NOTICE

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

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

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

AVIS

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

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

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

ISBN 0-315-55331-0

THE UNIVERSITY OF ALBERTA

THE EFFECTIVENESS OF THE BRIGANCE K AND 1 SCREEN FOR PREDICTING ABILITY AND ACHIEVEMENT IN THE EARLY GRADES

BY

ANNE T. SHILLINGTON ©

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE

OF MASTER OF EDUCATION IN SCHOOL PSYCHOLOGY

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

FALL, 1989

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Anne T. Shillington

TITLE OF THESIS: The Effectiveness of the Brigance K and 1 Screen for Predicting Ability and Achievement in the Early Grades

DEGREE: Master of Education in School Psychology

YEAR THIS DEGREE GRANTED: Fall, 1989

Anne T. Shillington
(Student's Signature)

Box 263
Stony Plain, Alberta
TOE 2G0
(Student's permanent address)

Date: September 28/89

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **'The Effectiveness of the Brigance K and 1 Screen for Predicting Ability and Achievement in the Early Grades'** submitted by Anne T. Shillington in partial fulfillment of the requirements for the degree of Master of Education.

....
(Supervisor)

....

....

Date ..August..21, 1989...

ABSTRACT

This retrospective study investigated the validity of the Brigance K and 1 Screen for First Grade (Brigance, 1982) (BK1-1) for predicting indices of academic ability and achievement in grades one through three at an English language, urban-rural elementary school. The total sample of 149 children (80 boys and 69 girls) were enrolled in grades one to five during the 1988 - 1989 school year and had been administered the BK1-1 either at the end of their kindergarten year (n = 114) or during their first month of grade one (n = 35) at the school.

The childrens' total test and subtest scores from the BK1-1 were used to predict their standing on the following indices of academic ability and achievement: 1) the Canadian Cognitive Abilities Test (end of grade one); 2) the Gates-MacGinitie Reading Test (grades one, two and three); 3) the Canadian Achievement Test (end of grades two and three); and 4) teacher-assigned grades (end of grades one, two and three). Following the calculation of descriptive statistics, these data were subjected to both correlational and classificational analysis to establish the predictive effectiveness of the BK1-1. A stepwise multiple regression analysis was also used to determine which of the thirteen BK1-1 subtests were maximally effective for predicting academic ability and achievement.

The descriptive statistics indicated that girls scored significantly higher than boys both in terms of their total test score on the BK1-1 and grade one reading achievement. Gender differences also emerged from the correlational analysis which established that correlations were generally stronger for girls than boys and that the BK1-1 was a significantly better predictor of grade three reading achievement for girls than boys. In terms of predictive effectiveness, the correlational analysis indicated that the BK1-1 was moderately predictive of academic ability and achievement for the total group in the early grades with correlations comparable to those for many existing school readiness tests. However, despite these moderate correlations, the BK1-1 was found to possess an extremely low level of predictive accuracy for making screening decisions about individual students. Finally, of the thirteen BK1-1 subtests only one, Recognition of Lowercase Letters, consistently emerged as a potentially valuable predictor of later academic achievement. The implications of these findings, their relationship to existing literature, and recommendations for both future practice and research in the field of school readiness testing were discussed.

ACKNOWLEDGEMENTS

The realization of this research study was made possible through the encouragement, assistance, and cooperation of many individuals. I wish to express my sincere thanks and appreciation to Dr. Lorraine Wilgosh for her excellent guidance and encouragement in supervising this thesis. I would also like to thank Dr. E. Conn-Blowers and Dr. Janis Blakey for their participation on my thesis committee.

I owe a special thanks to the parents of my research subjects for allowing me to obtain the necessary information from their childrens' records. I also wish to acknowledge the administrative, secretarial, and elementary teaching staff at the participating school for their pleasant and patient cooperation with me during all phases of this study.

I also wish to thank my parents, Jim and Bev Orieux, who have always encouraged me to pursue my dreams and whose love and moral support have been invaluable to me. Finally, I wish to express my very special thanks and love to my husband Keith for his patient support and understanding, his gracious tolerance of my compulsive work habits, and his assistance with many aspects of my graduate program.

Table of Contents

Page

Page

List of Tables

Table	Description	Page

List of Figures

CHAPTER I

Introduction and Statement of the Problem

At the turn of the twentieth century, Alfred Binet and his colleagues developed a series of tasks that could be used to identify children who would be unable to profit from regular instruction (Adelman, 1978; Kaplan & Saccuzo, 1982; Tsushima, Onorato, Okumura & Sue, 1983). Adelman (1978) stated that, "since that time, the topic of prediction and early identification of psycho-educational problems generally has been discussed in terms of its extreme importance and desirability" (p. 148). The term "early identification" is not limited to identification at an early age, but conceptually refers to the assessment of antecedent conditions predictive of later disorders, and can be applied to any age group and almost any disorder (Adelman, 1978). However, for the purpose of this investigation, the use of the term "early identification" will be limited in its application to the practice of school readiness testing that seeks to predict, prior to grade one, which children are likely to experience difficulty in school.

