1-1-1990

Predicting the educational achievement of preschool and kindergarten children from the cognitive subtests of early screening profiles.

Mary-Elizabeth Cohn
University of Massachusetts Amherst

PREDICTING THE EDUCATIONAL ACHIEVEMENT OF PRESCHOOL AND KINDERGARTEN CHILDREN FROM THE COGNITIVE SUBTESTS OF EARLY SCREENING PROFILES

A Dissertation Presented

by

MARY-ELIZABETH COHN

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF EDUCATION

February, 1990

School of Education

PREDICTING THE EDUCATIONAL ACHIEVEMENT OF PRESCHOOL AND KINDERGARTEN CHILDREN FROM THE COGNITIVE SUBTESTS OF EARLY SCREENING PROFILES

A Dissertation Presented

by

MARY-ELIZABETH COHN

Approved as to style and content by:

Ena Vazquez Nuttall

Ena Vazquez-Nuttall, Chairperson of Committee

Marla R Brassard

Marla R. Brassard, Member

Nylda L. Glickman

Nylda L. Glickman, Member

Marilyn E. Haring

Marilyn Haring-Hidore, Dean
School of Education

ACKNOWLEDGMENTS

I wish to thank my advisor and the chairperson of my dissertation committee, Dr. Ena Vazquez Nuttall. It has been a pleasure working with you, Ena. I also want to thank the members of my committee, Dr. Marla R. Brassard and Dr. Nylda L. Glickman, for their time, effort, and suggestions.

Special thanks is extended to American Guidance Service for their support: to Dr. Gary Robertson, to Gary Tinney, and to to all the people there who helped make this research possible. I am especially grateful to Dr. George McCloskey, the Associate Director of Test Development at AGS, who served as an outside consultant for this project. Thank you, George, for your help and your suggestions, always graciously given.

Appreciation is extended to Carol Armstrong, Nancy DiNardo, Dr. Lester Dubnick, and Mary Ann O'Rourke. Thank you for helping me locate research participants.

A special thank you to my family and friends for their interest and encouragement.

This work is dedicated to my husband, Robert Cohn, for his love, encouragement, and patience. Thank you, Harry!

ABSTRACT

PREDICTING THE EDUCATIONAL ACHIEVEMENT OF PRESCHOOL AND KINDERGARTEN CHILDREN FROM THE COGNITIVE SUBTESTS OF EARLY SCREENING PROFILES

FEBRUARY, 1990

MARY-ELIZABETH COHN, B.A., DIOCESAN TEACHERS COLLEGE

M.A., FAIRFIELD UNIVERSITY

Ed.D., UNIVERSITY OF MASSACHUSETTS

Directed by : Dr. Ena Vazquez-Nuttall

The purpose of the study was to collect predictive validity data on the cognitive subtests and composite of Early Screening Profiles, a screening instrument that will be published in 1990. Data collection involved 135 children, ages 3-6 through 6-11. The scores on Early Screening Profiles were compared to scores on the Achievement Scale of the Kaufman Assessment Battery for Children (K-ABC), the Peabody Picture Vocabulary Test-Revised (PPVT-R), and, for the 85 childrn in kindergarten or grade one at the time of follow-up testing, a teacher rating scale, Teacher Rating of Academic Performance (TRAP). Time between testing ranged from 5-1/2 to 8 months.

For the population studied, statistically significant, strong correlations of .75, .73, and .70 were found between the composite of Early Screening Profiles

and K-ABC Achievement, PPVT-R, and TRAP ($p<.01$). Strong or moderate correlations, all significant at the .01 level, resulted when Early Screening Profiles cognitive subtests were compared to criterion subtests. High agreement rates were found for standard scores of one standard deviation above the mean (82%) and one standard deviation below the mean (84%). Comparison of the Early Screening Profiles cognitive composite score with the total scores of all three criterion measures yielded average specificity and sensitivity rates of .80 and .74, respectively, for scores of 115 or higher. For scores of 85 or lower, the average specificity was high (.97) and the average sensitivity rate was modest (.32). No significant differences emerged based on sex. The older group of children scored higher than the younger on the K-ABC Achievement Scale.

Research results indicate that the cognitive subtests and composite of Early Screening Profiles show promise of becoming useful and valid additions to the field of early childhood screening.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Schools are responsible for making program decisions regarding children. To the extent possible, decisions are tailored to the needs of individual children. In order to make these decisions, children are often given nationally normed, standardized tests which, by design, compare children to others of their own age and grade placement. In the case of young children entering school for the first time, test scores, coupled with observation of behavior during testing, information from the parent, and the child's developmental history, often provide school personnel with the information they need to make appropriate and informed initial decisions about groups of children and about individual children within a given group.

Program and placement decisions are so powerful that they may impact a child's entire life. These decisions cannot be made casually. When decisions are based, at least partly, on test scores, the tests must be valid and reliable. While reliability refers to the dependability of the score a child obtains, validity, "the most important consideration in test evaluation" (American Psychological Association, p. 5), concerns "what the test measures and how well it does so" (Anastasi, 1988, p. 139). Predictive validity refers to the ability of

an instrument to predict over a time interval. If a preschool age child is given an early childhood screening test, the purpose of which is to predict the future success of that child in the school setting, it follows that the test must have good predictive validity. The test needs to be a valid measure of the child's future school success. Indeed, Standard 1.1 of the APA Standards (1985) states the need to present evidence of validity "for the major types of inferences for which the use of a test is recommended" (p. 13). As Satz and Fletcher (1979) indicated, unnecessary risk to the individual child is caused if the predictive utility of the early detection device is inadequately assessed. Unfortunately, a number of currently available screening instruments do not contain predictive validity data in their manuals.

The Research Problem

Early Screening Profiles is a newly developed early childhood screening instrument slated to be published in early 1990. National standardization has been conducted by the instrument's publisher, American Guidance Service. At publication, the test manual will provide age based standard score norms and appropriate related derived scores. The domains that are measured by Early Screening Profiles are cognition, motor development, and adaptive behavior.

Standardization, of itself, cannot determine the predictive validity of Early Screening Profiles, nor is it meant to determine the predictive validity of any one or several of the instrument's subtests. Predictive validity studies have to be conducted apart from the standardization, using either children whose results are included in the standardization, or a separate population of children. These children must be first evaluated using Early Screening Profiles and tested again, at a later date, using instruments which have previously demonstrated validity or using other measures such as teacher reports or peer ratings.

Purpose and Significance of the Study

The purpose of this research is to conduct a short term predictive validity study of the cognitive subtests of Early Screening Profiles in order to assess the ability of the instrument to predict school success in the area of cognition.

This research has three specific hypotheses relating to the predictive validity of Early Screening Profiles.

Hypothesis 1. Children's scores on the Cognitive Profile and the subtests of the Cognitive Profile of Early Screening Profiles will demonstrate a predictable, positive relationship with scores on the criterion measures.

Hypothesis 2. Correlations between the subtest and composite scores of the Cognitive Profile and scores on the criterion measures administered five-and-a-half to eight months later will show no difference due to sex.

Hypothesis 3. Correlations between the subtest and composite scores of the Cognitive Profile and scores on the criterion measures administered five-and-a-half to eight months later will demonstrate no difference in score pattern due to age. The score pattern for children between the ages of 3-6 and 4-11 will be similar to the score pattern for children between the ages of 5-0 and 6-11.

As was previously stated, test users need to have access to instruments with proven predictive validity. Users typically look for information on predictive validity in the test manual. They are frequently disappointed because such information is not always available when a test is first published. Results of this research, conducted prior to test publication, will appear in the manual of Early Screening Profiles. The research, then, has immediate significance to both the publisher of the test and to potential users of Early Screening Profiles.

Description of the Remaining Chapters

Chapter I has talked about the need for early childhood screening instruments to have good predictive

value since they can strongly impact decisions regarding young children. Additionally, this chapter has briefly described a new, yet-to-be-published early childhood screening instrument, Early Screening Profiles. The objectives of the research relative to Early Screening Profiles were described in terms of the hypotheses on which this study is based. The significance of the study for the publisher and for the future user of Early Screening Profiles was stated.

Chapter II provides an overview of early childhood screening: its history, content, purpose, value, and general characteristics. There is a short description of predictive validity. Seven early childhood screening instruments are briefly overviewed, particularly in terms of the predictive validity characteristics described in their manuals.

Chapter III presents the specific questions on which this research is based. Limitations of the research are presented. The instruments used in the research are described. The characteristics of the sample population participating in this research are given. The specific methods to be used for treatment of the data are discussed.

Chapter IV presents the research data, reports on the statistical analyses of the data, and draws conclu- related to the research questions.

Chapter V summarizes the data from this research, draws conclusions related to the findings, and makes some suggestions for future research.

CHAPTER II

REVIEW OF THE LITERATURE

This chapter begins by presenting the reason for early childhood screening and giving an overview of the characteristics of early childhood screening instruments. Since the results of this research will be included in a test manual, the chapter then discusses predictive validity through an examination and evaluation of the manuals of seven commercially available early childhood screening instruments.

Reason for Early Childhood Screening

Kindergartens in this country were introduced into the public schools in St. Louis in 1873 as a social service to the poor (Educational Resource Service, Inc., 1986). They have since come to be accepted as an integral part of most elementary school programs.

Screening, especially of kindergarten entrants, has become important in this country, given particular impetus by the passage of Public Law 94-142: The Education for All Handicapped Children Act of 1975 (U. S. Code, 1975). In line with its goal of helping the states to educate handicapped children beginningat age three, PL 94-142 focused attention on the early identification of high risk children by mandating that the schools, beginning in September, 1978, must identify children with potential learning problems.

Hence the importance of screening instruments used to determine whether a child may or may not be at risk of doing poorly in school.

PL 98-199 (U.S. Code, 1983) reaffirmed PL 94-142 by recognizing the necessity of early diagnosis followed by appropriate instruction. In 1986, Title II of PL 99-457 (U.S. Code, 1986) made federal grants available to states providing special education and related services to handicapped children ages three to five (Sec. 619 (C) amended).

Although screening instruments may be used at any grade during the child's school career and thereafter, most are geared toward the young child new to the school environment, the child on whom no other academic data has been collected. Early childhood screening can be seen as the child's introduction to school. Early childhood screening instruments are usually administered before the child begins kindergarten and are designed to briefly assess abilities associated with school success (Meisels, 1987).

Early Childhood Screening Instruments

An early childhood screening instrument usually looks at one or several performance areas. Areas most frequently assessed in the screening of young children within the schools include, but are not limited to, cognition, expressive and receptive language, motor

development, and articulation (University of the State of New York, 1982). These important areas are generally assessed by regular and special education personnel. Examiner or parent reports and observations on the child's social and emotional maturity and developmental history are often solicited as part of the screening process. Separate gross screenings of vision, hearing, and physical health are frequently conducted by medical doctors, whereas the other important areas of focus are generally covered by regular and special education personnel (Ysseldyke, Thurlow, O'Sullivan, & Bursaw, 1985).

The purpose of screening is to find those children who might be expected to have problems in school or who may be gifted (The University of the State of New York, 1982). As opposed to a readiness test, which measures learned accomplishments, a screening instrument assesses the child's ability to acquire skills (Meisels, Wiske, & Tivnan, 1984; Meisels, 1987). Clearly, this knowledge, of itself, does nothing unless the predictions made about the child afford better educational opportunities (Wilson & Reichmuth, 1984). Close teacher observation of those children identified by screening as being at risk of having learning difficulties is needed. This assessment can be an informal record keeping system such as a

checklist with consistent and convenient procedures (Board of Education, City of New York, 1983).

When observation verifies screening results, or to obtain verification of screening results prior to lengthy observation, further assessment is necessary. This often involves a full educational evaluation and typically includes the administration of a battery of relevant tests, usually including both achievement tests and clinical instruments. Additional testing helps to clarify the child's diagnosis.

The value of screening lies not only in finding children who, for a variety of reasons, may either do poorly or may perform at a higher than average level in a formal instructional program. It also demands that something be done for the at risk child in order to maximize learning potential. Although the primary purpose of screening is to find children who may be at risk, screening results can be used to design suitable educational programs. If screening indicates that ten children seem to have difficulty with motor skills, the teachers need to be flexible and know how to adjust their curricula to fit the current needs of these children to bring their skill development to normal levels (Board of Education of the City of New York, 1986).

The need to use test results to help in designing instruction for the child is addressed in Section 300.13,

Related Services, of PL 94-142. In PL 94-142, "related services" means the support services necessary to help children benefit from special education. These related services include "Preventing, through early intervention, initial or further impairment or loss of function" (Section 300.13(iii)).

One thing to be very aware of in the interpretation and use of screening instruments is that their results are not infallible. Therefore, screening test results cannot be used as the sole criterion for deciding that a child may have academic difficulties. This could easily result in inappropriate placement or in a self-fulfilling prophecy. There is no substitute for close teacher and parent observation to confirm, or to contradict, screening results.

