finger Localization Test; (o) Auditory-Discrimination Test; (p) Dichotic Listening Test, Right Channel Recall; (q) Dichotic Listening Test, Left Channel Recall; (r) Dichotic Listening Test, Ear Symmetry; (s) Dichotic Listening Test, Total Recall; (t) Auditory-Visual Integration Test; (u) Behavioral Checklist; and (v) Socioeconomic Status.

The criterion measure in this study was based upon a 10-item scale of reading which was rated by the teacher at the end of first grade. The scale ranged from "No Readiness" to "Advanced Third Reader." Those whose reading level was rated as below Primer were assigned to the "High Risk" group (N = 73); those whose reading was judged to be at the Primer level or above comprised the "Low Risk" group (N = 400). The two criterion groups (High and Low Risk) were further subdivided into an extreme and conservative reading subgroups.

The results indicated substantial differences between the criterion groups. The Low Risk group consistently performed better than the High Risk group. Those variables which did not differentiate

criterion groups were: Age, Handedness, Finger Tapping Difference, Auditory-Visual, and the Dichotic Listening Difference (Ear Asymmetry). Some of the tests revealed striking mean differences between criterion groups. The Finger Localization and Recognition-Discrimination tests were much lower in the High Risk group (56.1% and 44%, respectively) than in the Low Risk group (77.8% and 67.3%, respectively). Similarly, on the Beery Developmental Test of Visual Motor Integration, the High Risk group had almost a year's lag between their chronological age and performance age. By contrast the Low Risk group matched their chronological age on this test. The authors noted that the mean Verbal IQ on the Peabody Picture Vocabulary Test was much lower in the High Risk (IQ = 90.9) than in the Low Risk group (IQ = 106.1).

A discriminate function analysis was performed on the 22 predictor variables and the two criterion groups (High and Low Risk). Results revealed that 57 of the High Risk group (N = 73) were classified correctly, yielding a valid positive rate of 78.1% and a false negative rate of 21.9%. Similarly, 342 of the Low Risk group (N = 400) were classified correctly, yielding a valid negative rate of 85.5% and a false positive rate of 14.5%. The total hit-rate was 84.4% (399/473). Accuracy was higher in the extreme categories of each criterion group. Virtually all of the misclassification errors were confined to the conservative categories.

The step-wise regression analysis used in this study showed that the Finger Localization Test revealed the highest discriminable ranking which accounted for 76.1% of the total correct prediction. Recognition-Discrimination, Day of Testing, and Alphabet Recitation ranked second, third and fourth, respectively. Inclusion of these variables along with the Finger Localization Test increased the hit-rate of 81.6%. The authors also decided to elevate the cutoff in order to reduce the likelihood of misclassifying a normal child as High Risk.

The four highest ranking variables all loaded heavily on the factor, sensory-perceptual-motor mnemonic ability. The authors note that this same factor emerged in a study done by Adkins, Holmes & Schnackenberg (1971) using the de Hirsch predictive battery. This factor was also one of the most predictive measures of later reading ability in the original de Hirsch, Jansky, and Langford longitudinal study (1966).

The authors concluded that the results of the step-wise regression analysis provided additional support for the theory that delays in early developing skills will forecast later handicaps in reading and that intervention techniques which focus on those behaviors which already depict the subsequent handicap will help to alter the incidence of prevalent childhood disability.

A fourth, comprehensive study yielding favorable outcomes regarding early screening was that done by M. Glazzard (1980). The

purpose of her study was to compare the predictive effectiveness of kindergarten teacher ratings with reading readiness and reading achievement tests over a six-year span using reading achievement at the completion of the first through the sixth grade as the criterion.

The original subjects in this study consisted of 107 kindergarten students, 50 boys and 57 girls, with a mean chronological age of 6.2 years. The following predictor variables were used: (a) The Gates-MacGinitie Reading Tests: Readiness Skills (1968), (b) The Gates-MacGinitie Reading Tests, Primary A, Form 1 (1969), and (c) The Teacher Estimate of Kindergarten Pupils' Abilities (Kirk, 1966).

The criterion variables were the vocabulary and comprehension test scores from the Gates-MacGinitie Reading tests obtained at the completion of each grade, first through sixth.

A multiple-regression analysis with an iteration sequence was chosen for the data analysis since this procedure selects the smallest test or item battery that shows the maximum predictive efficiency. A second step in the data analysis consisted of a pairwise analysis of covariance to determine the residual correlation (the uniqueness) of a predictor variable with the criterion of achievement after the effect of the other variables had been partialled out.

Glazzard reported that both the Gates Reading Readiness Test and the Kirk Teacher Rating Scale are significant predictors of vocabulary achievement in grades one through six. Results showed that the Kirk Teacher Rating Scale was a more efficient predictive measure than either the Gates Reading Readiness Test or the Gates Reading Test for first grade comprehension achievement. However, the Gates Reading Readiness Test was found to be a more efficient predictive measure for comprehension achievement in grades two and three. For a fourth, fifth, and sixth year prediction, both the Kirk Teacher Rating Scale and the Gates Reading Readiness Test were uniquely predictive of comprehension achievement. From the results of this long-range study it appears that the Kirk Teacher Rating Scale of academic achievement is a more expedient, predictive measure than reading readiness test scores for a one year prediction of reading achievement. If the focus is on intervention, the earlier prediction is more efficient than a long-range prediction of two or more years before remedial help is offered.

The authors concluded that when a one-year prediction is desirable for early intervention, the following subtests could be eliminated from the Gates-MacGinitie Reading Readiness Test: Visual-Motor Coordination, Auditory Discrimination, and Listening Comprehension. Similarly, the areas of Creativity, Perceptual

Discrimination, and Abstract Ideas could be eliminated from the Kirk Teacher Rating Scale.

A recent study done by Butler et al. (1982) also supports the need for establishing early screening programs. The purpose of this study was to determine how well the Sheppard School Entry Screening Test (SSEST), administered in kindergarten, predicted reading performance in grades 1, 2, and 3 (N = 320 pupils). The specific questions explored were: (a) Does the SSEST predict reading achievement beyond that which can be explained by IQ alone?, (b) Does the level of predictability in the three grades differ, and why?, (c) Does the impact of different predictor variables vary in different grades?, and (d) Does the predictability vary for students of different reading abilities?

Multiple correlations, based on the three SSEST factors (Figure Drawing, Language, Perceptual-Motor Skills) and two background variables (Pupils' sex and whether or not at least one of their parents spoke English), were .49, .56, and .61 for reading scores in Grades 1, 2, and 3 respectively. Partialling out the effect of IQ reduces the magnitude of the correlations between the SSEST factors and the reading scores, but each of the relationships is still substantial and statistically significant. The results also indicated that 24% (Grade 1), 31% (Grade 2), and 38% (Grade 3) of the variance in reading scores is predictable. The relationships were substantially stronger in second and third grade

than in first grade. Not all the SSEST factors were equally effective at forecasting reading performance. Language was consistently the best predictor. Figure Drawing, although contributing somewhat less than Language, consistently made a substantial contribution as well. The authors constructed an expectancy/classification table to demonstrate the adequacy of decisions based on the reading predictions. The conclusions were that predictions were superior for students with the poorest reading abilities. The bottom 10% of predicted scores classified children as at risk. Using this decision rule the study correctly classified 20 (of 48) poor readers and 260 (of 272) normal readers for an overall hit rate of 87.5% (280/320). Prediction was much better for pupils with the lowest reading ability. A path analysis indicated that the direct impact of the predictor variables was limited primarily to reading in the first two grades. The authors concluded that the findings of this study demonstrated the validity of the SSEST as an early indicator of later reading achievement.

Integration of Research Findings

Stepping back and looking at the studies reviewed above, it is possible to view the similarities among the various studies as well as to analyze some of the research problems associated with making predictions based on early screening measures.

In doing so a comparison of the various instruments and the problems associated with these instruments will be reviewed.

Secondly, the types of statistical analyses used and the difficulties associated with each will be summarized. Lastly, a listing of the critical elements to consider when studying the effectiveness of kindergarten screening will be stated.