A review of the literature clearly indicates that, over the past two decades, there has been growing social and educational interest in and support for the early identification of children who are likely to experience academic difficulties (Adelman, 1978; Friedman, Fuerth & Forsythe, 1980; Hase, 1977; Horn & Packard, 1985; Keogh & Becker, 1973; Lewis, 1980; Lindsay & Wedell, 1982; Maitland, Nadeau & Nadeau, 1974; Raccioppi, 1982; Satz & Fletcher, 1979; Steinbauer & Heller, 1978; Telegdy, 1974a). Lewis (1980) and Lindsay and Wedell (1982) presented evidence indicative of this trend in England. In 1972 only 25% of local educational authorities (LEAs) were carrying out educational screening whereas in 1976, 47% of the LEAs were using educational screening procedures and another 27% of the LEAs were planning to implement them.

The widespread support for early identification practices rests on the assumption that through early identification, predicted academic difficulties may be minimized or prevented by prompt intervention (Evans & Ferguson, 1974; Friedman et al., 1980; Gullo, Clements & Robertson, 1984; Hase, 1977; Horn & Packard, 1985; Lindsay & Wedell, 1982; Mercer, Algozzine & Trifiletti, 1979a,1979b; Raccioppi, 1982; Swanson, Payne & Jackson, 1981; Telegdy, 1974a; Vacc, Vacc & Fogleman, 1987). According to Lindsay & Wedell (1982), there has been a "move away from attempts to identify children who are failing to acquire basic educational attainments. . . toward attempts to identify these children before their failure becomes apparent" (p. 212). Furthermore, it is also assumed that the early identification of learning problems will not only facilitate academic growth but prevent the occurrence of secondary emotional and behavioral disorders associated with children who have experienced repeated frustration and school failure (Horn & Packard, 1985; Raccioppi, 1982; Reinherz, 1977). Clearly, the emphasis is on the concept of early prevention rather than later remediation (Evans & Ferguson, 1974; Telegdy, 1974b).

The keystone in the development of effective early identification programs is the selection of appropriate and accurate screening instruments (Mercer et al., 1979a; Reinherz, 1977; Swanson et al., 1981). The practice of screening refers to the assessment of large groups of children with brief, low-cost procedures so that students with potential problems can be referred for more extensive evaluation and follow-up (Lewis, 1980; Paget & Nagle, 1986; Reinherz, 1977). Screening children for school readiness, at the end of kindergarten or beginning of grade one, routinely occurs in most school districts (Bremer, 1959; Flynn & Flynn, 1978; Keogh & Becker, 1973; Lindsay & Wedell, 1982; Maitland et al., 1974) and several instruments have been developed for this purpose (Klein, 1977). Maitland et al.(1974) have stated that the increased interest in early identification has been reflected in the number of readiness tests cited in the Mental Measurements Yearbook. To illustrate, in the sixth edition of the Mental Measurements Yearbook (Buros, 1965), only eight readiness tests were cited whereas in the Seventh Mental Measurements Yearbook (Buros, 1972), twenty-nine school readiness tests appeared.

The underlying assumption of instruments used for school readiness screening is that test performance is a valid predictor of future academic performance (Evans & Ferguson, 1974; Piersel & Kinsey, 1984; Telegdy, 1975). Standard 1.1 of the American Psychological Association's manual of 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). Because school readiness tests make inferences about future academic performance, evidence of predictive validity is of primary importance. Therefore, in order to determine whether or not a specific school readiness test is technically adequate and whether its findings are of practical value to educators, its predictive validity must be evaluated using longitudinal data. In other words, scores obtained by children on a readiness test must be compared with appropriate indices of academic outcomes at a later point in time (Adelman, 1978; Flynn & Flynn, 1978; Salvia & Ysseldyke, 1981).