To guard against screening giving inaccurate results, care must be taken in the selection of a screening instrument. Major characteristics to be considered in the selection of an appropriate screening instrument include technical adequacy, a national standardization, recently developed norms, curriculum or program relevance, and practical considerations.

Practical considerations include factors such as space, time, and personnel requirements. A screening instrument should be relatively inexpensive as determined by the school or agency's varying constraints of budget,

pupil/teacher ratio, staffing strengths, time, and space. A screener should take only a short time to administer (20 to 30 minutes), be easy to administer, and be capable of administration and scoring by non-clinical educational personnel.

The selected instrument should be in accord with district or agency goals for the education and the development of young children. It should be practical and lead to an intervention program which the district, school, or agency can feasibly implement and follow up with curriculum adjustments on an individual or group basis if necessary. A screening instrument is best if it is multi-dimensional, assessing several areas of a child's development.

A screening instrument that is standardized should have a national standardization that reflects the diversity found in the population as a whole. If a screening instrument is not norm referenced, it should at least have a strong national field testing built into its development. The population on which the usefulness of the instrument is verified should match the national population in a number of areas usually including, minimally, geographic region, age, race, size of school district, and socio-economic status (SES). Tests whose norming populations are limited, for example, to one area of the country or to one race or SES group have results

which cannot accurately reflect how children outside the norming group may perform.

The size of the standardization population and the age of the norms are important to consider. Salvia and Ysseldyke (1988) recommend a minimum of one hundred children per age or grade.

An important quality of a screening instrument is its technical adequacy. Even if not nationally standardized, a screening test should have good psychometric properties such as sound reliability and validity estimates. Since this research relates to predictive validity in relation to a specific early childhood screening instrument, Early Screening Profiles, the property of predictive validity is the focus of the next section.

Predictive Validity

A test has predictive validity when its scores accurately indicate what a child's score will be on a criterion measure administered at a future date. Since the goal of a screening test is prediction of school success, predictive validity is a very important psychometric quality of an early childhood screening test.

Implicit in prediction are the concepts of sensitivity and specificity, that is, how sensitive and specific an instrument is in making a prediction about

the future. An instrument with high sensitivity correctly identifies at risk children; one with high specificity correctly identifies children who are not at risk. Sensitivity and specificity relate to both false negatives and false positives. False negatives are incorrect exclusions of children from an at risk group; false positives are incorrect inclusions of children in an at risk group (Salvia & Ysseldyke, 1988). A screening instrument with high specificity and sensitivity predicts accurately for the majority of children tested. As Maloney and Ward (1976) have noted, however, there is no such thing as a 100 percent sure indicator. Which is the most important error to avoid, false positives or false negatives, is a value judgment (American Psychological Association, 1985).

Predictive Validity of Seven Screeners

The user of an early childhood screening instrument needs to know how well the instrument is able to predict the future school success of the children being tested. Since examiners typically look for validity information in the publication manual accompanying the test, and since the results of this research will appear in the publication manual of Early Screening Profiles, the following section will examine the manuals of seven commercially available early childhood screeners,

focusing specifically on information relating to predictive validity.

Boehm Test of Basic Concepts-Revised

The Boehm Test of Basic Concepts-Revised (Boehm-R), developed by Ann E. Boehm, and published in 1983, is a group administered instrument available in two alternate forms. Both forms assess a kindergarten, grade 1, or grade 2 child's mastery of fifty basic concepts such as top, last, several, half, and fewest: concepts dealing, for the most part, with time, quantity, and space. These concepts are considered by the author to be essential for successful early school achievement (Boehm, 1986).

The primary goal of the Boehm-R is instructional screening through the identification of individual children whose concept mastery level is low and through the identification of specific concepts with which the group as a whole may be unfamiliar (Boehm, 1986, p. 2).

The Boehm-R manual contains correlation coefficients showing predictive validity based on a comparison of Boehm-R test scores and scores on other achievement measures administered a year later. Criterion measures were the Comprehensive Tests of Basic Skills, reading level attained in the Bookmark Reading Program, the California Achievement Tests, and the Iowa Tests of Basic Skills. The seventeen correlation coefficients have a median of .44 and range from .28 to .64 which, the

manual states, are supportive of the use of the Boehm-R as a screening measure (Boehm, 1986, p. 59).

Denver Developmental Screening Test-Revised

The Denver Developmental Screening Test-Revised (DDST-R) was published in 1975 (with ancillary materials copyright 1981) and authored by William K. Frankenburg, Josiah B. Dodds, Alma W. Fandal, Elynor Kazuk, and Marlin Cohrs. A multi-dimensional screening instrument, it provides information in several developmental areas on children from birth through age 5.

The goals of the DDST-R are to screen asymptomatic children for possible problems, to confirm intuitive suspicions with an objective measure, and to monitor high risk children (Frankenburg, Dodds, Fandal, Kazuk, & Cohrs, 1975, p. 1).

The manual reports a high degree of agreement between the original DDST ratings of 236 children and the quotients of the Stanford-Binet and Bayley Scales-Revised (Frankenburg, et al., 1975). Time between testings is not given, so it is impossible to tell if these results relate to concurrent or predictive criterion validity.

Developmental Indicators for the Assessment of Learning-Revised

Normed for ages two through five, the Developmental Indicators for the Assessment of Learning-Revised (DIAL-R) was developed by Dr. Carol D. Mardell-Czudnowski

and Dr. Dorothea S. Goldenberg for screening in the three domains of motor, language, and concepts.

As stated in the manual, the primary goal of DIAL-R is to satisfy the obvious and continued need for an adequately standardized, valid, and reliable measure of early motoric, conceptual, and language development (Mardell-Czudnowski & Goldenberg, 1983, p. 63).

The manual reports a predictive validity study conducted on the original DIAL by Hall, Mardell, Goldenberg, and Wick in 1976. Two years after original testing, 249 children from the DIAL standardization were tested on the Metropolitan Reading Readiness Test and one of either the Iowa Test of Basic Skills, the Metropolitan Achievement Test, or the Stanford Achievement Test. A teacher rating scale was also used as a criterion measure. Multiple correlations ranged from .45 to .73; all were significant (Mardell-Czudnowski & Goldenberg, 1983). The sensitivity and specificity of DIAL-R compared to Stanford-Binet is presented as evidence of concurrent validity.

Early Screening Inventory

Early Screening Inventory (ESI), a developmental measure, was written by Samuel J. Meisels and Martha S. Wiske and published in 1983. Normed for ages four through five, its three main sections test visual-motor

coordination, gross motor/body awareness, and language/ cognition.

The goal of ESI is to identify children who may need special education services in order to perform adequately in school; it is meant to be one phase of a complete screening process (Meisels & Wiske, 1988, p. 1).

To investigate short term predictive validity, 472 randomly selected children screened on ESI prior to the start of the kindergarten year, were tested on the 1976 Metropolitan Readiness Test seven to twelve months later, at the end of their kindergarten year. This study found agreement between the two instruments for 391 children (83%). Results showed 44 children scoring high on screening but low on Metropolitan Readiness and 38 children who scored poorly on ESI but later did well on the Metropolitan Readiness Test. Based on this data, the manual states that these results indicate that ESI is a good predictor of reading readiness at the end of kindergarten (Meisels & Wiske, 1988).

A long term predictive validity study is reported for 115 children who were administered the ESI prior to their kindergarten year. Criterion measures included parent questionnaires, other screening results, and school records through grade four. This study showed that the ESI correctly classified between 64% and 79% of the children (Meisels, et al., 1984).

McCarthy Screening Test

The McCarthy Screening Test (MST) is an adaptation of the McCarthy Scales of Children's Abilities authored by Dorothea McCarthy. Published after Dr. McCarthy's death, the MST responded to a perceived need for a non-clinical screening instrument. Its eighteen subtests measure facility with language and concepts, visual perception, auditory memory, fine and gross motor coordination, and orientation in space.

The primary goal of the MST was the development of a large scale, non-specific screener to identify children with learning disabilities and other kinds of handicaps (Psychological Corporation, 1978, p. iii). The manual further states that the MST is able to screen out quickly, for further assessment or diagnosis, children whose low performance renders them at risk with respect to probable school success (Psychological Corporation, 1978, p. 9).

Since the MST is an adaptation of the McCarthy Scales, the validity discussion in the MST manual is, for the most part, linked to validation studies of the McCarthy Scales. However, predictive validity of the MST was calculated on 52 children using the Metropolitan Readiness Tests (MRT) as the basis for comparison. One year elapsed between testings. Moderately strong, statistically significant correlations, ranging

between .31 and .57, were found between the Pre-Reading Skills Composite and the Quantitative Skill Area of the MRT and Verbal Memory, Draw-a-Design, Numerical Memory, and Conceptual Grouping on the MST (Psychological Corporation, p. 12).

Miller Assessment for Preschoolers

As stated by its author, Lucy Miller, the Miller Assessment for Preschoolers (MAP) is a short but comprehensive screening tool with a broad range of items, designed to be sensitive to moderate as well as severe developmental delays (Miller, 1982). The MAP consists of twenty-seven core items designed to assess sensory and motor abilities, cognitive abilities, and combined abilities.

The two MAP goals are to provide a statistically sound screening tool useful in the identification of children in need of further evaluation and to provide a clinical framework helpful in defining a child's strengths and weaknesses (Miller, 1982, p. xiv).

Four years after the 1980 standardization, 338 children who had participated in the initial screening took part in follow-up testing to help establish predictive validity of MAP. Criterion instruments were the Wechsler Intelligence Scale for Children-Revised (WISC-R), Woodcock-Johnson Psychoeducational Battery, Bruininks-Oseretsky Test of Motor Proficiency, and the

Developmental Test of Visual Motor Integration (VMI). Another set of criterion measures included school retention, teacher observations, placement in self-contained special education classes, and report card grades in language, reading, and math. The manual reports that all correlations were significant at the .001 level with particularly high correlations between MAP Total Score and WISC-R scores, between MAP scores and the Woodcock-Johnson Math, Reading, and Language scores, and between MAP and the Bruininks (Miller, 1988, p. 115).

Minnesota Child Development Inventory

The Minnesota Child Development Inventory (MCDI) was written by Harold Ireton and Edward Thwing and published in 1974. It consists of 320 items on eight developmental scales of general development, comprehension-conceptual, situation comprehension, self help, and personal-social. It is in questionnaire format and is completed by the parent, most commonly by the child's mother.

The goals of the MCDI are to use the mother's observations to measure the present development of her child on a standardized scale and to serve as a good supplemental source of information in the identification of a child whose development is below age expectation (Ireton & Thwing, 1974, pp. 1, 3).

Predictive validity is not addressed in the manual. In line with its goal of identification as opposed to prediction, validity is presented in terms of age discrimination. This is both appropriate and useful informa information, especially when MCDI is used as it should be, as a corollary to a screening instrument administered to the child.

Evaluation of Seven Screeners

Of the seven instruments presented in the above section, only one, Boehm-R, is a group administered test, and, for that reason, has a larger standardization sample than individually administered tests. The predictive validity information reported in the Boehm-R manual refers both to studies conducted on the original Boehm Test of Basic Concepts and the Boehm-R. The manual reports predictive validity data collected on more than one thousand children from three school districts who participated in the standardization. Criterion testing was done a year after the administration of the Boehm-R. Other than stating that the subjects were part of the standardization sample, the manual does not report where these children lived, the sex of the children, the size of their school districts, their racial and ethnic characteristics, or their SES. Consequently, it is difficult to determine whether the Boehm-R is useful for any particular group of children. Although it is useful

to have the predictive validity information provided in the manual, a more detailed description of the sample population would have made the information more valuable.

Of the six other instruments, the predictive validity information provided in the manual does not always apply to the instrument itself. In the case of the MST, the predictive validity of its parent, the McCarthy Scales of Children's Abilities, is cited, based on the performance of fifty-two public school children in the northeast. As with the Boehm-R, no information is provided on the characteristics of the 52 children, other than that they lived in the northeast and were between the ages of 5 and 6.

The MST is an unusual case, since all of the MST items appear also on the McCarthy Scales. The assumption is made that children would score the same way on both instruments. This may or may not be a valid assumption; no empirical data was found to support the premise directly. Results of a study by Naglieri and Harrison (1982) estimated that the General Cognitive Index of the Kaufman Short Form, another abbreviated form of the Scales, was virtually identical to the McCarthy General Screening Index of the Scales. The fact that the MST norms are more than twelve years old gives the instrument questionable validity for today's children.

The manuals of the DDST-R and the DIAL-R report predictive validity information for their parent instruments, the DDST and the DIAL.

Studies of the DDST do not point to it as being an instrument with high predictive validity when used to screen for early school performance. Post publication research reviews have raised questions about the DDST's ability to identify educationally at risk children (Ireton, 1988). A year after publication of the DDST-R, Nugent (1976) found the DDST-R to be relatively inefficient in the detection of preschool children with IQs below 70. Meisels (1989), citing a number of concurrent and predictive validity studies, states that there is evidence that the DDST overlooks numbers of children at risk for developmental problems. More than two thousand children participated in a 1980s predictive validity study of DDST conducted in Canada by Cadman and others. The results showed the DDST to have only modest predictive validity (Cadman, Chambers, Walter, Feldman, Smith, & Ferguson, 1984).