First, it has been noted that there are a wide variety of screening instruments used in screening programs as well as a diverse selection of tests used as criteria for successful achievement at a later date. Some of the instruments employed have been used frequently and have demonstrated good reliability and validity but others fall well short of this measure. Many of the screening measures used were new and as a consequence have only a small amount of information available on which they can be properly evaluated. The majority of the studies reviewed did not include the evidence of reliability or validity. A few, such as the study by Flynn and Flynn (1978) did include brief summaries of each of the five measures used in their study thus making the results of their study more acceptable to the reader.

Among the predictor and criterion instruments used in the studies, single test, test batteries, and teacher ratings were the most common. Single instruments methods generally used one standardized test or one variable (e.g., Denver Developmental Screening Test utilized by Lindquist, 1982) as either the predictor or the criterion measure. When a battery of tests (or variable)

was used, the prediction was made from a pattern of scores or one global score. Eaves, Kendall, and Crichton (1974) for example, used the de Hirsch Modified Predictive Index to predict teacher recommended grade placement. A third type of prediction study used to establish prediction information involved obtaining teacher ratings of children through a checklist, rating scale, or questionnaire. Teachers' ratings tended to be most effective in identifying those children in need of intervention and those not likely to need special programming. Keogh and Smith (1970) suggested that teachers do recognize developmental and behavioral characteristics important in school performance. Glazzard (1980) reported that the predictive ability of teacher ratings was found to be quite high for prediction of reading achievement one, two, three, and four years later. She concluded that the excess of administration time would not be as expedient a prediction measure as having the kindergarten teacher rate the youngsters and thereby identify potentially "high-risk" students.

Statistical analysis of the results were usually done in two ways. First, correlation coefficients were computed to determine the extent of the relationship between them. Secondly, a predictive-performance matrix was compiled to demonstrate the overall hits and misses of the study.

The correlational format used in the evaluation of early identification projects involved administering a screening (i.e.,

prediction) instrument(s) in either kindergarten or during kindergarten and a criterion (i.e., performance) instrument(s) at a later date. Most of the studies reviewed here ranged from screening during the kindergarten year with a two to three year follow-up evaluation. Correlation scores between the predictor and criterion measures were then evaluated as evidence of program utility. Not all of the studies reviewed reported correlational data. For example, Satz and Friel (1974) and Lewis (1980) discussed their findings without ever actually reporting a correlational coefficient thus making their findings difficult to compare with other studies that only report correlation coefficients. Those that did, reported simple correlations (rs) averaging between .08-.71 and/or multiple correlations (Rs) between .28 and .86. Sometimes the correlations were between total test scores but more frequently between subtest scores of the predicted and predictor variables. Inconsistent reporting made it difficult for the reader to make comparisons across studies.

The second type of analysis of the data utilized in some of the studies reviewed was the prediction-performance matrix where some form of horizontal and/or vertical analysis was performed. In most of the studies, it was anticipated that the screening procedure would produce a high predictive validity. For those studies which quoted only the percentage of overall hits, the figure tended to produce a spuriously high figure as it was generally

easy to predict that children who are initially successful will remain successful. This group usually represents the large majority. With the studies which also reported the percentage misclassified, it was sometimes unclear whether the figure is a percentage of the criterion (how well they performed) or the prediction ("at risk" or "ready") (Lewis, 1980; Rubin et al., 1978). Satz and Friel (1974) used a composite figure based on the number of false positives as a percentage of the predictor and false negatives as a percentage of the criterion. This facilitated the understanding of the accuracy of prediction for the groups as a whole but tended to obscure the accuracy of predictions for individuals.

Another problem with the incomplete use of prediction-performance matrix mainfested itself when some of the studies used either the vertical or horizontal computations alone when reporting their hits and misses. Sometimes the information reported was even less than that. An example of this was seen in the Eaves et al. (1974) study where they generalized that the Modified Predictive Index (MPI) predicted 40-63% of the Grade 2 teachers' opinions. Such information does not tell the reader anything useful.

In order to make comparisons among studies both the horizontal and vertical columns should be reported (i.e., using the horizontal method for computing false positives and the vertical method to

establish false negatives (Gallagher & Bradley, 1972). Overall hit rates were reported by Butler (1982) to be 87.15% and Satz and Friel (1974) to be 84.4%. Virtually all of the studies which used the prediction performance matrix reported that misses were generally confined to the conservative categories where the children initially screened looked mildly risky. Butler et al. (1982) also reported that predictions are superior for students with the poorest reading abilities. Generally, the studies showed that it was not very difficult to identify children who can be expected to fail in school because of their high risk status.

Another concern expressed by many of the researchers was the use of a cutoff at which children are designated at risk. Again most of the studies did not report their criterion for cutoff. Butler et al. (1982) assumed that the target group was 15% of the poorest readers and that a criterion of the botton 10% of predicted scores was chosen to classify children as at risk. In using this decision rule they minimized the number of normal readers who were misclassified but at the expense of missing children who were poor readers. If they had used a criterion of the bottom 30%, they would have correctly identified more poor readers, but only at the expense of misclassifying more normal readers. The use of expectancy tables helps make the consequences of different decision rules more explicit. The best decision rule will depend on the consequences of the decision.

The conclusions which can be drawn from these studies seem to indicate that early screening can be useful when certain elements are included. On the other hand, if the prerequisites are not a part of the early screening procedure the benefits are likely to be lacking.

Salient information delineated includes:

1. According to Keogh (1977) predictive validity is increased when predictions are made to outcomes which are close in time and which are directly related to the particular programs in which the child will be placed. Readiness and risk are in part related to how the child and program interact.

2. A second major finding is that there are at least two partially independent components of risk in kindergarten: the first is related to academic aptitude, the second to behavioral adaptability. According to Keogh (1977), many of the behaviors apparently related to immature social and motor behavior at kindergarten tended to drop out as the child experienced the school program. Limitations of academic aptitude became a more powerful influence or determiner of risk as the program demands become greater. Thus, it may be important to include both behavioral and readiness for academic inventories when doing kindergarten screening.

3. In the studies reviewed, the sample size ranged from 49 (Keogh & Smith, 1970) to 732 (Rubin et al., 1978). As Butler

et al. (1982) pointed out, the studies that employ small sample sizes and also consider a large number of predictor variables may cause the results to be seriously inflated estimates of the predictability of later achievement.

4. Another interesting aspect of the studies reviewed was how the authors arrived at the conclusions regarding their studies. For example, Eaves et al. (1974) concluded that their findings did not enable them to make satisfactory predictions based on the relationship between earlier performance on motor, perceptual, and language tasks and later progress in reading skills. However, the multiple correlations they reported between the Modified Predictive Index and the Cooperative Primary Tests ranged from .55 to .86. On the other hand, Glazzard (1980) concluded that from the results of her study, both a reading readiness test and a kindergarten teacher rating scale were predictive of reading achievement one, two, three, and four years later. The multiple correlations reported (R = .74-.81) were very similar to the Eaves et al. study mentioned above, yet the opposite conclusion was drawn. Such contradictory conclusions make it difficult for the reviewer to make comparisons between studies.

CHAPTER 3

METHODS AND PROCEDURES

The subjects in this study, the instruments used, the procedures followed and the method of analyzing the data are described in this chapter.

Subjects

One hundred and sixty-four white, middle-class children from a midwestern city (population approximately 350,000) were the original subjects of this study. All subjects were enrolled in one of four elementary parochial schools in the Fall of 1980 and were currently in either the first or second grades in their respective schools.

Because of attrition, not all of the original subjects who were involved in the 1980 prekindergarten screening were able to be included in the final data. All students still residing in the original parochial schools were included in the study. The final group consisted of 113 from 164 subjects. Enrollment was distributed among the four schools as follows: (a) School 1 - 37; (b) School 2 - 34; (c) School 3 - 24; and (d) School 4 - 18. One hundred of the remaining subjects were enrolled in second grade while the remaining 13 subjects were enrolled in first grade as a result of being held back one year. The mean age of the children at the time of the kindergarten screening was 5.2.

Instruments

The following is a description of the four predictor screening instruments and the two criterion follow-up instruments used in the study.

Developmental Test of Visual-Motor Integration (VMI)

The VMI (Berry & Buktenica, 1967) assesses the degree to which visual perception and motor behavior are integrated. The author postulates five developmental levels on the road to such integration: motor proficiency, tactual-kinesthetic tracing, visual perception, and visual-motor integration. The integration of relevant functions is considered a prerequisite to academic success; the VMI is specifically concerned with the visual-motor aspect.