In regard to DIAL, as with Boehm-R and MST, the characteristics of the predictive validity population are not given; further, it is not even possible to tell if the information, based on the original DIAL, not DIAL-R, applies to concurrent or predictive validity. Salvia and

Ysseldyke (1988) conclude that the validity of DIAL-R is not clearly established.

In spite of their shortcomings, DDST and DIAL were the most frequently used screening instruments, used by greater than fifty per cent of respondents reported in the University of Minnesota Research Report #2 (Ysseldyke, et al., 1985). Lichtenstein and Ireton (1984) report that the DDST has been widely used in special education early identification programs. These findings point to the need not only for the development of screening instruments with good predictive validity, but also imply a need for education of the test user.

The publication manual of the MAP acknowledges that predictive validity information is not yet available (Miller, 1982). The 1988 edition of the manual provides predictive validity information on 338 children from the standardization sample tested four years later. Children change considerably over four years. Meisels (1985) states that a time lapse of two or more years can render tests with good initial predictive ability less accurate. The large number of children in the predictive validity sample, representing approximately one quarter of the standardization sample, would seem to largely offset the four year time factor. Characteristics of both the standardization and predictive validity samples are comprehensively and meticulously reported in Table 18 of

the revised manual (Miller, 1988, p. 113). Two sets of criterion measures were used, one comprised of teacher reports and class placement, and the other a set of standardized tests. As reported in the manual, the overall results of the study indicated that the MAP Total Score is a better indicator of performance than any specific MAP Performance Index Score (p. 115).

Additionally, a predictive validity study of MAP was conducted in Michigan by Lemerand (1985) on 273 children, and one was conducted in Colorado by Cohn (1986) with 134 participants. Both the researchers reported MAP to have reasonably good predictive validity characteristics.

The manual of the MCDI does not provide predictive validity information. However, a number of longitudinal studies have been conducted on the MCDI since its publication. Guerin and Gottfried report on a predictive validity study of MCDI involving 89 mothers of 2-1/2 year old children. At age five, these children were given the Kaufman Assessment Battery for Children (K-ABC), the Wechsler Intelligence Scales for Children-Revised (WISC-R), and the Wide Range Achievement Test-Revised (WRAT-R). The multiple correlations displayed in this study ranged from .45 (MCDI to K-ABC Mental Processing) to .69 (MCDI to K-ABC Achievement), all significant at the .01 level, pointing to the usefulness of the MCDI in

clinical, educational, and pediatric settings (Guerin & Gottfried, 1987).

Colligan (1982) explored the usefulness of parent reports in preschool screening as an alternative or a supplement to direct testing. Reviewing seven research studies involving MCDI with a total of 1,413 children, Colligan concluded that the parent questionnaire is a good means of obtaining useful information about possible academic problems a child may encounter in school.

In the manual of Early Screening Inventory (ESI) there is detailed information regarding short term and long term predictive validity and about the specificity and sensitivity of the instrument. Since predictive validity is of high importance in determining the usefulness of an instrument used to predict school performance, the information given in the ESI manual is both important and useful.

Summary

This chapter began with an overview of the history, content, purpose, value, and general characteristics of early childhood screening and the instruments used for early childhood screening. It then presented and discussed seven currently available early childhood screeners, particularly in light of their predictive validity characteristics.

APA Standards (1985) state that if a test is to be used for prediction, evidence of predictive validity must be presented. A look at the manuals of seven current screening instruments shows that this standard is not always followed, perhaps because the necessary data are not available at the time the test is published. When predictive validity information is presented, it does not always provide the test user with sufficient information regarding the characteristics of the predictive validity sample to give the user a sense of security in applying results to the user's target population.

The need to report predictive validity in test manuals, where it is readily accessible to the test user, is a problem that needs to be overcome if the selection of inappropriate or worthless measures is to be prevented (Lehr, Ysseldyke, & Thurlow, 1986). It is a problem of time because, by its nature, predictive validity data can be collected only over time, sometimes not before the publication of a screening instrument's manual. The research reported in this study attempts to allay that problem in the case of Early Screening Profiles by having predictive validity data collected and analyzed prior to publication of the test and its manual.

CHAPTER III

METHODS AND PROCEDURES

The primary purpose of this research was to conduct a predictive validity study of the cognitive subtests of Early Screening Profiles, a new early childhood screening instrument. Early Screening Profiles is scheduled to be published in early 1990. The results of this research will be included in the manual of Early Screening Profiles, enabling the instrument's publisher to fulfill the APA requirement of providing validity information for the major use of Early Screening Profiles, prediction of school success. Of primary consideration was the correlation between the Cognitive Profile of Early Screening Profiles and the scores on three criterion measures. Additionally, the research asked whether significant differences existed between scores on the cognitive subtests of Early Screening Profiles and the criterion instruments based on sex and on age.

Research Questions

The three hypotheses stated in Chapter I resulted in the following six research questions which were examined in this study.

Question 1. Do children's scores on the composite standard score of the Cognitive Profile of Early Screening Profiles correlate highly with children's

scores on the criterion measures? The composite standard score represents the composite of all four cognitive subtests.

Question 2. Do children's scores on the subtests of the Cognitive Profile of Early Screening Profiles correlate at least moderately with children's scores on the criterion measures?

Question 3. Do children who score significantly low on the subtests and the composite of the Cognitive Profile of Early Screening Profiles score significantly low five-and-a-half to eight months later on the criterion measures? Significantly low will be defined as one standard deviation or more below the mean: a standard score less than or equal to 85.

Question 4. Do children who score significantly high on the subtests and the composite of the Cognitive Profile of Early Screening Profiles score significantly high five-and-a-half to eight months later on the criterion measures? Significantly high will be defined as one standard deviation or more above the mean: a standard score greater than or equal to 115.

Question 5. Are there differences in the ability of Early Screening Profiles to predict criterion measure performance based on sex?

Question 6. Are there differences in the ability of Early Screening Profiles to predict criterion measure performance based on age?

Assumptions

1. Early Screening Profiles will have good psychometric qualities, including a strong national standardization and good technical adequacy in the areas of reliability and validity.

2. The content of Early Screening Profiles will be appropriate in all three of its profiles: cognitive, adaptive behavior, and motor.

3. The criterion measures have adequate predictive validity.

4. The data reported on the Parent Permission Forms is correct.

Scope and Limitations

The scope of this study was to collect predictive validity data regarding the cognitive subtests of Early Screening Profiles, a test which will be used to predict the future school performance of children ages three through seven.

The study has the following specific limitations:

1. Sample size: the sampled population consisted of 136 children attending school in Fairfield County, Connecticut, and Huntington, New York. There is no way of knowing if this is a representative sample of the population.

2. Age and grade range: tested children ranged in age from three years six months through six years five months at the time of initial testing. At the time of initial testing, these children were attending either prekindergarten or kindergarten. At criterion testing, the children were attending either prekindergarten, kindergarten, or first grade.

3. Residence within the Northeast: the research is limited to those school districts, schools, and parents who agreed to let their children participate.

4. Physical conditions: it was not possible to control the variable of physical conditions under which testing took place.

5. The research is limited by the reliability and validity of the instruments used.

Population

A total of 136 children participated in the research. Of these, 64 resided in Connecticut and 72 in New York. In Connecticut, 58 of the participants attended the Bridgeport Public Schools, the Child Care Center of Stamford, Inc., or the Greenwich Christian Day School. The remaining 6 children attended miscellaneous private and public schools in Fairfield County. In New York, all 72 of the children were enrolled in the Huntington Public Schools on Long Island.

Table 1 provides a detailed description of the characteristics of the research sample.

Table 1

Characteristics of the Research Sample by Sex and by Number and Percent of Participants

Characteristics	Boys	Girls	Total	Percent
Race				
Black	14	10	24	18%
Hispanic	9	9	18	13
White	49	45	94	69
	72	64	136	100%
Socioeconomic Status*				
elementary school only	3	4	7	5%
attended high school	2	2	4	3
high school graduate	15	14	29	21
attended college	19	7	26	19
college graduate	25	24	49	36
graduate school	5	12	17	13
information not given	3	1	4	3
	72	64	136	100%
Primary Language**				
English	50	56	116	85%
Spanish	5	6	11	8
Other	3	2	5	4
information not given	4	0	4	3
	72	64	136	100%

* Education of parent who has gone farthest in school.
** Language spoken at home.

The information for Table 1 was compiled from self reports of parents who completed the Permission Forms agreeing to the testing of their children. The languages other than English or Spanish included one each of Greek, French, Italian, German, Persian, and Polish. In each of those cases, the child tested appeared to the researcher to be age appropriately fluent in English as judged by

the child's expressive and receptive language before, during, and following testing.

Whether residents of New York or Connecticut, all tested children resided within a sixty mile radius of New York City at the time of testing.

Table 2 presents the age range of the subjects at the time of initial testing. The first category, Prekindergarten to Pre-K, for example, means that the child was in prekindergarten at the time of initial testing and still in prekindergarten at the time of criterion testing. For the purpose of this study, prekindergarten and day care attendees were grouped together and are referred to as prekindergarten children.

Table 2

Age and Grade Placement of Children
at Times of Initial and Follow-up Testing

Age and Grade	Boys	Girls	Total
Age Range at Initial Testing:			
3-6-0 through 3-11-30	4	1	5
4-0-0 through 4- 5-30	9	10	19
4-6-0 through 4-11-30	14	11	25
5-0-0 through 5- 5-30	12	5	17
5-6-0 through 5-11-30	19	24	43
6-0-0 through 6- 5-30	14	13	27
	72	64	136
Grade Placement at Initial and Follow-up Testing:			
Prekindergarten to Pre-K	24	15	39
Pre-K to Kindergarten	6	7	13
K to Transitional K	4	1	5
K to Special Ed Grade One	0	2	2
K to Grade One	38	39	77
	72	64	136

Table 2 indicates a reasonably even distribution by sex and by age, with the age range 5-6 through 5-11 containing the largest number of cases. This is reflected in grade placement. Children moving from kindergarten to grade 1 were most heavily represented.

Measures Used

Initial testing was conducted on the 136 research participants using the Cognitive Subtests of Early Screening Profiles (ESP). The criterion instruments administered five-and-one-half to eight months later were the Achievement Scale of the Kaufman Assessment Battery for Children (K-ABC), and Form L of the Peabody Picture Vocabulary Test-Revised (PPVT-R). The classroom teachers of 96 of the 136 children completed the Teacher Rating of Academic Performance (TRAP). This section describes both Early Screening Profiles and the criterion instruments.

Early Screening Profiles

Early Screening Profiles is the instrument for which predictive validity is being tested in this research. The standardization edition used in the research was authored by Alan S. Kaufman, Robert H. Bruininks, and Sara S. Sparrow, with Nadeen L. Kaufman, Patti Harrison, Steven Ilmer, John Rynder, and George McCloskey. Early Screening Profiles was standardized by American Guidance Service for use with children ages three through seven and will be published in early 1990. The instrument is

comprised of three profiles: Cognitive, Adaptive Behavior, and Motor. Only the Cognitive Profile is the subject of this research.

The Cognitive Profile of Early Screening Profiles is made up of four subtests: Verbal Concepts, Visual Discrimination, Logical Relations, and Basic School Skills. Each item on each subtest is presented to the child on an easel; one side of the easel contains the examiner's plate, the child's side of the easel contains the visual stimulus for the item. Testing of each subtest begins with sample items. Each subtest has specific starting points and discontinue rules by age. Testing on all subtests combined takes approximately twenty minutes.

Verbal Concepts is the first subtest of the ESP Cognitive Profile. It contains four item types relating to receptive vocabulary, expressive vocabulary, receptive riddles, and expressive riddles. The child is presented with a visual and/or auditory stimulus. Some items require a verbal response, others a motoric (pointing) one.

Visual Discrimination is the second subtest of the Cognitive Profile. Here the child is shown a stimulus picture and is asked to match it to the same picture within a row of different response pictures. No verbal responses are required.

Logical Relations is the third subtest of the Cognitive Profile. In the first seven items, the child is shown a stimulus picture set apart from a row of different pictures. The child is asked to find the picture in the row that goes with the stimulus picture. For the remaining items, the child is shown visual analogies with the fourth element missing. A number of possible responses are printed below the analogy; the child selects the one that best completes the analogy.

Basic School Skills is the fourth subtest of the Cognitive Profile. This subtest contains number and quantity concepts; number, letter, and word naming; and number, letter, and word recognition.

Since Early Screening Profiles is not yet published, the information regarding its technical merit is not yet available.