The VMI consists of a series of 24 geometric forms which the subject is asked to copy without erasures or corrections. The test is for ages 2 to 15 years and scores provide a visual-motor integration-age equivalent. It takes about 10 minutes to administer, and minimal training is necessary.

The standardization group consisted of 1,039 normal children, 3 to 14 years of age, from suburban, urban, and rural areas in Illinois. No other information was given about the standardization sample in the manual. Test-retest realiabilities, with intervals of two to eight weeks, range in the low .80s, while internal consistency reliabilities range from the .70s to the low .90s.

The concurrent validity of the test is satisfactory, using such criteria as chronological age (r = .89), reading achievement (r = .50), mental age (r = .38 to .59), perceptual skill (r = .80), and psycholinguistic skills (r = .20 to .81) (Sattler, 1982).

Peabody Picture Vocabulary Test (PPVT)

The PPVT (Dunn, 1965) is assumed to provide an estimate of an individual's receptive language ability. It is administered individually to each child. The child is asked to identify, from among four alternatives, a correct pictorial representation as the examiner speaks a word corresponding to each picture. Items are arranged from simple to complex. This test is suitable for use with children of preschool age and beyond and is easily administered. The PPVT requires little special training for scoring and interpretation.

The PPVT was standardized on a population of 4,012 white subjects, aged 2-6 years through 18-0 years. The coefficients, using raw scores, ranged from .67 at age 6-0 to .84 at ages 17-0 and 18-0. The standard errors of measurement in IQ points range *from 6.0 at ages 17-0 and 18-0 years to 8.61 at age 6.0 years.* The alternate form reliability coefficients range from .37 to .97 with a median of .77. The highest coefficients are with cerebral palsied and mentally retarded groups, whereas the lowest coefficients are with preschool children and with Head Start children. Test-retest studies suggest that the PPVT generally provides stable IQ's.

The PPVT manual indicates that content validity is high because Webster's New Collegiate Dictionary (G & C Merriam, 1953) was used as the source for all words whose meaning could be depicted by a picture. The median correlation of studies comparing the PPVT and Stanford-Binet is .66. The findings of studies which correlate the PPVT with the WISC indicate that the highest correlations are found with the Verbal Scale (.36 to .94, median correlation of .66), next highest with the Full Scale (range of .30 to .84 median correlation of .63), and lowest with the Performance Scale (range of .21 to .74, median correlation of .54) (Sattler, 1982).

Preschool Language Scale (PLS)

The PLS (Peotone, 1973) attempts to evaluate preacademic readiness by measuring the following skills: language, auditory and vocal-motor integration (comprehension and speech), auditory-visual integration (reading), and auditory-visual-manual-motor integration (writing).

During the three years of development, the PLS had been administered to over 2,500 prekindergarten children. The PLS has gone through a series of six revisions leading to its final form. Four Illinois communities were included in the testing validation study. According to the manual, the population ranged from lower income families living in trailer courts to wealthy white-collar workers living in expensive new housing developments. The information on the population does not give actual numbers or

percentages of race, sex, or SES. The Pearson coefficient for the total population comparing the pre and post test language scores resulted in a correlation of .77.

The correlation coefficient for the total population between the PLS raw score and the Binet mental age was .77 (Peotone, 1973).

Motor Activity Scale (MAS)

The MAS (Peotone, 1973) is administered individually to each child. It attempts to assess non-verbal development. Perceptual-motor skills are one part of a child's non-verbal development and involves both an awareness of objects and information through the senses and the ability to perform coordinated movements. This test requires approximately 10 to 15 minutes for administration. The examiner can learn how to administer the test quite easily. It is best if an aide is present to assist with the scoring, timing, and giving of directions.

The MAS is divided into four subtests: (a) Body Imagery; (b) Manual Dexterity; (c) Body Movement; and (d) Eye-motor Coordination. The MAS has not been standardized. There are no available reliability and validity studies.

Gates-MacGinitie Reading Test (Gates)

The Gates is a reading achievement test published in 1978 (Vocabulary and Comprehension) which yields raw scores, stanines, percentile ranks, grade equivalents, and extended scale scores.

Kuder-Richardson Formula 20 reliability coefficients were computed for each test level and ranged from .90 to .95 for Vocabulary and from .88 to .94 for Comprehension. The Gates is appropriate for grades 1 through 12. Levels A and B, Form 1 were used in this study.

Behavior Rating Profile (BRP)

According to the test manual, the BRP (Brown & Hammill, 1978) has as its purpose to provide an ecological evaluation of students' behaviors that is well standardized, highly reliable, experimentally validated, and norm-referenced. It is described by the authors as an ecological/behavioral assessment device which permits students' behaviors to be examined in a variety of settings and from several pertinent points of view.

The BRP is comprised of six components: The Student Rating Scales: Home, School, and Peer; the Teacher Rating Scale; the Parent Rating Scale; and the Sociogram. The 30 items of the scale are descriptive sentence items which the teacher or the parents are asked to classify into four categories ranging from, "Very Much Like the Student," to "Not At All Like the Student."

The internal consistency reliability estimate for the Teacher Rating Scale for second and third grades is .97. The BRP has been correlated with the Quay-Peterson Behavior Problem Checklist and the Vineland Social Maturity Scale and this correlation is .84 for both. There was not a significant correlation between the BRP and

the Walker Problem Behavior Identification Checklist. Construct validity is also considered to be good. The coefficients range is size from .49 to .96 with a median coefficient of .81 (Brown & Hammill, 1978).

Procedures

In late April and early May of 1980, 164 prekindergarten children were screened at four different parochial schools in a large midwestern metropolitan city. The screening instruments used and described earlier were the VMI, PPVT, PLS, and the MAS. Each child was seen individually by a six-member team which included two learning disability teachers, two kindergarten teachers, a physical education teacher, and a speech therapist. Three aides assisted in the screening program but were not involved in the actual assessment process.

Prior to the kindergarten screening, average ranges had been established as estimates of acceptable or unacceptable achievement on each of the four screening instruments. These scores were based on the average range of scores for the kindergarten screening tests which had been given to 332 children in a neighboring district in the spring of 1978. Although these scores were used in the identification process, they were not the sole criteria on which decisions were made. The child's total performance was taken into consideration. This included his/her attention, ability to concentrate, cooperations, and other behaviors subjectively evaluated by team members.

Each child to be screening rotated among the five stations and remained at each station from 10 to 15 minutes. It took approximately 60 to 90 minutes to screen about 25 children. After the screening, members of the team met to discuss the results of their respective tests as well as to share their observations concerning the behavior of each child. Decisions were then made to place the child in one of the following categories: (a) Readiness good in all areas; (b) Readiness good in most areas; (c) At-Risk developmentally (delay kindergarten); and (d) At-Risk (further testing required).

Parents were notified regarding the results of the screening if their child fell in categories "b," "c," or "d." They were not contacted if their child fell in category "a." For parents whose child was categorized "b" a list of suggestions was provided for them to work with their child in whatever areas tended to be low. The strengths of each child were also shared with the parents. The deficits noted in category "b" children were not considered significant enough to warrant further testing. For category "c", one member of the screening team discussed with the parents the reasons why the recommendation was made to delay kindergarten entry for another year. Parents were given the option to follow this advice, to start the child, or to get a second opinion. The parents of category "d" children were advised to permit further testing with their child because of significant problems noted on

the initial screening measures. Students could receive further diagnostic testing by a school psychologist later in the spring or parents had the option of taking their child to any other agency for diagnostic testing. Based on the results of the diagnostic tests administered later, it was determined if the child needed a public school special class placement, counseling, or some other type of intervention within the regular classroom. The numbers from each school are shown in Table 1.

Table 1

Readiness Categories of Children Originally Screened, Spring, 1980

	Schools				
Category	1	2	3	4	Total
a	32	33	20	16	101
b	6	10	8	6	30
c	12	5	4	3	24
d	5	2	1	1	9
Totals	55	50	33	26	164

No special classes were available in any of the schools included in the study. All subjects who started kindergarten in the Fall of 1980 were enrolled in one of the four regular kindergarten classes with varying degrees of supportive assistance available to the kindergarten teacher.

A follow-up study was done in March of 1983, almost three years later, to determine how well these same children were doing both academically and behaviorally. One hundred and thirteen of the original 164 subjects continued to be enrolled in one of the four schools in either the first (N = 13) or second (N= 100) grades. Of the 19 children who were identified as category "c" (delay kindergarten entry), 13 were either delayed or held back a year but were not delayed (by parent demand). Table 2 summarizes the numbers and categories of the children who were available for follow-up.

The criterion measures, described earlier, were the Gates-MacGinitie Reading Tests and the Behavior Rating Profile (Teacher Rating Form). The second grade teachers from each of the four schools administered the Gates (Level B, Form 1) sometime during the week of March 14-18, 1983. They also rated their students' behavior according to the Teacher Rating Form of the BRP. The first grade Gates (Level A, Form 1) was administered by the researcher while the first grade teachers rated their students using the Teacher Rating Form of the BRP. In addition, parents of the children who were originally categorized as "At-Risk" (Category c) were contacted

by the researcher so as to obtain their degree of satisfaction regarding the current academic status of their children.

Table 2

Children Included in Follow-Up, Spring, 1983

	Schools				
Category	1	2	3	4	Total
a	23	22	12	10	67
b	4	7	8	6	25
c	10	3	4	2	19
d	0	2	0	0	2
Totals	37	34	24	18	113

Data Analysis

The Statistical Package for the Social Sciences (SPSS), a system of computer programs available at Academic Computing Services, University of Northern Iowa, was used for the statistical analysis of this study.

The study first evaluated the relationship between the prekindergarten screening measures and reading achievement three

years later. To test the first null hpothesis that there is no significant correlation between prekindergarten measures and reading achievement three years later, the Pearson product-moment correlation coefficient was calculated. Significance was set at $p < .05$. Multiple regression was used to analyze the relationship between the dependent or criterion variables (Gates and BRP) and the set of independent or predictor variables (PPVT, VMI, PLS, MAS). The forward (stepwise) inclusion procedures was used to isolate a subset of available predictor variables. Independent variables were entered only if they met two statistical criteria. First, an F ratio was computed in a test for significance of a regression coefficient. Significance was set at .01. Secondly, the tolerance of an independent variable considered for inclusion was set at T = .001. According to Kim and Kohart (1975), tolerance is " . . . the proportion of the variance of that variable not explained by the independent variables already in the regression equation" (p. 346). Since a tolerance index has a range of 0 to 1, a tolerance of .6 would suggest that 60% of the variance is unexplained by the predictors already entered.

The second null hypothesis, which stated that there is no significant correlation between prekindergarten measures and behavioral functioning three years later, was also tested using the Pearson product-moment correlation coefficient ($p < .05$).

Multiple regression analysis as described above was used again to test the second hypothesis.

To answer research questions 1 through 4, prediction-performance matrices were constructed to determine the hit rate in percentages of valid positives, false positives, valid negatives, and false negatives. An example of a prediction-performance comparison matrix is presented in Figure 1 (Mercer, Algnozzine, & Trifiletti, 1979). Four possible outcomes are indicated.

Use of vertical and horizontal computations were used to report the conclusions of this study. Figure 2 (Mercer, Algnozzine, & Trifiletti, 1979) presents a numerical example of data collected from a hypothetical prediction study. By using the horizontal analysis method, percentages of correct and incorrect outcomes can be noted. For example 80 of the 100 children (i.e., 80% who were "at-risk" actually did poorly while 20% (20 of 100) did well. Figure 2 also includes vertically computed percentages. For example, 67% (80 of 120) of the low performing children were accurately predicted and 7% (20 of 280) of the average of high performers were predicted to do poorly (i.e., false positives).

Research question 5 sought to determine how the students who were in first grade at the time of the study performed on a reading criterion measure three years later. By analyzing the scores received on the Gates, the students were grouped into categories of successful reading achievement (at or above the 50th percentile

Figure 1. Example of prediction-performance comparison matrix.

	Performance	
	Poor	Good
Prediction		
"At-Risk"	Predicted poor Performed poor (Valid positive) A	Predicted poor Performed good (False positive) B
"Ready"	Predicted good Performed poor (False negative) C	Predicted good Performed good (Valid negative) D

Figure 2. Numerical example of prediction-performance comparison matrix with examples of vertical (V) and horizontal (H) hit rates.

	Poor	Good	
	80	20	100
"At-Risk"	(80%) H	(20%) H	predicted
	(67%) V	(7%) V	poor
Prediction			
	40	260	300
"Ready"	(13%) H	(87%) H	predicted
	(33%) V	(93%) V	good
	120 performed poorly	280 performed well	400 total

rank on the Gates) or unsuccessful reading achievement (below the 50th percentile rank on the Gates).

To answer research question 6, which asked how the students who were in first grade at the time of the study were doing behaviorally, the criterion data on the BRP was analyzed. Students who were rated at or above normal (Standard score of 8 or above) in their behavioral functioning were considered as having no behavioral problems while those who received a score of seven or less on the BRP were considered to have behavioral problems.

Chapter 4 analyzes the results of the data collected from the predictor and criterion variables.

CHAPTER 4

RESULTS AND DISCUSSION

Chapter 4 presents the empirical results and analysis of this study. The means and standard deviations for all related test scores are initially presented. This data will describe the overall performance of the children used in this study. The Pearson product-moment coefficients, multiple regression correlation, prediction-performance matrices, and the reading achievement and behavior ratings of the related first grade students are calculated according to each corresponding hypothesis. Because results are somewhat detailed, results and discussion for each hypothesis appears together.

The means and standard deviation for the subjects involved in this study as well as the average range of scores from a previous study done in a neighboring district are presented in Table 3. These average scores served as a general guide for determining acceptable average performance in the present screening procedure. They are, in fact, very similar to the average scores of the current study.

As a group, the average performance on the early screening measures were about average when compared with the average ranges for the neighboring district. Scores for both criterion measures (Gates and BRP) were above average. The Gates-MacGinitie mean grade equivalent was 3.8 or approximately one year above actual

grade placement, while the BRP yielded had an average score of 11.8 which is above the BRP's mean of 10 for the norm sample.

Table 3

Average Range, Mean, and Standard Deviation of Subjects on Test Scores

Variable	Average range in previous study	Mean	Standard deviation
PPVT (IQ)	93 - 127	114.34	13.63
PLS (Raw score)	24 - 40	32.53	5.46
VMI (Age in months)	-	66.83	10.81
MAS (Raw score)	9 - 15	12.79	1.22
Gates (Grade level)	-	3.8	1.11
BRP (Standard score)	-	11.82	2.56

Note. Dash (-) indicates that the variable was not used in the previous study.

Hypotheses

Hypothesis 1 stated that there is no significant relationship between prekindergarten measures and reading achievement three years later. Pearson product-moment coefficients were computed between

the four predictor variables (PPVT, PLS, VMI, PLS) and the Gates and were statistically significant ($p < .05$). Table 4 presents these single correlations as well as a summary of results for the multiple regression analysis. Multiple correlation ($R = .48$) showed approximately 23% of the variance in reading achievement was accounted for by four independent variables which leaves 77% of the variance unexplained.

Table 4

Single and Multiple Correlations between Predictor Variables and Gates Reading Scores

Predictor Variable	Multiple R	R Square	Simple r
PPVT	.34	.12	.34
VMI	.42	.17	.25
MAS	.46	.21	.24
PLS	.48	.23	.32

Hypothesis 1 was rejected since a significant ($p < .05$) statistical correlation was found between each of the predictor variables and the criterion variable, reading achievement. Although the relationships were statistically significant, the

practical significance of these relative small correlations (range: .24-.34) is debatable. Minimal confidence can be put in a set of predictors whose R = .23 and whose unexplained variance is 77%. These results are similar to studies (Eaves et al., 1974; Lindquist, 1982; Rubin, 1978) that reported negative results concluded negatively and dissimilar to studies (Glazzard, 1982; Keogh & Smith) that reported positive results.

The second hypothesis stated that there is no significant relationship ($p < .05$) between prekindergarten measures and behavioral functioning three years later. The Pearson product-moment correlations indicate that there are no significant single correlations between the BRP and the independent variables (PPVT, PLS, VMI, MAS). Therefore, hypothesis 2 could not be rejected. Table 5 presents the single correlations as well as a summary of results for the multiple regression analysis. It should be noted that the VMI dropped out of the multiple regression analysis since it did not meet statistical criteria for inclusion.