Kaufman Assessment Battery for Children (K-ABC)

The K-ABC was authored by Alan S. Kaufman and Nadeen L. Kaufman. Yielding age based standard score norms for children ages two and one-half through twelve, it was nationally standardized and published by American Guidance Service. Of its three scales, Sequential Processing, Simultaneous Processing, and Achievement, only the last, the Achievement Scale was used in this research. The Achievement Scale of K-ABC contains six subtests: Expressive Vocabulary, Faces and Places,

Arithmetic, Riddles, Reading Decoding, and Reading Understanding. The last of these, Reading Understanding was not used because its norms are outside the age range of the children tested.

Expressive Vocabulary is the first subtest of the K-ABC Achievement Scale. In it, the child is shown a picture and must name the picture accurately.

Faces and Places is the second subtest of the K-ABC Achievement Scale. In this subtest, the child is shown a picture of a fictitious or real person or place and must tell the examiner who or what the picture represents.

Arithmetic is the third subtest of the K-ABC Achievement Scale. Here the child is shown a picture and is asked a question regarding the picture that relates to an arithmetic concept such as one-to-one correspondence, counting, number recognition, sequencing, addition, or subtraction.

Riddles is the fourth subtest of the K-ABC Achievement Scale. In this subtest, the child is given a verbal stimulus. The child is read a three-part sentence and must name the item defined in that sentence without the aid of a visual stimulus.

Reading Decoding is the fifth subtest of the K-ABC Achievement Scale. In this subtest the child must correctly identify upper case and lower case letters and words from a visual stimulus.

Published in 1983, the K-ABC underwent a national item tryout prior to standardization. The national standardization, conducted in 1981, involved 2000 children stratified by sex, age, geographic region, race or ethnic group, community size, the educational placement of the child, and socioeconomic status. Parental educational attainment, an excellent estimate of socioeconomic status (Kaufman & Kaufman, 1983, p.65) was used for determining SES. The K-ABC Interpretive Manual (Kaufman & Kaufman, 1983) reports that reliability coefficients of internal consistency, computed using Guilford's formula, range from .93 to .96 for the Achievement Scale for children ages 3-0 through 6-11 (p. 83). Test-retest reliability coefficients for the Achievement Scale for children through age 8-11 is .95 (p. 83).

The K-ABC Interpretive Manual (Kaufman & Kaufman, 1983, p. 121) reports the results of five predictive validity studies of school age children validating K-ABC against the Peabody Individual Achievement Tests (PIAT), the California Achievement Tests (CAT), and the Iowa Test of Basic Skills (ITBS). Correlations between total score on the criterion test and the K-ABC achievement score are given. Both the mean and the median correlations of these five are .77. These five studies included a total of 151 children ranging in age from 5-5 to 12-6. Of the

151, 30 children were defined as culturally different, Navajo Indian children (r=.82), and 29 were described as educable mentally retarded (r=.67). The remaining 92 children are described as normal (r=.79).

A sixth predictive validity study reported in the K-ABC Interpretive Manual involved 31 preschool children ranging in age from 3-0 to 4-11. Children were given the Woodcock-Johnson Psycho-Educational Battery (W-J) eleven months after administration of K-ABC. Correlation between the K-ABC achievement subtests and the W-J preschool cluster was .73; correlation between K-ABC Achievement Scale and W-J Knowledge Cluster was .84.

Kamphaus and Reynolds (1987) report on two post-publication predictive validity studies of the K-ABC. The first, conducted by Murray and Bracken (1984), reported a .88 correlation between K-ABC achievement subtests and the Peabody Individual Achievement Test (PIAT) total test score. This study was conducted on 29 elementary grade children over an eleven month period.

The second predictive validity study reported by Kamphaus and Reynolds was conducted in North Carolina in 1981. The criterion instrument here was the California Achievement Test (CAT), a group administered instrument, and the time between initial and follow-up testing was six months. The correlation coefficient between the K-ABC Achievement Subtests and the CAT total score is .77

based on testings of 44 children (Childers, Durham, Bolen, & Taylor, 1985).

Both the six studies reported in the Interpretive Manual as APA Standards direct, and the two studies reported by Kamphaus and Reynolds, indicate that the K-ABC has sound psychometric qualities, including good predictive validity.

Further, a stability study was conducted by Lyon and Smith involving 53 at-risk preschool children who were administered the K-ABC twice, with nine months between testings. This study reports a stability coefficient of .82 for the two administrations of the K-ABC Achievement Scale. The study results support the concept that the K-ABC global scales, of which Achievement is one, are stable over time for preschoolers (Lyon & Smith, 1987).

Peabody Picture Vocabulary Test-Revised (PPVT-R)

Form L of the PPVT-R was also used as a criterion instrument. This test provides an age based standard score measure of receptive vocabulary for ages two and one-half through adult. It was authored by Lloyd M. Dunn and Leota M. Dunn, standardized by American Guidance Service, and published by American Guidance Service in 1981.

To administer the PPVT-R, the examiner shows the subject a series of age appropriate plates, each of which contains four pictures. The examiner gives the child a

stimulus word which is illustrataed by one of the four pictures on the plate; the child responds by pointing to, or by otherwise indicating, the picture the word represents.

The PPVT-R manual (Dunn & Dunn, 1981) reports that the standardization sample, based on 1970 census data, included 200 participants for each six month age group, ages 2-1/2 through 18. The stratification variables were age, sex, ethnicity, geographic distriction, size of community, and socioeconomic status based on the occupation of the major wage earner.

The manual reports split-half reliability coefficients for Form L for ages 3-1/2 - 6-1/2 ranging between .70 and .84 (Dunn & Dunn, 1981, p. 54). The mean for this age group is .76. The manual reports delayed test-retest coefficients for 232 children between the ages of 3-0 and 6-11 who were in the standardization. Time between testing ranged from 9 to 31 days. Alternate form reliability coefficients for standard scores for this group ranged from .58 to .77 with a mean of .70 (Dunn & Dunn, 1981, p. 56).

At the time of publication of the PPVT-R manual, predictive validity information was not available relating directly to the PPVT-R. The PPVT-R manual summarizes concurrent and predictive validity studies

of the parent instrument, the Peabody Picture Vocabulary Test (PPVT), reporting a .71 median correlation based on 55 correlations with ten criterion instruments and concluding that the PPVT correlates moderately well with verbal intelligence and most highly with other vocabulary measures (Dunn & Dunn, 1981, pp. 67, 68).

Predictive validity correlations given in the PPVT-R manual are based on 27 comparisons and range from .24 (PPVT to Wide Range Achievement Test, Reading) to .62 (California Achievement Test, Total Test) (Dunn & Dunn, 1981, p. 67). Unfortunately, The PPVT-R Technical Supplement (Robertson and Eisenberg, 1981) does not give further data on the predictive value of the PPPVT-R.

Vance, Kutsick, and West (1987) conducted a concurrent validity study of the PPVT-R, comparing scores of 51 children tested on the PPVT-R and the Expressive One-Word Picture Vocabulary Test (EOWPVT) to scores on the Wechsler Preschool and Primary Scale of Intelligence (WPPSI). They report a .71 correlation between the PPVT-R and the WPPSI Performance Score for non-language delayed children. They also report the PPVT-R standard score as being significantly lower that the WPPSI Performance and Full Scale IQs.

Fletcher and Satz (1982) conducted a seven year longitudinal study, following 195 children from kindergarten through grade 6. The subjects were given

the PPVT-R and three other instruments. It was found that the four test battery had continuing usefulness and predicted achievement outcomes through grade 6.

Altepeter (1985) used PPVT-R for the intellectual screening of 74 preschool children, ages 2-6 through 5-11, who had been referred for psychological evaluation. Comparison of the PPVT-R scores of these children with their Stanford-Binet scores yielded a correlation coefficient of .72. However, only 55% of the cases were correctly classified by PPVT-R. Tarnowski (1987) reported similar data resulting from a comparison of the PPVT-R results of 217 subjects, ages 2-0 through 15-11, with their scores on the Stanford-Binet. Although the correlation was .88, Tarnowski recommends PPVT-R be used with caution since only 98 of the cases were correctly classified by PPVT-R.

Insufficient data were found to support the predictive validity of PPVT-R for use as the sole criterion instrument in this study. In fact, research suggests that the PPVT-R as a screener should be used with caution and only as part of a comprehensive psychoeducational battery of tests (Vance, et al., 1987; Bracken, Prasse, & McCallum, 1984; Altepeter, 1985; Tarnowski, 1987). It was included as a criterion measure in this research for practical reasons. For one thing, there was sufficient time spent with each child to

allow for its administration. Additionally, it is widely used in making placement decisions (Lehr, et al, 1986).

Teacher Rating of Academic Performance (TRAP)

The third criterion instrument, used with children attending kindergarten or first grade at time of follow up testing, was a teacher rating scale, Teacher Rating of Academic Performance (TRAP) (Gresham, Reschly, & Carey, 1987). TRAP contains five questions, each of which is answered on a five point scale. Responded to by the classroom teacher, the questions relate to the performance of the child in the classroom in terms of general academics and classroom performance in reading and in mathematics.

A study involving a total of 200 children (100 learning disabled and 100 non-handicapped), ages 7-1/2 to 11-1/2, was conducted in Iowa. The study reported that TRAP accurately classified 85.7% of the non-handicapped group, and 96.2% of the learning disabled group. This yielded an overall correct classification rate of 91%: TRAP correctly classified 99 of 109 children (Gresham, et al., 1987). In addition to the children's teachers completing TRAP, children were administered the Wechsler Intelligence Scales for Children-Revised (WISC-R) and the Peabody Individual Achievement Test (PIAT). The high classification data supports the use of TRAP as a criterion instrument in this research.

Data Collection

In the Fall of 1987, a number of day care centers and public school districts in New York, Connecticut, and Massachusetts were invited to participate in the research. Those who responded positively were the Bridgeport (Connecticut) Public Schools, the Huntington (New York) Public Schools, the Child Care Center of Stamford (Connecticut), Inc., and the Greenwich (Connecticut) Christian Day School. The publisher of Early Screening Profiles offered participating schools an incentive of catalog materials or cash for each testing session for each child who was tested.

Once a school had agreed to participate in the study, parents of prekindergarten and kindergarten children in those schools were sent a letter explaining the project and a permission slip. Consenting parents completed permission slips on each child. The permission slip requested information from the parent regarding race, primary language spoken in the home, and education levels of the parents.

Initial testing, using the Cognitive Subtests of the standardization edition of Early Screening Profiles, was conducted in late Fall, 1987, and during the Winter and Spring of 1988. Follow up testing on the criterion instruments was conducted five-and-a-half to eight months later. The criterion instruments for all children were

the appropriate subtests of the Achievement Scale of the Kaufman Assessment Battery for Children (K-ABC), and Peabody Picture Vocabulary Test-Revised (PPVT-R), Form L. Teachers of children who were in a public school first grade or kindergarten setting at the time of follow-up testing were asked to complete the Teacher Rating of Academic Performance (TRAP). TRAP data was collected on 96 public school kindergarten and first grade children. All testing was completed by mid-December, 1988.

Of the 136 children who comprised the predictive validity study, 119 were tested by the researcher. The remaining 17 were tested by a certified school psychologist employed by the Bridgeport Public Schools.

Table 3 shows the time lapse between initial and follow-up testing for each age group in the sample.

Table 3

Time Lapse between Initial and Follow-up Testing by Age

Age Range/Initial Testing	Number of Children	Number of Months between Testing 5-1/2	6	7	8
3-6-0 through 3-11-30	5	0	4	1	0
4-0-0 through 4- 5-30	19	3	14	1	1
4-6-0 through 4-11-30	25	6	18	0	1
5-0-0 through 5- 5-30	17	0	5	10	2
5-6-0 through 5-11-30	43	0	6	33	4
6-0-0 through 6- 5-30	27	0	5	22	0
	136	9	52	67	8

Table 3 shows that the time lapse between testing on Early Screening Profiles and the criterion instruments ranged from 5-1/2 to 8 months. It is important to note that, for 119 of the 136 participants, the time between testing was either 6 or 7 months.

Data Analysis

The purpose of the research is to establish the predictive validity of Early Screening Profiles. This section relates back to the research questions posed at the start of this chapter and describes the specific ways in which the data was analyzed.

Question 1. Do children's scores on the composite standard score of the Cognitive Profile of Early Screening Profiles correlate highly with children's scores on the criterion measures? The composite standard score represents the composite of all of the four cognitive subtests.

Using the SPSS-X Pearson CORR Program (SPSS, Inc., 1983), Pearson product-moment correlation coefficients were computed between the composite of the Cognitive Profile and children's scores on the criterion instruments: the K-ABC Achievement Scale subtests and composite, the PPVT-R, and TRAP. Each correlation coefficient was tested by the SPSS-X program to determine if the degree of relationship between the Cognitive Profile subtest or composite score and the criterion measure

significantly differed from chance level. Statistical significance was tested at the .01 level.

Question 2. Do children's scores on the subtests of the Cognitive Profile of Early Screening Profiles correlate at least moderately with children's scores on the criterion measures?

Using the SPSS-X Pearson CORR Program (SPSS, Inc., 1983), Pearson product-moment correlation coefficients were computed between each subtest of the Cognitive Profile and children's scores on the following criterion instruments: K-ABC Achievement Scale subtests and composite, PPVT-R, and TRAP. Each coefficient of correlation was tested by the SPSS-X program to determine if the degree of relationship between the Cognitive Profile subtest score and the criterion measure significantly differed from chance level. Statistical significance was tested at the .01 level.

Question 3. Do children who score significantly low on the subtests and the composite of the Cognitive Profile of Early Screening Profiles score significantly low five-and-a-half to eight months later on the criterion measures? Significantly low was defined as one standard deviation or more below the mean: a standard score less than or equal to 85.

Question 4. Do children who score significantly high on the subtests and the composite of the Cognitive

Profile of Early Screening Profiles score significantly high five-and-a-half to eight months later on the criterion measures? Significantly high was defined as one standard deviation or more above the mean: a standard score greater than or equal to 115.

Using SPSS-X Crosstabs Program, 2 x 2 contingency tables indicating the relationship between Cognitive Profile subtests and composite scores and scores on the criterion measures were generated to respond to both questions 3 and 4. An index of agreement was obtained for each table by dividing the number of cases listed in cells 1 and 4 by the total number of cases. The percentage of overreferrals and underreferrals was then established and tabled as were the rates of sensitivity and specificity.

Question 5. Are there differences in the ability of Early Screening Profiles to predict criterion measure performance based on sex?

Using SPSS-X Regression Program, correlations between each Cognitive Profile score and each criterion measure for females and for males were plotted and fitted to a regression line. The slope of the regression line for males was compared with the slope of the regression line for females using a procedure described by Neter and Wasserman (1974). The statistical significance of the difference between the slopes of the female and male

regression lines of each score comparison was tested at the .01 level.

Question 6. Are there differences in the ability of Early Screening Profiles to predict criterion measure performance based on age?

Using SPSS-X Regression Program, correlations between each Cognitive Profile score and each criterion measure for children ages 3-6 through 4-11 and for children ages 5-0 through 6-11 were plotted and fitted to a regression line. The slope of the regression line for children ages 3-6 through 4-11 was compared to the slope of the regression line for children 5-0 through 6-11 using a procedure described by Neter and Wasserman (1974). The statistical significance of the difference between the slopes of the regression lines for the younger and the older groups for each score comparison was tested at the .01 level.

Significance

Standard 1.1 of the American Psychological Association's Technical Standards for Educational and Psychological Testing (1985) states that "evidence of validity should be presented for the major types of inferences for which the use of a test is recommended" (p. 13). This information does not usually include predictive validity data because, by definition, the collection of predictive validity data involves data

collection over time. By the time predictive validity data is collected, the test has been published.

In the case of Early Screening Profiles, initial data collection for predictive validation was done by this research at the same time as standardization. Follow-up data was collected prior to test and test manual publication, allowing for predictive validity coefficients to be reported in the publication manual. The research, then, has immediate significance to both the publisher of the test and to potential users of Early Screening Profiles.

CHAPTER IV

RESEARCH RESULTS

The goal of this chapter is to present the collected data, and to describe the results by responding to the six research questions raised in Chapter III.

Descriptive Statistics

As a reference source to be used as background for an interpretation of the statistical data, Table 4 presents the characteristics of the sampled population in terms of n-counts, the standard score means, and the standard deviations for the instruments used in data collection: the Cognitive Subtests of Early Screening Profiles (ESPCog), K-ABC, PPVT-R, and TRAP. Table 5 displays the ranges of the standard scores for the data collection instruments and indicates how many cases fell at or above one standard deviations below or above the mean.

The data displayed in Table 4 suggest that the group sampled was, on the whole, several points above the expected mean of 100 for a sampled population for both ESPCog and its subtests and for K-ABC Achievement (KAch) and its subtests, while for PPVT-R the mean of the sampled population was 1 point below the test mean of 100. This discrepancy gave rise to Table 5 which displays the ranges of the standard scores for subtests and composites of Early Screening Profiles and the criterion instruments. It will be noted that Table 5

indicates a larger number of low (at or greater than two standard deviations below the mean) scores for PPVT-R than for the other instruments which suggests the reason for a mean of 99 rather than of 100 or higher. The 14 children who obtained scores on the ESP Verbal Concepts subtest of 130 or higher probably influenced the ESP overall mean toward the high side.

Table 4

N-Counts, Standard Score Means, and Standard Deviations for Subtests and Composites of ESPCog, K-ABC Achievement, PPVT-R, TRAP

	n	Mean	Standard Deviation
ESP Subtests			
Verbal Concepts (VC)	136	106	16
Visual Discrimination (VD)	135(1)	103	13
Logical Reasoning (LR)	136	102	14
Basic School Skills (BS)	136	102	12
ESP Cognitive Profile (ESPCog)	135	105	14
K-ABC Achievement Subtests			
Expressive Vocabulary (EV)	21(2)	105	16
Faces and Places (FP)	136	102	12
Arithmetic (Ari)	136	102	15
Riddles (Rid)	136	103	13
Reading (Rd)	115(2)	101	14
K-ABC Achievement Total (KAch)	136	102	13
PPVT-R	136	99	19
TRAP Questions			
Question 1	96	3	1.1
Question 2	85	3	1.2
Question 3	96	3	1.1
Question 4	85	3	1.2
Question 5	96	3	1.0
TRAP Composite (TRAP)	85(3)	16.1	5.3

(1): one child not tested in this area, no ESPCog score computed. (2): subtests are age based, not all children took EV and Rd. (3): all statistics based on 85 children who were rated on all five TRAP questions.

Table 5

Standard Score Ranges and Numbers of Outliers for Subtests and Composites of ESP, K-ABC Achievement, and PPVT-R

	n	Range	outliers n<=70	outliers n>=130
ESP Subtests				
Verbal Concepts	136	71-148	0	14
Visual Discrim	135	74-150	0	4
Logical Relations	136	69-133	1	4
Basic Sch Skills	136	71-132	0	1
ESPCog	135	75-138	0	4
K-ABC Achievement Subtests				
Expressive Vocab	21	74-133	0	1
Faces & Places	136	73-132	0	2
Arithmetic	136	63-149	1	5
Riddles	136	66-131	1	2
Reading	115	56-131	4	2
K-ABC Achieve Total	136	72-128	0	0
PPVT-R	136	45-140	13	6

Table 5 data, particularly when viewed in conjunction with the data displayed in Table 4, suggest that the sampled population performed close to, but a little above, the expected mean for the population.

Results of Statistical Analyses

Question 1. Do children's scores on the composite standard score of the Cognitive Profile of Early Screening Profiles correlate highly with children's scores on the composites of the criterion measures?

Using the SPSS-X Pearson CORR Program (SPSS, Inc., 1983), Pearson product-moment correlation coefficients were computed between the composite of the Cognitive Profile and children's scores on the K-ABC Achievement

Scale Composite (KAch), PPVT-R, and TRAP. Each correlation coefficient was tested by the SPSS-X program to determine if the degree of relationship between the ESPCog score and the criterion measure differed significantly from chance level. Statistical significance was tested at the .01 level. Table 6 indicates the results.

Table 6

Pearson Product-Moment Correlation Coefficients of ESP Cog Profile Standard Scores with Standard Scores on KAch, PPVT-R, and TRAP Administered 5-1/2 to 8 Months Later

	KAch	PPVT-R	TRAP
ESP Cognitive Profile	.75**	.73**	.70**

One-tailed significance. **p < .01

As Table 6 shows, statistically significant, strong correlations were found between the Cognitive Profile of Early Screening Profiles and all three criterion measures administered 5-1/2 to 8 months later. All three of the correlations are significant at the .01 level, indicating that they are not due to chance. It is interesting to note that the correlation coefficients of ESPCog to K-Ach and PPVT-R are very close.

Question 2. Do children's scores on the subtests of the Cognitive Profile of Early Screening Profiles correlate at least moderately with children's scores on the subtests of the criterion measures?

Using the SPSS-X Pearson CORR Program (SPSS, Inc., 1983), Pearson product-moment correlation coefficients were computed between each subtest of the Cognitive Profile and children's scores on the following criterion instruments: K-ABC Achievement Scale subtests and composite, PPVT-R, and TRAP. Each correlation coefficient was tested by the SPSS-X program to determine if the degree of relationship between the Cognitive Profile subtest score and the criterion measure significantly differed from chance level. Statistical significance was tested at the .01 level.

Table 7 displays data related to question 2, indicating correlations among ESP subtests and and total and the subtests and totals of the criterion instruments.

Table 7

Pearson Product-Moment Correlation Coefficients between ESP Cognitive Subtests/Cognitive Profile Standard Scores and Standard Scores on Subtests/Composites of Criterion Instruments Administered 5-1/2 to 8 Months Later

		ESP Cognitive Subtests				
		Verbal Concepts	Visual Disc	Log Rel	Basic Skills	ESPCog Profile
K-ABC Subtests						
Exp Voc	n= 21	.48*	.26	.39*	.61**	.48*
Faces/Pl	n=136	.46**	.32**	.29**	.48**	.49**
Arith	n=136	.51**	.53**	.46**	.74**	.72**
Riddles	n=136	.72**	.34**	.41**	.50**	.65**
Reading	n=115	.38**	.35**	.26**	.64**	.53**
KAch Total	n=136	.64**	.48**	.46**	.73**	.75**
PPVT-R	n=136	.71**	.42**	.43**	.67**	.73**
TRAP Total	n= 85	.49**	.52**	.36**	.71**	.70**

One-tailed significance *p<.05 **p<.01

The data in Table 7 show that both the ESP Subtests and the ESP Cognitive Profile have statistically significant predictive correlations to the criterion measures. Most of these correlations are significant at the .01 level. The correlations between the ESP Cog Profile and the total scores of each of the three criterion measures are all highly significant. Statistically strong (r=.60 and above) or moderate correlations were found in most cases. The correlation of Expressive Vocabulary to Visual Discrimination was the only instance of a correlation that was not significant.

Question 3. Do children who score significantly low on the subtests and the composite score of the Cognitive Profile of Early Screening Profiles score significantly low five-and-a-half to eight months later on the criterion measures? Significantly low will be defined as one standard deviation or more below the mean: a standard score which is less than or equal to 85.

Using SPSS-X Crosstabs program, 2 x 2 contingency tables indicating the relationship between Cognitive Profile subtests and composite scores and scores on the criterion measures were generated. An index of agreement was obtained for each table by dividing the number of cases listed in cells 1 and 4 by the total number of cases. The percentage of both overreferrals and

underreferrals was established and tabled as were the rates of sensitivity and specificity.

The data responding to this question are displayed in four tables, all relating to standard scores <=85. Table 8 displays the percent agreement, the false positives, and the false negatives among ESP subtest scores and K-ABC subtest scores. Table 9 shows the percent agreement, the false positives, and the false negatives among the ESP Cognitive Profile, K-ABC Achievement Total, and PPVT-R. Table 10 indicates the percent agreement between the ESP Cognitive Profile, the K-ABC Achievement Total, and PPVT-R. Table 11 displays the range of percents for agreement, overreferrals, and underreferrals.

The generally high agreement rates seen in all of Tables 8 through 11 indicate that the ESP subtests and the Cognitive Profile have good specificity and sensitivity.

There is a noticeably higher underreferral rate than overreferral rate: false negatives (children who were not identified but turned out to be at risk) outnumber false positives in 21 of the 30 data displays in Tables 8 through 10. When the criterion instruments, K-ABC and PPVT-R are considered separately, the difference between the identified false positives and false negatives is

greater when ESP is looked at as predicting at risk performance on PPVT-R rather than as a predictor of K-ABC performance.