Correlations were not significant as the screening measures used were not designed to predict behavioral functioning but rather were designed to measure language, perceptual-motor and fine and gross motor skills.

Research Questions

Research questions 1 and 2 asked what percent of the children identified as either "at-risk" or "ready" lagged in reading

Table 5

Single and Multiple Correlations between Predictor Variables and Behavior Rating Profile

Predictor Variable	Multiple R	R Square	Simple r
PPVT	.15	.02	-.15
PLS	.22	.05	.10
MAS	.22	.05	.03
VMI	-	-	.04

achievement measured by the Gates-MacGinitie and what percent of the same children were achieving at or above the 50th percentile rank on the Gates. Table 6 displays the actual numbers of children falling within each cell of the prediction-performance matrix (comparing reading achievement with early identification) and the column and grand totals. The overall "hit rate" for predicting reading achievement was 86%. By applying the horizontal analysis method, percentages of correct and incorrect outcomes were obtained. Three out of the eight children (i.e., 37.5%) who were identified as "at-risk" actually did poorly while 62.5% (5 of 8) did well. A primary fact states that 83 of the 92 children (i.e., 90%) who were identified as "ready" actually did well and 10% (9 of 92) did poorly.

Table 6

Prediction-Performance Matrix for Reading Achievement with Hit Rate Percentage for Horizontal (H) and Vertical (V) Analysis

	Performance		
	Poor	Good	
Predictions "At-Risk"	N = 3 37.5% (H) 25% (V)	N = 5 62.5% (H) 6% (V)	TN = 8
"Ready"	N = 9 10% (H) 75% (V)	N = 83 90% (H) 94% (V)	TN = 92
	TN = 12	TN = 88	GT = 100

Vertical analysis was also performed to consider the relationship between the observed values (i.e., within the cells) and the actual performance levels. While only 10% of those identified as "ready" actually performed poorly, the predictor tests were incorrect for 75% (9 of 12) of those who performed

poorly. In addition, 25% (3 of 12) of the poorly performing children were accurately predicted and 6% (5 of 88) of the good performers were identified as "at-risk" (i.e., false positives).

This overall "hit rate" compares favorably with the "hit rates" reported in the literature. In Satz and Friel's study (1974) a "hit rate" of 84.4% was considered good as was the 87.5% "hit rate" reported by Butler et al. (1982) and the 85% reported by Mercer, Algnozzine and Trifiletti (1979). The major errors made were false positives and false negatives. No special class or remedial program was initiated for these children so that the screening did not result in special education programming or labeling. Since no control was set on the variables of teacher effectiveness, curriculum strengths, or parental knowledge and intervention, the effects of extra help or special attention to some areas of weakness are not known.

It should also be mentioned here that the 86% overall "hit rate" reported in this study did not include those 13 of the 19 students who were originally identified in the "c" category, "at-risk developmentally." These students waited a year to enter kindergarten or were later held back a year and subsequently were in first grade at the time of the criterion testing. No effort was made by this researcher to locate the other 12 in category "c" or "d" who were involved in the original screening program and identified at that time as "at-risk". Some of these children

transferred to the public school where special classes were available.

Research questions 3 and 4 asked what percent of the children originally identified as "at-risk" or "ready" displayed significant behavior problems and what percent are within the "normal range" or above according to the BRP administered three years later.

Table 7 displays a prediction-performance matrix comparing early prediction with second grade behavioral functioning. Using the horizontal analysis method, percentages of correct and incorrect outcomes were obtained. Three of the eight children (i.e., 37.5%) who were identified as "at-risk" were rated as having behavior problems while 62.5% (5 of 8) were rated satisfactorily in behavioral functioning. Also, 85 of 92 children (92%) who were identified as "ready" were rated satisfactorily in behavioral functioning and 8% (7 of 92) were rated as having behavior problems.

Vertical analysis showed that while only 8% of those identified as "ready" actually were rated as having behavioral problems, the prediction was incorrect for 70% (7 of 10) of those who were rated as having behavioral problems. In addition 30% (3 of 10) of the children who were rated as having behavior problems were accurately predicted and 6% (5 of 90) of the children rated with satisfactory behavior were identified as "at risk" (i.e., false positives). The overall "hit rate" for predicting behavioral functioning was 88%.

Table 7

Prediction-Performance Matrix for Behavioral Functioning with Hit Rate Percentages for Horizontal (H) and Vertical (V) Analysis

	Performance		
	Poor	Good	
Prediction			
"At-Risk"	N = 3 37.5% (H) 30% (V)	N = 5 62.5% (H) 6% (V)	TN = 8
"Ready"	N = 7 8% (H) 70% (V)	N = 85 92% (H) 94% (V)	TN = 92
	TN = 10	TN = 90	GT = 100

As in the previous discussion regarding research questions 1 and 2, the largest number of misses was confined to the false positive cell where five out of the eight identified as "at-risk" were not found to have significant behavioral problems and seven of the 10 (70%) identified as "ready" were later rated as having behavioral problems.

The "hit rates" for each cell when using the criterion BRP are similar to the results when using the Gates as a criterion. This finding is not surprising as the decision to classify the children during the initial screening process as either "ready" or "at-risk" was not based on performance scores of the set of independent variables alone. The observations of each examiner regarding the child's behavior as he or she performed was an integral part of the final decision in the identification process.

Research question 5 asked how the children who waited a year to begin kindergarten or who had repeated a grade were doing in their reading achievement. Table 8 shows the first grade students' reading achievement in grade equivalents and percentile rankings. It also shows the behavioral functioning according to standard scores and percentile rankings.

Out of the 13 children who were identified as "at-risk" and who were in first grade at the time of the study, eight of them (62%) were at or above the 50th percentile rank in reading achievement. Five of the 13 children (38%) were measured by the Gates as below the 50th percentile rank. Of the remaining eight who scored above the 50th percentile rank, three of them received equivalents in the third grade or the 90th percentile rank. Questions arise whether or not these three children may not have been just as well off starting kindergarten a year earlier.

Table 8

First Grade Students Reading Achievement and Behavioral Functioning as Measured by the Gates-MacGinitie and the Behavior Rating Profile (Teacher Rating Form)

Student	Gates-MacGinitie		Behavior Rating Profile	
	Grade equivalent	Percentile rank	Scaled score	Percentile rank
1	2.0	63%	15	96%
2	3.2	90%	15	96%
3	1.5	33%*	10	50%
4	1.2	22%*	11	63%
5	2.5	78%	11	63%
6	1.5	37%*	12	75%
7	1.7	59%	9	37%
8	1.9	70%	15	96%
9	1.6	44%*	9	37%
10	2.1	73%	12	75%
11	1.4	28%*	7	16%*
12	3.6	95%	9	37%
13	3.8	96%	12	75%

Note. * signifies below the acceptable criterion.

The final research question, number 6, asked how the 13 children who were identified as "at-risk" and were in first grade at the time of the study, functioned behaviorally. Table 8 reveals that only one of the 13 children is rated by the BRP as having significant behavior problems.

Keeping the "c" category children back a year may have been a good decision. However, as mentioned above, how well they would have performed had they started kindergarten with the other children who were screened at that time cannot be determined.

At the end of the study the researcher contacted the parents of those children who were initially identified as being "at-risk" (Category c) to ascertain how they regarded the decision they had made three years ago to either keep their child out of kindergarten for a year or to start them in kindergarten regardless of the screening program's recommendations. (See Appendices A & B for actual parental comments.)

Of the 13 children who were in first grade or held back a year, virtually all of the parents contacted were moderately happy to extremely satisfied with their decision. Regarding the six children who were identified as "developmentally at risk" but whose parents decided to start them in kindergarten regardless of the screening recommendations, the parental comments reflected less satisfaction with their decision. The comments indicated that some parents mentioned "immaturity," "need for extra help,"

and "we've learned a lesson." One parent was extremely satisfied with her decision.

Summary

The major findings of this chapter include the following:

1. There was a small but statistically significant correlation between each of the early screening measures and reading achievement.