Table 8

Percent Agreement among ESP Subtest Scores and K-ABC Subtest Scores for Standard Scores <=85

ESP-Criterion Subtest to Subtest Comparisons	Total n	Agreement n	Agreement %	Overrefer False Positives n	Overrefer False Positives %	Underrefer False Negatives n	Underrefer False Negatives %
Verbal Concepts to							
Expres Vocab	21	17	81	1	5	3	14
Faces/Places	136	116	85	8	6	12	9
Arithmetic	136	113	83	8	6	15	11
Riddles	136	123	91	7	5	6	4
Reading	115	102	89	4	3	9	8
Visual Discrim to							
Expres Vocab	20	14	70	3	15	3	15
Faces/Places	135	113	84	9	7	13	9
Arithmetic	135	109	81	9	7	17	12
Riddles	135	116	86	10	7	9	7
Reading	115	98	85	5	4	12	11
Logical Relations to							
Expres Vocab	21	18	86	2	9	1	5
Faces/Places	136	116	85	11	8	9	7
Arithmetic	136	109	80	13	10	14	10
Riddles	136	113	83	15	11	8	6
Reading	115	94	82	9	8	12	10
Basic School Skills to							
Expres Vocab	21	19	90	1	5	1	5
Faces/Places	136	119	87	5	4	12	9
Arithmetic	136	118	87	4	3	14	10
Riddles	136	122	90	6	4	8	6
Reading	115	102	89	2	2	11	9

Table 9

Percent Agreement among ESP Cognitive Subtests and K-ABC Achievement Totals, PPVT-R, and TRAP for Standard Scores <=1 Standard Deviation below the Mean

ESP Cog Subtests to Total Score Comparisons	Total n	Agreement n	Agreement %	Overrefer False Positives n	Overrefer False Positives %	Underrefer False Negatives n	Underrefer False Negatives %
VC to KAch	136	117	86	7	5	12	9
VC to PPVT-R	136	112	82	3	2	21	16
VC to TRAP	85	67	79	2	2	16	19
VD to KAch	135	110	82	10	7	15	11
VD to PPVT-R	135	99	73	9	7	27	20
VD to TRAP	85	66	78	2	2	17	20
LR to KAch	136	115	85	11	8	10	7
LR to PPVT-R	136	104	77	10	7	22	16
LR to TRAP	85	65	77	2	2	18	21
BS to KAch	136	120	88	4	3	12	9
BS to PPVT-R	136	113	83	1	1	22	16
BS to TRAP	85	68	80	0	-	17	20

Table 10

Percent Agreement between ESP Cognitive Profile Standard Score and K-ABC Achievement Total, PPVT-R, and TRAP for Standard Scores <=1 Standard Deviation below the Mean

ESPCog Total Compared to:	Total n	Agreement n	Agreement %	Overrefer False Positives n	Overrefer False Positives %	Underrefer False Negatives n	Underrefer False Negatives %
K-ABC Ach	135	122	90	5	4	8	6
PPVT-R	135	111	82	4	3	20	15
TRAP	85	68	80	1	1	16	19

Table 11
Range of Percent Agreement, Overreferrals, and Underreferrals for Subtests and Totals of ESP to K-ABC Achievement and PPVT-R for Standard Scores <=85

	Range of % Agreement					
	- K-ABC Achievement -			- - PPVT-R - - -		
	Agree	False Pos	False Neg	Agree	False Pos	False Neg
ESP Sub/ KAch Sub	70-91%	2-15%	4-15%	-	-	-
ESP Sub/ Criterion Totals	82-88	3- 8	7-11	73-83	1-7	16-20
ESPCog/ Criterion Totals	90	4	6	82	3	15

The data suggest that ESP subtests and Cognitive Profile may be better predictors of low achievement in specific school related tasks, such as those measured by K-ABC Achievement, than of a more global skill such as receptive vocabulary, tested by PPVT-R. For standard scores <=85, Table 11 indicates a lower overall agreement and a higher rate of underreferrals for PPVT-R than for K-ABC Achievement.

When a test is used for prediction, which is a primary use of ESP, it is helpful to know how sensitive and specific the test is in its predictions. Does it correctly identify at risk children; that is, to what extent is it sensitive to at risk children? Does it correctly identify children who are not at risk; that is, to what extent does it specify children who are not

at risk? The purpose of Table 12 is to display the sensitivity and specificity of the various subtests and the total ESP Cognitive score relative to the total scores of the three criterion instruments administered 5-1/2 to 8 months after administration of ESP.

Table 12

Sensitivity and Specificity of ESP Subtest and Cognitive Profile Standard Scores Compared to Performance on KAch, PPVT-R, and TRAP for ESPCog Scores <=1 SD below the Mean

	K-Ach n=135		PPVT-R n=135		TRAP n=85	
	Sen.	Spec.	Sen.	Spec.	Sen.	Spec.
Verb Con	29%	94%	30%	97%	20%	97%
Vis Disc	6	92	6	92	11	97
Log Rel	41	91	27	91	5	97
B Sch Sk	29	97	26	99	15	100
ESPCog	50	97	31	96	16	98

The data displayed in Table 12 indicate that, for this study, ESP Cognitive subtests and the Cognitive Total were highly successful in specifying children who would not be at risk of doing poorly in school.

Question 4. Do children who score significantly high on the subtests and the composite of the Cognitive Profile of Early Screening Profiles score significantly high five-and-a-half to eight months later on the criterion measures? Significantly high will be defined as one standard deviation or more above the mean: a standard score greater than or equal to 115.

Using SPSS-X Crosstabs Program, 2 x 2 contingency tables indicating the relationship between Cognitive Profile subtests and composite scores and scores on the criterion measures were generated. An index of agreement was obtained for each table by dividing the number of cases listed in cells 1 and 4 by the total number of cases. The percentage of both overreferrals and underreferrals was established and tabled as were the rates of sensitivity and specificity.

As with Question 3, the data responding to this question are displayed in four tables; in this case, relating to standard scores >=115. Table 13 displays the percent agreement and the false positives and false negatives among ESP subtest scores and K-ABC subtest scores. Table 14 shows the percent agreement, the false positives, and false negatives among the ESP Cognitive Profile, K-ABC Achievement Total, and PPVT-R. Table 15 indicates the percent agreement between the ESP Cognitive Profile, the K-ABC Achievement Total, and PPVT-R. Table 16 displays the range of agreement percents.

The data from the tables indicate a reasonably high rate of agreement between ESP and the criterion instruments for scores >=115, with the majority of the percent agreements in the 70s. This agreement is not as high as was found for scores <=85, where the majority of the percent agreements fell in the 80s.

Table 13

Percent Agreement among ESP Subtest Scores and K-ABC Subtest Scores for Standard Scores >=115

ESP-Criterion Subtest to Subtest Comparisons	Total n	Agreement n	Agreement %	Overrefer False Positives n	Overrefer False Positives %	Underrefer False Negatives n	Underrefer False Negatives %
Verbal Concepts to							
Expres Vocab	21	18	86	0	0	3	14
Faces/Places	136	94	69	32	24	10	7
Arithmetic	136	101	74	24	18	11	8
Riddles	136	106	78	22	16	8	6
Reading	115	77	67	29	25	9	8
Visual Discrimination to							
Expres Vocab	20	14	70	1	5	5	25
Faces/Places	135	97	72	24	18	14	10
Arithmetic	135	108	80	14	10	13	10
Riddles	135	101	75	8	13	16	12
Reading	115	89	77	17	15	9	8
Logical Relations to							
Expres Vocab	21	14	67	4	19	3	14
Faces/Places	136	107	79	5	11	14	10
Arithmetic	136	106	78	11	8	19	14
Riddles	136	105	77	12	9	19	14
Reading	115	101	90	4	4	7	6
Basic School Skills to							
Expres Vocab	21	16	76	0	0	5	24
Faces/Places	136	103	76	7	12	16	12
Arithmetic	136	114	84	7	5	15	11
Riddles	136	109	80	10	7	17	30
Reading	115	96	83	10	9	9	8

Table 14

Percent Agreement among ESP Cognitive Subtests and K-ABC Achievement Totals, PPVT-R, and TRAP for Standard Scores >=1 Standard Deviation above the Mean

Subtest to Total Score Comparisons	Total	Agreement		Overrefer False Positives		Underrefer False Negatives	
	n	n	%	n	%	n	%
VC to KAch	136	107	79	23	17	6	4
VC to PPVT-R	136	105	77	20	15	11	8
VC to TRAP	85	57	67	22	26	6	7
VD to KAch	135	108	80	16	12	11	8
VD to PPVT-R	135	104	77	14	10	17	13
VD to TRAP	85	63	74	13	15	9	11
LR to KAch	136	112	83	10	7	14	10
LR to PPVT-R	136	98	72	13	10	25	18
LR to TRAP	85	65	76	5	6	15	18
BS to KAch	136	116	85	8	6	12	9
BS to PPVT-R	136	108	79	8	6	20	14
BS to TRAP	85	68	80	8	9	9	11

Table 15

Percent Agreement between ESP Cognitive Profile Standard Score and K-ABC Achievement Total, PPVT-R, and TRAP for Standard Scores >=1 Standard Deviation above the Mean

ESPCog Total Compared to:	Total	Agreement		Overrefer False Positives		Underrefer False Negatives	
	n	n	%	n	%	n	%
K-ABC Ach	135	112	83	18	13	5	4
PPVT-R	135	110	82	15	11	10	7
TRAP	85	61	72	19	22	5	6

Table 16

Range of Percent Agreement, Overreferrals, and Underreferrals for Subtests and Totals of ESP, K-ABC Achievement, and PPVT-R for Standard Scores >=115

	Range of % Agreement					
	- K-ABC Achievement -			- - PPVT-R - - -		
	Agree	False Pos	False Neg	Agree	False Pos	False Neg
ESP Sub/ KAch Sub	67-90%	0-25%	6-30%	-	-	-
ESP Sub/ Criterion Totals	79-85	6-17	4-10	72-79	6-15	8-18
ESPCog/ Criterion Totals	83	13	4	82	11	7

The lower overall rate of agreement here, with scores >=115, as compared to the data in Tables 12-14, suggest that ESP appeared to be more sensitive to the at risk child in this study than to the child with above average academic potential.

The large percent of agreement discrepancy between PPVT-R and K-ABC Achievement when each is compared to the ESP Cognitive total for scores <=85 was not observed here with scores >=115. The latter case is more reflective of the similarity between the Pearson Product Moment Correlations of K-ABC Achievement and PPVT-R to the ESP Cognitive Total seen in Table 7.

Table 17 displays the sensitivity and specificity of ESP subtests and the ESP Cognitive Total for children

whose standard scores on ESPCog and the criterion instruments fell at or above 1 standard deviation above the mean.

Table 17

Sensitivity and Specificity of ESP Subtest and Cognitive Profile Standard Scores Compared to Performance on KAch, PPVT-R, and TRAP for Scores >=1 SD above the Mean

	K-Ach n=135		PPVT-R n=135		TRAP n=85	
	Sen.	Spec.	Sen.	Spec.	Sen.	Spec.
Verb Con	74%	80%	65%	81%	67%	67%
Vis Disc	52	86	45	87	53	80
Log Rel	39	91	19	88	17	93
B Sch Sk	48	93	35	92	50	88
ESPCog	78	84	68	86	77	70

As was seen in Table 12, ESP subtests and Cognitive Profile were found to be highly specific. In this case, ESP specifically identified high percentages of children who would later score above average on the criterion instruments.

Question 5. Are there differences in the ability of Early Screening Profiles to predict criterion measure performance based on sex?

Using SPSS-X Regression Program, correlations between each Cognitive Profile score and each criterion measure for females and for males were plotted and fitted to a regression line. The slope of the regression line for males was compared with the slope of the regression line for females using a procedure described by Neter and

Wasserman (1974). The statistical significance of the difference between the slopes of the female and male regression lines of each score comparison was tested at the .01 level.

Table 18 displays the correlations between the ESP Cognitive Profile Total Scores for boys versus girls relative to the three criterion instruments: K-ABC Achievement, PPVT-R, and TRAP.

Table 18

Pearson Product Moment Correlation Coefficients and F-Test Results for Females Versus Males on ESPCog Total and the Criterion Instruments

Variable	Number of Cases	Mean	Standard Deviation	r	F
Females					
ESP-Cog	64	106	14.6		
				.80	
K-Ach Total	64	102	13.6		
					.714 NS
Males					
ESP-Cog	71	104	13.0		
				.69	
K-Ach Total	71	102	12.0		
Females					
ESP-Cog	64	106	14.6		
				.77	
PPVT-R	64	99	20.8		
					1.167 NS
Males					
ESP-Cog	71	104	13.0		
				.70	
PPVT-R	71	100	18.0		
Females					
ESP-Cog	41	108	13.8		
				.70	
TRAP	41	16	5.4		
					.323 NS
Males					
ESP-Cog	44	109	12.3		
				.70	
TRAP	44	17	5.4		

NS: not significant (p>.01)

F-tests conducted at the .01 level on differences between the scores of males and females were conducted for each of the three criterion measures. Even though there are differences between correlations of ESPCog and KAch and between ESP Cog and PPVT-R (.69 versus .80 and .70 versus .77 respectively), the differences are not statistically significant. ESPCog predicts KAch, PPVT-R, and TRAP equally well for males and females.

Question 6. Are there differences in the ability of Early Screening Profiles to predict criterion measure performance based on age?

Using SPSS-X Regression Program, correlations between each Cognitive Profile score and each criterion measure for children ages 3-6 through 4-11 and for children ages 5-0 through 6-11 were plotted and fitted to a regression line. The slope of the regression line for children ages 3-6 through 4-11 was compared to the slope of the regression line for children 5-0 through 6-11 using a procedure described by Neter and Wasserman (1974). The statistical significance of the difference between the slopes of the regression lines for the younger and the older groups for each score comparison was tested at the .01 level.

Table 19 displays the coefficients between the ESP Cognitive Total Score and two of the criterion instruments for children who were ages 3-6 through 4-11 at the

time of ESP testing versus children who were 5-0 through 6-11 at ESP testing. No data is presented for TRAP since the TRAP items were not appropriate for children in the younger group.