2. There was no significant correlation between the early screening measures and behavioral functioning.

3. The overall hit rate was considered good as was the valid negative rate.

4. False positives and false negatives tended to be overrepresented.

Chapter 5 will summarize this study, draw conclusions and offer recommendations for utilizing this screening program in the future.

CHAPTER 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

On the basis of this study, it can be concluded that the early screening program evaluated here has both strengths and limitations. In this chapter is a summation of the shortcomings and positive aspects of the study and recommendations for future application.

Summary

This study evaluated the kindergarten screening program used in four elementary schools in a large metropolitan midwestern community. Subjects were 113 primarily white middle-class students who participated in the kindergarten screening during the Spring of 1980 and who were in either second (N = 100) or first (N = 13) grade at the time of this study.

Data were collected on early screening measures (PPVT, VMI, PLS, MAS) which comprised a set of predictor variables used to identify students as "at-risk" or "ready" to begin kindergarten. Three years later (1983) these same subjects were evaluated using the criterion measures of reading achievement (Gates-MacGinitie) and behavioral functioning (Behavior Rating Profile - Teacher Rating Form).

Data were analyzed by using the Pearson product-moment correlations and multiple regression analysis. A comparison of levels of prediction (i.e., "at-risk" or "ready") with levels of criterion performance (i.e., "above or below the 50th percentile

rank in reading") and behavior (three or more scaled scores below the mean or above) in behavioral functioning.

Results indicated that each of the predictor variables were statistically correlated with later reading performance ($p < .05$) and thus the first hypothesis was rejected. However, the Pearson product-moment correlations $r = .24$ to $.34$) and the multiple correlation ($R = .48$) between the screening instruments and the reading criterion were relatively low since the amount of variance accounted for was 23%. Therefore, the predictive validity of this set of screening instruments is problematic. The second hypothesis was substantiated when there were no significant correlations ($p < .05$) between the predictor variables and the behavioral criterion (BRP).

The study also made use of the prediction-performance matrix in order to evaluate the "hits" and "misses." The overall hit rate for analysis between the set of predictor variables and the criterion variable was 86% for reading achievement and 88% for behavioral functioning. Both matrices revealed a large percentage of false positives and false negatives. Thus, even though the overall efficiency was high, 62.5% of the children who later functioned satisfactorily in reading and behavior had been originally classified as "at-risk," 75% who were identified as "ready" actually performed poorly in reading, and 70% who were identified as "ready" performed poorly in behavioral functioning.

The first grade students were also evaluated according to their reading achievement and behavioral functioning. Out of the 13 students available at the time of the study, five were below the acceptable criterion in reading and one in behavior. An informal survey of their parents indicated a greater satisfaction rate than of the parents who were originally cautioned to keep their child back a year but who decided to start them. Subjective perceptions indicate that generally parents are satisfied with the decision they made and do their best to help their children achieve successfully.

It was concluded that while the overall hit rate was satisfactory, it included large percentages of "misses" (false negatives and false positives). Also, the particular set of predictor variables used (PPVT, VMI, PLS, MAS) should be discontinued as the correlations were not strong enough to warrant continued used as a set of early screening measures.

Conclusions and Recommendations

The present study raises a serious note of caution regarding prekindergarten identification of potential reading or behavioral problems. Hypothesis 1 was rejected as a significant statistical relationship ($p < .05$) was found between the prekindergarten measures (PPVT, VMI, PLS, MAS) and reading achievement three years later. The Pearson product-moment coefficients yielded scores that were statistically significant, yet, the correlations are relatively

small (.24 to .34) and only 23% of the variance in reading achievement was accounted for by the four screening tests.

Therefore, it is recommended that the continued use of this particular set of screening instruments be discontinued and a search be made to include instruments which tap a greater variety of readiness skills and have greater predictive power. The researcher considered the early screening measures to have tapped a variety of readiness skills (e.g., language, vocabulary, visual and auditory perception, and motor skills). Yet multiple regression analysis did not indicate that each test contributed significantly to the total amount of variance (77% of the variance was still unaccountable).

Hypothesis 2 was accepted as there was not a significant relationship ($p < .05$) between prekindergarten measures and behavioral functioning three years later. It is recommended that the screening program adopt a more structured behavior rating inventory where teachers could rate behaviors on a well-defined scale. Another option would be to have the BRP or some similar behavior rating scale completed by the parent or the preschool teacher, if the child had been in a preschool program. Both parents and former teachers could be good sources of information regarding the child's readiness to enter kindergarten. Combining the objective screening measures mentioned above along with the more subjective behavioral

observations and rating scales could provide a good balance to the decision-making process.

Research questions 1 through 4 asked about the predictive validity of the screening program. A comparison of levels of prediction (i.e., "at-risk" or "ready") with levels of criterion performance (i.e., poor or good) was used as the model for evaluating the overall "hit rate" as well as the percentages of false positives, false negatives, valid positives, and valid negatives.

The overall "hit rate" of this screening program was good (86%-88%). Being able to classify 90% of those who would do well in reading and behavioral functioning and 37.5% of those who did not was a reasonably good prediction rate. It can be argued here that one could make almost as good of a prediction by utilizing a much simpler screening procedure (i.e., knowing the child's age and spending some informal time with the child). In most groups of children one can usually assume that 85% to 90% of them will not have either learning or behavioral problems. Therefore, it is a fairly easy process to predict which ones will continue to do well in school even without administering screening measures.

The significantly high percentage of false positives (62.5%) and false negatives (70% and 75%) is of concern. As indicated in Chapter 4, no labeling or special programming resulted as a consequence of this screening program. However, it does represent

a potentially serious cost to the schools and to society as well as causing possible alarm and unnecessary concern to the parents. Lowering the cutoff scores would reduce the number of false positives and at the same time increase the percentage of false negatives, thereby running the risk of overlooking some children who may need special attention. Thus, it is recommended that the screening results always be viewed only as a preliminary process to be followed by continued observation and further assessment and intervention if necessary.

Research question 5 and 6 asked how the children who were either kept back from entering kindergarten or who later repeated a year were doing in reading achievement and in behavioral functioning. In other words, just how good were the recommendations which were made to keep some of the children out of kindergarten for a year?

The criterion variables, Gates, indicated that out of the 13 children, five of them (38%) fell below the 50th percentile rank in reading achievement, while three of the 13 scores in the 90th percentile rank. It is difficult to conclude just how valuable was the recommendation to hold the child back. Judging from the parental comments and teacher ratings on the BRP, it appears that the screening program was successful in preventing some children from starting school before they were ready. On the other hand, the possibility of their getting into school and being exposed to

the proper program they needed may have also been helpful. The ability to discern the difference between "immaturity" and "deficit skills" remains confusing; test results per se do not usually make the distinction that clear. The focus of a good screening program should not be on the "sorting" process alone (i.e., "at-risk" or "ready") but also on the guidance and information which can be given to the parents and teachers of those children who have been screened. The critical elements here are the total concern of the child and the experience of the examiners who are making the decisions.

Finally, it is recommended that those responsible for the preparation and evaluation of screening programs continue to evaluate the predictive accuracy of screening procedures through follow-up studies such as this one. While it is important to have a high satisfaction rate among parents, principals, and teachers, it is not enough to provide evidence that the program is as reliable and valid as it should be to protect the rights of all children.

REFERENCES

Abeson, A., & Zettel, J. (1977). The end of the quiet revolution: The education for all handicapped children act. Exceptional Children, 44, 114-128.

Adkins, P. L., Holmes, G. R., & Schnackenberg, R. C. (1971). Factor analysis of the de Hirsch predictive index. Perceptual Motor Skills, 33, 1319-1325.

Ahr, A. E. (1967). The development of a group preschool screening test of early school entrance potential. Psychology in the Schools, 4, 59-63.

Ames, S. G., Becker, L. D., & Dalton, S. (1977). The predictive validity of the Stanford Early Achievement Test. Educational and Psychological Measurement, 37, 505-507.

Ammons, R. B., & Ammons, C. H. (1962). The Quick Test (QT) provisional manual. Psychological Reports, 11, 111-161. (Monographs Supplement I-VII).

Anderson, B. (1970). Classroom diagnosis of R.R. factors. Journal of Learning Disabilities, 3, 260-263.

Berry, K. E., & Buktenica, N. A. (1967). Developmental Test of Visual-Motor Integration. Chicago: Follett Publishing Company.

Brown, L. L., & Hammill, D. D. (1978). Behavior Rating Profile: An Ecological Approach to Behavioral Assessment. Austin, Texas: Pro-ed.

Bryan, Q. R. (1964). Relative importance of intelligence and visual perception in predicting reading achievement. California Journal of Education Research, 15, 44-48.

Butler, S. R., Marsh, H. W., Sheppard, M. J., & Sheppard, J. L. (1982). Early prediction of reading achievement with the Sheppard School Entry Screening Test: A four-year longitudinal study. Journal of Educational Psychology, 74, 280-290.

Buttram, J., Covert, R. W., & Hayes, M. (1976). Prediction of school readiness and early grade achievement by classroom teachers. Educational and Psychological Measurement, 36, 543-546.

Camp, B. W., & Zimet, S. G. (1974). The relationship of teacher rating scale to behavior observation and reading achievement of first grade children. The Journal of Special Education, 8, 353-359.

Carver, R. (1978). The case against statistical significance. Harvard Educational Review, 48, 379-399.

Chomsky, C. (1976). The acquisition of syntax in children from 5 to 10 (Research Monograph No. 57). Cambridge, Mass.: The M.I.T. Press.

Cowgill, M. L., Friedland, S., & Shapiro, R. (1973). Predicting learning from kindergarten reports. Journal of Learning Disabilities, 6, 577-582.

de Hirsch, K., Jansky, J. J., & Langford, W. S. (1966). Predicting reading failure. New York: Harper & Row.

Dunn, L. M. (1965). Peabody Picture Vocabulary Test Manual. Circle Pines, Minnesota: American Guidance Service.

Durrett, M. E., & Henman, J. (1972). Concurrent validity of the PPVT, Draw-A-Man, and Children's Embedded Figures Test with four year old children. Educational and Psychological Measurement, 32, 1089-1093.

Eaves, L. C., Kendall, D. C., & Crichton, J. U. (1974). The early identification of learning disabilities: A follow-up study. Journal of Learning Disabilities, 7, 632-638.

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

Ernhart, C. B., Jordan, T. E., & Spaner, S. D. (1971). Maternal Quick Test scores in child development research. Psychological Reports, 28, 669-670.

Ernhart, C. B., Spaner, S. D., & Jordan, T. E. (1977). Validity of selected preschool screening tests. Contemporary Educational Psychology, 2, 78-89.

Federal Register. (August 23, 1977, pp. 42474-42518). Education of handicapped children. Regulations Implementing Education for all Handicapped Children Act of 1975 (Public Law 94-142).

Feshbach, S., Adelman, H., & Fuller, W. (1974). Early identification of children with high risk of reading failure. Journal of Learning Disabilities, 7, 639-644.

Flynn, J. M., & Flynn, L. A. (1978). Evaluation of the predictive ability of five screening measures administered during kindergarten. Journal of Experimental Education, 46, 65-69.

Gallagher, J., & Bradley, R. (1972). Early identification of developmental difficulties. In I. Gordon (Ed.), Early Childhood Education. Chicago: University of Chicago Press.

Gallerani, D., O'Regan, M., & Reinherz, H. (1982). Prekindergarten screening: How well does it predict readiness for first grade? Psychology in the Schools, 19, 175-182.

Glazzard, M. (1977). The effectiveness of three kindergarten predictors for first grade achievement. Journal of Learning Disabilities, 10, 95-99.

Glazzard, M. (1979). The long-range effectiveness of three kindergarten predictors for first, second, third, and fourth grade achievement. Journal of Learning Disabilities, 12, 55-60.

Glazzard, M. (1980). Teacher rating and reading readiness as predictors of vocabulary and comprehension achievement in first, second, third and fourth grade. Learning Disability Quarterly, 3, 35-45.

Glazzard, M. (1982). Long-range kindergarten prediction of reading achievement in first through sixth grade. Learning Disability Quarterly, 5, 85-88.

Graham, F. K., Ernhart, C. B., Thurston, D. L., & Craft, M. (1962). Development three years after perinatal anoxia and other potentially damaging newborn experiences. Psychological Monographs, 76, (3, Whole No. 522).

Hammill, D. D. (1971). Evaluating children for instructional purposes. Academic Therapy, 6, 341-353.

Haring, N. G., & Ridgeway, R. W. (1967). Early identification of children with learning disabilities. Exceptional Children, 33, 387-395.

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, Ohio: Charles E. Merrill.

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.

Hunt, J., & Kirk, C. E. (1974). Criterion-referenced tests of school readiness. Genetic Psychology Monographs, 90, 144-182.

Ilg, L., Ames, L. B., & Aspell, R. J. (1965). School readiness as evaluated by Gesell developmental visual and projective tests. Genetic Psychological Monograph, 71, 61-91.

Jansky, J., & de Hirsch, K. (1973). Preventing reading failure: Prediction, diagnosis and intervention. New York: Harper & Row.

Kapelis, L. (1975). Early identification of reading failure: A comparison of two screening tests and teacher forecasts. Journal of Learning Disabilities, 8, 638-641.

Keogh, B. K. (1965). School achievement associated with successful performance on the Bender Gestalt Test. Journal of School Psychology, 3, 37-40.

Keogh, B. K. (1977). Early ID: Selective perception or perceptive selection? Academic Therapy, 12, 267-275.

Keogh, B. K., & Becker, L. D. (1973). Early detection of learning problems: Questions, cautions, and guidelines. Exceptional Children, 40, 5-11.

Keogh, B., & Smith, C. (1961). Group techniques and proposed scoring system for the Bender Gestalt Test with children. Journal of Clinical Psychology, 17, 172-175.

Keogh, B. K., & Smith, C. E. (1967). Visual-motor ability for school prediction: A seven year study. Perceptual and Motor Skills, 25, 101-110.

Keogh, B. K., & Smith, C. E. (1970). Early identification of educationally high-potential and high risk children. Journal of School Psychology, 8, 285-290.

Keogh, B. K., Tchir, C., & Windeguth-Behn, A. (1974). Teachers' perceptions of educationally high risk children. Journal of Learning Disabilities, 7, 43-50.

Keogh, B., Welles, M., & Hall, R. (1976). An approach to early identification of high potential and high risk pupils (Technical Report). University of California, Los Angeles.

Kim, J., & Kohart, F. J. (1975). Multiple regression analysis: Subprogram regression. In N. H. Nie, C. H. Hull, J. G. Jenkins, K. Steinbrenner, & D. H. Bent (Eds.), Statistical package for the social sciences (2nd ed.). New York: McGraw-Hill Book Co.

Kirk, W. D. (1966). A tentative screening procedure for selection of bright and slow children in kindergarten. Exceptional Child, 32, 235-241.

Koppitz, E. M. (1964). The Bender Gestalt Test for young children. New York: Grune & Stratton.

Koppitz, E. M., Sullivan, J., Blyth, D. D., & Shelton, J. (1959). Prediction of first grade achievement with the Bender-Gestalt Test and Human Figure Drawing. Journal of Clinical Psychology, 15, 164-168.

Kottmeyer, W. (1947). Readiness for reading. Elementary English, 24, 355-366.

Landsman, M., & Dillard, H. (1967). Evanston Early Identification Scale. Chicago: Follett.

Lessler, K., & Bridges, J. S. (1973). The prediction of learning problems in a rural setting: Can we improve on readiness tests? Journal of Learning Disabilities, 6, 90-94.

Lewis, A. (1980). The early identification of children with learning difficulties. Journal of Learning Disabilities, 13, 51-57.

Lilenfield, B. M. (1975). Teacher identification of children with school-related problems: Predictive validity six years later. Education Psychology Abstracts, 6557A. (Order No. 76-8498, 269 pages.)

Lindquist, G. T. (1982). Preschool screening as a means of predicting later reading achievement. Journal of Learning Disabilities, 15, 331-332.

Lindsay, G. A., & Wedell, K. (1982). The early identification of educationally "at risk" children revisited. Journal of Learning Disabilities, 15, 212-217.

Lindsay, G. A. quoted in Wedell, K. (1978). Early identification and compensatory interaction. Paper delivered at NATO International Conference on Learning Disorders, Ottawa.