Table 19

Pearson Product Moment Correlation Coefficients and F-Test Results for Younger Versus Older Children on ESPCog Total to K-Ach and PPVT-R

Variable	Number of Cases	Mean	Standard Deviation	r	F
3-6 thru 4-11					
ESP-Cog	46	98	12.6		
				.71	
K-Ach Total	46	101	12.7		
					8.49*
5-0 thru 6-11					
ESP-Cog	85	108	13.1		
				.82	
K-Ach Total	85	103	12.8		
3-6 thru 4-11					
ESP-Cog	46	98	12.6		
				.67	
PPVT-R	46	92	17.1		
					.51
5-0 thru 6-11					
ESP-Cog	85	108	13.0		
				.71	
PPVT-R	85	104	19.2		

*p<.01 F=4.79

F-tests were conducted to test differences between the scores of younger versus older children for both KAch and for PPVT-R. The ESPCog and K-Ach correlations for the two age groups are significantly different at the .01 level, indicating that ESPCog predicts K-ACh better for older children than for younger children.

PPVT-R age differences are not significant.

The results of the research questions will be discussed further in Chapter V.

CHAPTER V

DISCUSSION

The goal of this research was to examine the predictive validity characteristics of the Cognitive Subtests of Early Screening Profiles for ages 3-1/2 to 6-1/2. ESP is a nationally standardized early childhood screening instrument, to be published in 1990, designed to help identify children who, upon entering school, may be at risk of having academic problems in cognition, adaptive behavior, or motor skills.

This chapter summarizes and discusses the research results and makes some suggestions for future research.

Summary of Results

Research Question 1. Do children's scores on the composite standard score of the Cognitive Profile of Early Screening Profiles correlate highly with children's scores on the criterion measures?

Each of the three criterion measures to which ESPCog was compared was selected for a different reason. The K-ABC Achievement Scale was chosen because of its sound psychometric qualities and because the skills it measures (integrated language, arithmetic knowledge, background of information, and reading decoding) are all generally acknowledged to be important school skills. PPVT-R was selected due to its widespread use as a screener and because it measures the important global skill area of

receptive language. TRAP was chosen to provide a basis of comparison between the results of highly objective standardized instruments and the more personal, long term perception of the child by a trained observer, the child's classroom teacher.

It was anticipated that the correlations between ESPCog and each of the criterion measures would be positive and statistically strong, as they are (.75 to K-ABC Ach, .73 to PPVT-R, and .70 to TRAP). The coefficients are not only high and strong, but consistent with each other, indicating that ESPCog predicted equally well for highly objective and for less objective types of instruments. Additionally, although the two nationally standardized criterion measures are unlike each other (one broad based, one narrow in skill range), ESPCog predicted equally well for both of them, giving further support to the ability of ESPCog to serve as a valid screening instrument. Interestingly, Bing and Bing (1985), comparing the K-ABC and PPVT-R scores of thirty Head Start children, found high correlations between the K-ABC Achievement Scale and the PPVT-R.

Predictive validity coefficients were given in Chapter II of this research for five of the seven screeners reviewed in that chapter. In each case, correlation coefficients were cited as evidence of the instrument's ability to predict. In the Boehm-R manual

(Boehm, 1986) the median coefficient of predictive validity was .44. DIAL (not DIAL-R) predictive validity oefficients ranged from .45 to .73 (Mardell-Czudnowski & Goldenberg, 1983). The coefficients between MST and MRT ranged from .31 to .57 and were judged moderately strong (The Psychological Corporation, p. 12). The MAP correlation coefficient to WISC-R was .50; to various subtests of the W-J Psychoeducational Battery, correlations were between .35 and .38 (Miller, 1988, p. 115). Guerin and Gottfried (1987) reported that the MCDI correlations to criterion instruments (K-ABC, WISC-R, and WRAT-R), all significant at .01, ranged from .45-.69 (MCDI to K-ABC Achievement = .69). LaRoche (1989), reviewing predictive validity correlations of a number of screening instruments, concluded that correlations exceeding .50 appear to provide acceptable evidence of an instrument's predictive validity.

Based on this information, correlations found in this study indicate that children's scores on the composite standard score of the Cognitive Profile of ESP do indeed correlate highly with children's scores on the criterion measures of K-ABC Achievement (.75), PPVT-R (.73), and TRAP (.70); all correlations are significant at .01.

Research Question 2. Do children's scores on the subtests of the Cognitive Profile of ESP correlate at

least moderately with children's scores on the subtests of the criterion measures?

The correlations displayed in Table 7 divide into three groups: low, middle, and high. The lowest set of sixteen correlations, ranging from .26 to .53, was found when the two ESP subtests of Visual Discrimination and Logical Relations were compared to criterion performance. Though all but one of these is statistically significant, all of them are weak or, at best, moderately strong. This is not surprising because both Visual Discrimination and Logical Relations examine the least content related areas of the ESP subtests. These two subtests could be considered ability and concept related. In this group of correlations, the two highest, .52 and .53, compare Visual Discrimination to TRAP and to K-ABC Arithmetic, respectively. This suggests that the non-language based skill tapped by visual discrimination is important in the early grades and is perceived as important by classroom teachers.

A second set of correlations, those between the ESP Verbal Concepts subtest and the criterion measures, ranged from .38 to .72 with an average of .55. Here, high correlations (.72 and .71) are in the expected areas: Verbal Concepts compared to K-ABC Riddles (definitions) and to PPVT-R.

A third set of correlations is made up of the ESP subtest, Basic School Skills, and the ESP Cognitive Profile. These sixteen correlations range from .48 to .75. All are statistically significant. They are moderate to strong correlations. Basic School Skills predicted performance on TRAP, on the K-ABC subtests of Expressive Vocabulary, Faces and Places, Arithmetic, Reading Decoding, and on the Achievement Total about equally well or better than did the ESP Cognitive Profile. The ESP Cognitive Profile had stronger comparisons than Basic School Skills between K-ABC Riddles and PPVT-R, both of which tap receptive vocabulary. It will be noted that the ESP subtest, Verbal Concepts, predicted best for Riddles and PPVT-R.

With a few exceptions (six of forty correlations <=.35, all in Visual Discrimination and Logical Relations), children's scores on the subtests of the Cognitive Profile of ESP correlate moderately or strongly with children's scores on the subtests of the criterion measures. Of the subtests, the strongest overall was found to be Basic School Skills which predicted about as well as the ESP Cognitive Profile for both K-ABC Achievement and for TRAP. Verbal Concepts predicted about as well as the ESP Cognitive Profile for PPVT-R.

Research Question 3. Do children who score significantly low on the subtests and the composite of

the Cognitive Profile of Early Screening Profiles score significantly low five-and-a-half to eight months later on the criterion measures? Significantly low was defined as one standard deviation of more below the mean: a standard score which is less than or equal to 85.

Five tables of data were presented in Chapter IV to respond to this question. Tables 8, 9, and 10 display data from a narrow to a broad base and will be discussed together, along with Table 11 which merges the data in Tables 8-10.

Tables 8 9, and 10, resulting from crosstabulations, show impressive and consistent rates of agreement between subtests of ESP and K-ABC Achievement, between ESP subtests and the criterion totals, and between the ESP Cognitive Profile and criterion totals. Agreement rates below 75% occurred in two cases where the ESP subtest had little in common with the content of the criterion (Visual Discrimination merged with PPVT-R and Visual Discrimination merged with Expressive Vocabulary on K-ABC). In all other instances, at least three-quarters (75%) of the sampled population is captured in the agreement rate. This is higher than the rates of agreement range of 64% to 79% on ESI, a developmental screening test (Meisels & Wiske, 1988), and higher than MAP rates of agreement of 77% and 78% when the MAP 25% cutoff point is used (Miller, 1988). This suggests that

users of ESP, when 85 is used as a cutoff point, can be confident that approximately three quarters of their screened children will be accurately identified as at risk or not at risk of academic difficulty. Of course other factors influence a child's school performance; a screener can only point toward likely outcomes, not assure them.

Several points are of particular interest. In Table 8, the most consistently high agreement rates are between the K-ABC Achievement Subtests and Basic School Skills of the ESPCog subtests, pointing to Basic School Skills as the best of the ESPCog subtests as a single predictor of success or academic difficulty in kindergarten or grade 1. Table 9 indicates that any one of the ESPCog subtests predicts somewhat better for K-ABC Achievement than for PPVT-R performance or performance as assessed by the child's kindergarten or grade 1 teacher. However, the high agreement rates between the TRAP total and the ESPCog subtests support the use of this teacher rating scale as providing a useful and accurate appraisal of children's performance.

The agreement rates capture children who scored below or at 85 on both ESPCog and its subtests and the criterion instruments and those who scored above 85 both times. Of the remaining children, ESP tended to under-refer more frequently than it overreferred. As Table 11

shows, the percent of false negatives between the ESP subtests and the K-ABC Achievement subtests and composite ranges from 7% to 15%. When ESPCog is compared to the K-ABC Achievement total, only 6% of actually at risk children were not identified by ESP. Only eight of the 135 children with ESPCog Profile scores were underreferred. The percent of underreferrals for PPPVT-R and TRAP is higher (15% and 19% respectively) than for K-ABC Achievement, indicating that, for this study, ESP was a better predictor of specific school related tasks than of global skills such as, in this case, receptive language and overall classroom performance. This is of particular interest since classroom performance is the real world criterion on which children are rated by trained observers, their classroom teachers.

The rate of false positives for scores below or equal to one standard deviation below the mean is consistently low, ranging from 0% to 15% with only three data displays at or above 10%. This indicates that ESP is expected to have a low incidence of overreferrals.

Sensitivity and specificity data, shown in Table 12, indicate that ESPCog is highly specific in identifying children who are not at risk, with an average specificity rate of .97 for the Cognitive Profile. This is higher than the specificity rates of .82 for kindergarten and .72 for grade 1 reported by Meisels and Wiske (1988) for

ESI. It is about the same as the high specificity rate reported for DABERON by LaRoche (1989).

The sensitivity, that is, the extent to which ESPCog identified at risk children, is considerably lower. The ESPCog sensitivity for the prediction of performance on the three criterion measures ranges from weak (16%) to modest (31%) to moderately high (50%), resulting in an average of 32%. The DABERON, another kindergarten screening instrument, had a sensitivity of 31% in the study by LaRoche (1989), about the same as the combined ESPCog sensitivity. Meisels and Wiske (1988), on the other hand, report sensitivity rates of 88% and 92% for children in grades kindergarten and 1. This is of particular interest since its authors describe ESI as a developmental screener rather than an achievement based one.

The high specificity suggests that ESPCog may be more efficient at specifying children who are not at risk of academic failure than of locating at risk children. The underreferral rates indicate that some children who turned out to have academic difficulties were not referred by ESP. This is acceptable for a screener and implicit in the use of the term "screener". As Meisels (1988) points out, tests with high specificity lead to few overreferrals.

Research Question 4. Do children who score significantly high on the subtests and the composite of the Cognitive Profile of ESP score significantly high five-and-a-half to eight months later on the criterion measures? Significantly high was defined as one standard deviation or more above the mean: a standard score which is higher than or equal to 115.

As with question 3, five tables of data were presented in Chapter IV to respond to this question. Here, Tables 13, 14, and 15, displaying data from a narrow through a broad base will be discussed in conjunction with Table 16 which merges the data.

As with scores <=85, there are high rates of agreement seen between the subtests of K-ABC and the ESPCog subtests (67% to 90%), with the highest, on the average, in Basic School Skills. These percents, while high, are not as impressive as the ones for scores <=85. This is due to the higher overall rate of false positives for scores >=115. ESP tended to identify relatively large numbers of children as being capable of above average performance when, in fact, they scored lower than 115 on the criterion instruments. One conclusion this suggests is that children who score within the normal range (between 85 and 115), have the opportunity of doing well on ESP and, consequently, feel positive about what might well be their first formal school experience.

Since the overall rate of false positives is higher for scores >=115 than for scores <=85, this pattern reverses itself when underreferrals are compared because the rates of agreement remained relatively high. The underreferral rates for the three criterion instruments with scores > = 1 standard deviation above the mean are 4%, 7% and 6%, whereas they were 6%, 15%, and 19% when scores < = 1 standard deviation below the mean were compared to the same three criterion instrument scores.

When high ESPCog Profile scores are merged with high criterion measure total scores, ESP shows moderately high rates of specificity (84%, 86%, and 70%). This means ESP is able to specify children who are not likely to perform at an above average level in school. On the other hand, the sensitivity of ESP for high scores is also moderately high (78%, 68%, and 77%).

In the discussion of question 3, it was possible to compare ESPCog data with data from other screening instruments. That has not been the case here since currently available early childhood screening instruments tend to stress screening for the child at risk of academic failure and do not deal with children at the other end of the spectrum.

The strong agreement rates indicate that children who obtained ESP scores >=115 achieved correspondingly high scores on the criterion instruments. The higher

specificity rate than sensitivity rate suggests that ESPCog may be expected to be somewhat more efficient at specifying children who may perform at above average levels than of locating children who will perform at a level lower than one standard deviation above the mean. This is consistent with the results of question 3.

Research Question 5. Are there differences in the ability of Early Screening Profiles to predict criterion measure performance based on sex?

Bias in testing is an on-going issue, one that relates to the usefulness of a test. Reynolds (1980) suggested that test developers need to be aware of the issue of bias and demonstrate predictive validity as part of test development. Clearly, if a test should predict significantly better either for girls or for boys, its usefulness as a general screening measure is lessened. This is not the case with ESPCog. There were no significant differences found in the ability of ESPCog to predict criterion measure peformance based on sex.

Research Question 6. Are there differences in the ability of Early Screening Profiles to predict criterion measure performance by age?

As Table 19 shows, the correlations between ESPCog and PPVT-R for the younger and older children are not significantly different. For TRAP data, there were

insufficient numbers of children (n=2) in the younger age group to test for differences.

There is a significant difference between scores of younger and older children in the sampled population when ESPCog is compared to K-ABC Achievement. The older group of children scored significantly higher. The younger children were ages 3-6 through 4-11 at time of ESPCog testing; they were still in pre-kindergarten or early in their kindergarten year at the time of criterion testing. The older children were ages 5-0 through 6-5 at time of ESPCog testing; of these, 79 had at least two months of first grade experience behind them at the time of criterion testing. Since the two ESP subtests with the highest correlations with the criterion instruments, Verbal Concepts and Basic School Skills, are also the ones with the greatest amount of school related content, it is hypothesized that the reason for the difference lies in the fact that the older children had exposure to formal school instruction for a longer time period than the younger children. Another hypothesis is that differences are anchored in other factors outside the scope of this research such as curriculum differences, or differences in race, ethnicity, and SES.

Suggestions for Future Research

Since ESP is a new test, it affords numerous opportunities for research relating to predictive

validity and other areas. Several suggestions for research result from the present study.

ESP should be correlated with other measures of school achievement in order to confirm the results of this study and to provide support for the use of ESP as an early childhood screening instrument. These studies should concentrate not only on the Cognitive Profile, as this study did, but deal also with the other ESP subtests of Adaptive Behavior and Motor Skills.

Cutoff points other than +/- 1 standard deviation from the mean would be useful in research studies so that the ability of ESP to predict according to various cutoff points can be established, making the test a more flexible one, suited to the varying needs of school districts.

Since the ESPCog school achievement related subtests of Verbal Concepts and Basic School Skills correlated more highly with the criterion measures in this study than did the ability oriented subtests of Logical Relations and Visual Discrimination, research comparing performance on these two pairs of subtests would be useful.

Children who are initially screened on Early Screening Profiles should be tested or otherwise rated one and two years later to see if, in fact, their

educational placements correspond to those predicted by their scores on Early Screening Profiles.

Teacher rating scales should be included as criterion measures in predictive validity studies of ESP in order to corroborate the finding of this study which indicated that the Teacher Rating of Academic Performance (TRAP) was a good criterion measure. The longitudinal collection of TRAP data on children in this study is already in progress.

Since this preliminary research on ESP indicated significantly higher correlations for older than for younger children, the stability of the ESP Cognitive Profile over time should be tested.

This study did not address differences attributable to SES, race, ethnicity, or testing conditions. Research on ESP, incorporating information on these and other variables, would be useful.

It would be useful to have more data available on screening for children who are likely to perform at an above average level once they begin school. Research studies dealing with that screening area would be desirable.

Conclusions

ESP Cognitive subtests, particularly those of Expressive Vocabulary and Basic School Skills, as well as the Cognitive Profile, correlated highly with all three

criterion measures and evidenced high rates of agreement with the criterion instruments when scores of +/- 1 standard deviation from the mean were used as cutoff points. The highest correlations were found between ESP and K-ABC Achievement, suggesting that the ESP Cognitive Profile is a better predictor of specific school related tasks than of global skills.

Overall, the results of the research indicate that the Cognitive Profile and the cognitive subtests of Early Screening Profiles give promise of being useful and valid additions to the field of early childhood screening.

REFERENCES

Altepeter, T. (1985). Use of the PPVT-R for intellectual screening with a preschool pediatric sample. Journal of Pediatric Psychology, 10(2), 195-198.

American Psychological Association (1985). Standards for Educational and Psychological Testing. Washington, D.C: Author.

Anastasi, A. (1988). Psychological Testing, Sixth Edition. New York: Macmillan Publishing Co., Inc.

Bing, S.B. & Bing, J.R. (1985). Comparison of the K-ABC and PPVT-R with Head Start children. Psychology in the Schools, 22(3), 245-249.

Board of Education of the City of New York (1983). Getting Started in the All-Day Kindergarten. New York: Author.

Board of Education of the City of New York (1986). Three, Four, Open the Door. New York: Author.

Boehm, A.E. (1986). Boehm Test of Basic Concepts, Revised. San Antonio: The Psychological Corporation.

Bracken, B.A., Prasse, D.P., & McCallum, R.S. (1984). Peabody Picture Vocabulary Test-Revised: An appraisal and review. School Psychology Review, 13(1), 49-60.

Cadman, D., Chambers, L.W., Walter, S.D., Feldman, W., Smith, K., & Ferguson, R. (1984). The usefulness of the Denver Developmental Screening Test to predict kindergarten problems in a general community population. American Journal of Public Health, 74(10), 1093-1097.

Childers, J.S., Durham, T.W., Bolen, L.M., & Taylor, L.H. (1985). A predictive validity study of the Kaufman Assessment Battery for Children with the California Achievement Test. Psychology in the Schools, 22(1), 29-33.

Cohn, S.H. (1986). An analysis of the predictive validity of the Miller Assessment for Preschoolers in a suburban public school district. Unpublished doctoral dissertation. Denver: University of Denver.

Colligan, R.C. (1982). Prediction of school performance from the Minnesota Child Development Inventory: implications for preschool screening. Paper presented at the Annual Meeting of the American Psychological Association, Washington, D.C., August.

Dunn, L.M. & Dunn, L.M. (1981). Peabody Picture Vocabulary Test-Revised. Circle Pines, MN: American Guidance Service.

Educational Research Service, Inc. (1986). ERS Report: Kindergarten Programs and Practices in Public Schools. Arlington, VA: Author.

Fletcher, J.M. & Satz, P. (1982). Kindergarten prediction of reading achievement: a seven-year longitudinal follow-up. Educational and Psychological Measurement, 42(2), 681-685.

Frankenburg, W.K., Dodds, J.B., Fandal, A.W., Kazuk, E., & Cohrs, M. (1975). Denver Developmental Screening Test Reference Manual-Revised. Denver: University of Colorado Medical Center.

Gresham, F.M., Reschly, D.J., & Carey, M.P. (1987). Teachers as "Tests": classification accuracy and concurrent validation in the identification of learning disabled children. School Psychology Review, 16(4), 543-553.

Guerin, D. & Gottfried, A.W. (1987). Minnesota Child Development Inventories: predictors of intelligence, achievement, and adaptability. Journal of Pediatric Psychology, 12(4), 595-609.

Ireton, H. (1988). Developmental Screening Measures. Personal communication, in press.

Ireton, H. & Thwing, E. (1974). Minnesota Child Development Inventory. Minneapolis: Behavior Science Systems, Inc.

Kamphaus, R.W. & Reynolds, C.R. (1987). Clinical and Research Applications of the K-ABC. Circle Pines, MN: American Guidance Service.

Kaufman, A.S., Bruininks, R.H., Sparrow, S., Kaufman, N.L., Harrison, P.L., Ilmer, S., Rynder, J., & McCloskey, G. (1987). Early Screening Profiles, Standardization Edition. Circle Pines, MN: American Guidance Service.

Kaufman, A.S. & Kaufman, N.L. (1983). The Kaufman Assessment Battery for Children. Circle Pines, MN: American Guidance Service.

LaRoche, M.F. (1989). A Predictive Validity Study of the DABERON: A Screening Test Used For Identifying Children Who May Be At-Risk for Academic Failure. Unpublished doctoral dissertation. Amherst: University of Massachusetts.

Lehr, C.A., Ysseldyke, J.E., & Thurlow, M.L. (1986). Assessment Practices in Model Early Childhood Education Programs. Minneapolis: University of Minnesota.

Lemerand, P.A. (1985). Predictive validity of the MAP. Unpublished doctoral dissertation. Ann Arbor: University of Michigan.

Lichtenstein, R. & Ireton, H. (1984). Preschool Screening. Orlando: Grune & Stratton, Inc.

Lyon, M.A. & Smith, D.K. (1987). Stability of the Kaufman Assessment Battery for Children for a sample of at-risk preschool children. Psychology in the Schools, 24(2), 111-115.

Maloney, M.P. & Ward, M.P. (1976). Psychological Assessment: A Conceptual Approach. New York: Oxford University Press.

Mardell-Czudnowski, C.D. & Goldenberg, D.S. (1983). Developmental Indicators for the Assessment of Learning-Revised. Edison, NJ: Childcraft Education Corp.

Meisels, S.J. (1985). Developmental Screening in Early Childhood: A Guide (rev. ed.). Washington, D.C.: National Association for the Education of Young Children.

Meisels, S.J. (1987). Uses and abuses of developmental screening and school readiness testing. Young Children, 42(2), 4-6 & 68-73.

Meisels, S.J. (1989). Can developmental screening tests identify children who are developmentally at risk? Pediatrics, 83(4), 578-585.

Meisels, S.J. & Wiske, M.S. (1988). Early Screening Inventory Test and Manual, Second Edition. New York: Teachers College Press.

Meisels, S.J., Wiske, M.S., & Tivnan, T. (1984). Predicting school performance with the Early Screening Inventory. Psychology in the Schools, 21(1), 25-33.

Miller, L.J. (1982). Miller Assessment for Preschoolers. Littleton, CO: The Foundation for Knowledge in Development.

Miller, L.J. (1988). Miller Assessment for Preschoolers (rev. ed.). San Antonio: The Psychological Corporation.

Murray, A. & Bracken, B.A. (1984). Eleven-month predictive validity of the Kaufman Assessment Battery for Children. Journal of Psychoeducational Assessment, 2(3), 225-232.

Naglieri, J.A. & Harrison, P.L. (1982). McCarthy Scales, McCarthy Screening Test, and Kaufman's McCarthy Short Form correlations with the Peabody Individual Achievement Test. Psychology in the Schools, 19(2), 149-155.

Neter, J. & Wasserman, W. (1974). Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs. Homewood, IL: Richard D. Irwin, Inc.

Nugent, J.H. (1976). A comment on the efficiency of the Revised Denver Developmental Screening Test. American Journal of Mental Deficiency, 80(5), 570-572.

Psychological Corporation (1978). McCarthy Screening Test Manual. New York: Author.

Reynolds, C.R. (1980). Differential predictive validity of a preschool battery across race and sex. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, April.

Robertson, G.J. & Eisenberg, J.L. (1981). Peabody Picture Vocabulary Test-Revised Technical Supplement. Circle Pines, MN: American Guidance Service.

Salvia, J. & Ysseldyke, J.E. (1988). Assessment in Special and Remedial Education, Fourth Edition. Boston: Houghton Mifflin Company.

Satz, P. & Fletcher, J.M. (1979). Early screening tests: some uses and abuses. Journal of Learning Disabilities, 12(1), 56-60.

SPSS, Inc. (1983). SPSS-X User's Guide. New York: McGraw Hill Book Co.

Tarnowski, K.J. & Kelly, P.A. (1987). Utility of the PPVT-R for pediatric intellectual screening. Journal of Pediatric Psychology, 12(4), 611-614.

U.S. Code Congressional and Administrative News (1975). PL 94-142: Education for All Handicapped Children Act of 1975 (11-29-75). 89 Stat., 773.

U.S. Code Congressional and Administrative News (1983). PL 98-199: Education of the Handicapped Act Amendments of 1983 (12-2-83). 97 Stat., 1357.

U.S. Code Congressional and Administrative News (1986). PL 99-457: Education of the Handicapped Act Amendments of 1986 (10-8-86). 100 Stat., 1145.

University of the State of New York (1982). Screening in New York State. Albany: Author.

Vance, B., Kutsick, K., & West, R. (1987). Concurrent validity of the Peabody Picture Vocabulary Test-R and the Expressive One-Word Picture Vocabulary Test for language-delayed and non-language-delayed young children. Diagnostique, 13(1), 3-9.

Wilson, B.J. & Reichmuth, M. (1984). Early screening programs: when is predictive accuracy sufficient? Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, April.

Ysseldyke, J.E., Thurlow, M.L., O'Sullivan, P., & Bursaw, R.A. (1985). Current Screening and Diagnostic Practices for Identifying Young Handicapped Children. Minneapolis: University of Minnesota.