Lloyd, D. N. (1978). Prediction of school failure from third grade data. Educational and Psychological Measurement, 38, 1193-1200.

Lyons, J. S., & Bangs, T. (1972). Effects of preschool language training on later academic achievement of children with language and learning disabilities: A descriptive analysis. Journal of Learning Disabilities, 5, 585-592.

MacGinitie, W. H., et al. (1978). Gates-MacGinite Reading Tests Teacher's Manual (2nd ed.). Boston: Houghton Mifflin Company.

McCandless, B. (1972). Review of the Boehm Test of Basic Concepts. In O. K. Buros, Seventh Mental Measurements Yearbook. New York: Gryphon.

Mercer, C. D., Algozzine, B., & Trifiletti, J. (1979). Early identification - An analysis of the research. Learning Disability Quarterly, 2, 13-24.

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.

Mitchell, B. C. (1962). The Metropolitan Readiness Tests as prediction of first grade achievement. Educational and Psychological Measurement, 22, 765-772.

Motor Activity Scale. (1973). Poetone, Illinois.

Preschool Language Scale. (1973). Poetone, Illinois.

Robinson, H. A. (1966). Reliability of measures related to reading success of average disadvantaged and advantaged kindergarten children. Reading Teacher, 20, 203-209.

Rogolsky, M. M. (1968-69). Screening kindergarten children: A review and recommendation. Journal of School Psychology, 7, 18-27.

Rosenthal, R., & Jacobsen, L. (1968). Pygmalion in the classroom. New York: Holt, Rinehart & Winston.

Rubin, R. A., & Balow, B. (1977). Perinatal influences on the behavior and learning problems of children. In B. B. Lahey & Kazdin (Eds.), Advances in child clinical psychology. New York: Plenum.

Rubin, R. A., Balow, B., Dorle, J., & Rosen, M. (1978). Preschool prediction of low achievement in basic school skills. Journal of Learning Disabilities, 11, 664-667.

Ryan, J. J., & French, J. R. (1974). Long-term grade predictions for intelligence and achievement tests in schools of differing socioeconomic levels. Educational and Psychological Measurement, 34, 553-559.

Sattler, J. (1982). Assessment of children's intelligence and special abilities. (2nd ed.). Boston: Allyn & Bacon, Inc.

Satz, P., & Friel, J. (1974). Some predictive antecedents of specific reading disability: A preliminary two-year follow up. Journal of Learning Disabilities, 7, 49-55.

Silberberg, N., Iversen, I., & Silberberg, M. (1968). The predictive efficiency of the Gates Reading Readiness Tests. The Elementary School Journal, 68, 213-218.

Smith, C. E., & Keogh, B. K. (1962). The group Bender-Gestalt: A reading readiness screening instrument. Perceptual and Motor Skills, 15, 639-645.

Stevenson, H. W., Parker, T., & Wilkerson, A. (1976). Predictive value of teachers' ratings of young children. Journal of Educational Psychology, 68, 507-517.

Tobiessen, J., Duckworth, B., & Conrad, W. G. (1971). Relationships between the Schenectady Kindergarten Rating Scales and first grade achievement and adjustment. Psychology in the Schools, 8, 29-36.

U. S. Office of Education. (August 3, 1979). Education Daily. Washington, DC: Education News Service Division, Capitol Publications, Inc.

Webster's New Collegiate Dictionary. (1953). Springfield, MA: G & C Merriam Company.

Wedell, K. (1980). Early identification and compensatory interaction. In R. Knight & D. Bakker (Eds.), Treatment of hyperactive and learning disordered children: Current Research. Baltimore: University Press.

Weiner, P. S., & Wepman, J. M. (1971). The relationship of early perceptual level functioning to later school achievements in black disadvantaged children. (Division of Research Demonstration Grants, SRS, Department of HEW, Washington, DC).

Wendt, R. N. (1978). Kindergarten entrance assessment: Is it worth the effort? Psychology in the Schools, 15, 56-62.

Wolff, M., & Stein, A. (1967). Six months later: A comparison of children who had headstart with their classmates in kindergarten. New York: Ferkauf Graduate School of Education, Yeshiva University.

APPENDIX A

Parental Comments Regarding Decision to Keep Child Back One Year

(Mother) We had come to the decision not to start her even before the screening. We are so happy that we waited. Another year serves to establish the sense of self and bonding that she needed with us, her parents. Her emotional makeup is closer now to her intellectual. She seems to be a more integrated person. I have a bright fourth grader. Wish I had kept him out a year. He would have been better off. I have eight children so I have learned alot.

(Mother) We are really glad that we made that decision. It was a tough decision at the time. She just wasn't ready to settle down.

(Father) Generally, we are very happy. He now is in the top reading group. If he ever got sick, it might be a problem as he would be so old. Don't think I have to worry about that so much, however.

(Mother) Well, we did put him in kindergarten and then he had to repeat. We felt this was best rather than have him go to preschool another year. He's doing real well now. I think we made the right decision.

(Father) I feel that she is doing excellent now. We started her in kindergarten. She then went on to the first grade and then

repeated first grade. Her confidence seems to be better now. Her last report card was good and her teacher gave us a good report as well.

(Mother) We sent him on and then he repeated kindergarten. Feel good about it now. He is doing so much better except in reading, he still has a problem. We put him on ritalin and took him off all sugar. He is calmer now. We did this in the last three weeks but the teacher said that she can also notice a difference.

(Mother) Yes, we are happy to have held her back. She is doing quite well now.

(Mother) Very glad. Doing super. She has taken off in all of her subjects. If we hadn't held her back she would have really been struggling. I'm now facing this with another younger one. I think I'll keep her back too. She has a September 30 birthday.

(Mother) I think it was the best thing. He's doing great. He just checked out a book that is 120 pages long. Now he enjoys school and is quite a bit above the others. He is way ahead in math. His peer group is best now. It was the best decision. I sent him to preschool where he learned to play and share.

(Mother) Doing very well. Best thing I could have done. Had I sent her she would be lost now. She did go to preschool, then went to kindergarten for help in speech and language. Without the screening I wouldn't have known she needed special help.

(Mother) After the screening, we had him tested by the public schools. They recommended that he start. His kindergarten teacher worked with him. At the end of kindergarten, it was decided that he needed another year of kindergarten so we sent him to his original school. He is doing fine now. Is not sensitive about repeating because his older brother did too. There is no problem because of his personality. His reading skills are very poor.

(Mother) Her schooling is very easy now. Reading so good. Really glad she was held back. Would have had trouble confidence wise, I believe. Now not have to prod. She is more joyful and confident. I'm sure it was the right decision. I have an eighth grader now who is academically at the top. She is in speech contests and Wits class. But sports and physical strength are her main disadvantage. She had a hard time in first grade in the beginning but now is doing fine. She is challenged enough.

(Mother) We're real happy. She made a big turn around since the beginning of the year. Now she is taking school seriously. We have no regrets.

APPENDIX B

Parental Comments Regarding Decision to Send Their Child On Even Though the Recommendation was to Hold Them Back

(Father) I have mixed emotions about it. Funny you should ask. He had been enrolled in a preschool for two years. They recommended that he go on to kindergarten since he was academically ready. His problems now seem to be ones of immaturity. His birthday is in June. He is holding his own, I guess. I think he is in the middle group. If I had the decision to do over again, I would probably do the same.

(Father) She's doing all right.

(Mother) Doing real well. If I had to do it over again, I'd wait a year. Her birthday is in late August. She would have been better waiting a year. I've heard more people say they are happy when they decide to wait if they are in doubt. I have another little girl coming up this year. We will wait, we've learned a lesson. We help her an awful lot at home. We study her words every night.

(Mother) Doing well. Behaviorally and academically. First grade and kindergarten have both gone well. Her birthdate is October 15. After having had her teachers realize she was so young, she did better. I have a younger child whose birthday is in December. I wonder if they will let me start her.

(Mother) No reservations. In first grade she had trouble with blending. Her teacher and I gave her lots of extra help. We decided to send her to kindergarten summer school so that she wouldn't forget her skills. Now she is in the middle group in reading. She likes to read. The doctor checked her and said to start.

(Mother) As a whole pretty good. First grade was ok. Cursive and math the last couple of weeks have been causing her problems. She has been sick and missed some school. We have no regrets.