Evidence of the predictive validity of many commonly used school readiness tests has been established by several studies (e.g., Gullo et al., 1984; Nagle, 1979; Nichta, Federici & Schuereger, 1982; Piersel & Kinsey, 1984; Telegdy, 1975; Vacc et al., 1987). Generally, these studies report correlations in the range of .40 - .75 between readiness test scores and the results of standardized academic achievement tests administered in grades one through three. Despite the fact that many school readiness tests exist, with moderate to substantial levels of predictive validity for screening purposes, new instruments continue to be developed. It would seem reasonable to assume that unless a newly developed instrument is able to either equal or improve on the levels of predictive validity reported for existing school readiness tests, its use as a screening device is not supported. However, the effectiveness of many of these new screening tests has not been systematically evaluated (Adelman, 1978; Klein, 1977; Lewis, 1980; Lindsay & Wedell, 1982; Lindquist, 1982; Tsushima et al., 1983; Wilson & Reichmuth, 1985). This is problematic because, according to Horn & Packard (1985), the effectiveness of early identification and intervention programs is necessarily limited by the reliability and validity of the screening measures used to predict which children are at risk for future academic problems. Piersel & Kinsey (1984) succinctly state that:

> The need for well-developed instruments in screening is not disputed. What is of concern is the frequent lack of adequate standardization procedures and follow-up validity studies to assist in defining the strengths and limitations of these assessment devices. (p. 921)

Therefore, before any newly developed school readiness test can be appropriately used, evidence of its validity for making initial screening decisions must be established. A case in point is a recently published instrument, the Brigance K and 1 Screen for Kindergarten and First Grade (BK1) (Brigance, 1982). The BK1 is described as a criterion-referenced screening test of readiness skills which are considered to be prerequisite to success in kindergarten and grade one. The BK1 consists of forms to assess both kindergarten and grade one readiness and attempts to provide an overall picture of a student's development in five areas: language development, motor ability, number skills, body awareness, and auditory and visual discrimination. For the purpose of the present investigation, discussion will be limited to the grade one form of the BK1 (BK1-1) for which a comprehensive description will be presented in the following chapter.

The BK1-1 manual states that scores from the test can be used to identify children requiring a more comprehensive evaluation, to help determine appropriate placement and grouping of students, and to assist teachers with planning appropriate programs for students. However, at the present time, the technical manual does not supply information to support these recommended uses of the test results and a review of the literature to date has not revealed empirical studies that have investigated the predictive validity of this instrument.

The lack of evidence related to the predictive validity of the BK1-1 is problematic because, as previously stated, evidence of predictive validity is a basic requirement for school readiness screening instruments. The BK1-1 has the potential to be a useful screening instrument because it is easily administered and cost-efficient. However, its ability to predict school success and to identify students with potential learning problems has not been established empirically. Helfeldt (1984), as well as Robinson and Kovacevich (1985) have concluded that formal validation procedures are critical to the appropriate use of the BK1-1 as a screening instrument.

Based on the previous discussion, a clear rationale exists for an investigation of the predictive validity of the BK1-1. However, the predictive validity of any school readiness test is not absolute and must be assessed relative to the specific school environment in which it will be ultimately applied for practical use (Abrahamson & Bell, 1979; Salvia & Ysseldyke, 1981). Tsushima et al. (1983) have stated that, "even when validity coefficients are available, a local re-validation is worthy of consideration because of the invariably unique features found within each school, such as variation of socioeconomic, cultural characteristics, teaching emphases, and peer pressures" (p. 665). In addition, Lichenstein (1982), as well as Maitland et al.(1974) have advocated the development and use of local norms and cut-off scores for school readiness screening devices in order to maximize the predictive accuracy of instruments. Therefore, the establishment of local norms and cut-off scores for the use of the BK1-1 in a specific school setting is warranted. The main objective of this study was to determine how effectively scores from the BK1-1, administered at the end of kindergarten or beginning of grade one, predict indices of academic ability and achievement on a longitudinal basis; that is, at the end of grades one, two, and three. Secondary objectives included investigating whether the predictive accuracy of the BK1-1 is affected by gender differences and determining the best combination of BK1-1 subtests for predicting indices of academic ability achievement at the end of grades one, two, and three. Finally, if the BK1-1 was found to be an acceptable predictor of academic achievement, local norms and cut-off scores were to be established for the school in which the research study was conducted.

This study is of practical significance to the field of assessment because it establishes preliminary evidence of predictive validity for a relatively new testing device. In addition, the results of this investigation will be of practical value to educators who may wish to use the BK1-1 for screening purposes in schools similar to that in which this study was conducted.

CHAPTER II

Review of the Literature

As stated in the previous chapter, the BK1-1 is both easily administered and cost-efficient and as such, may have potential as a useful school readiness screening device. However, its predictive validity is not currently supported by empirical evidence. Several school readiness tests with acceptable levels of predictive validity for making screening decisions already exist. Unless a new test, such as the BK1-1, can demonstrate either similar or superior predictive power, its use as a screening instrument is not supported.

The review of the literature which follows will develop the framework of knowledge necessary for evaluating the predictive validity of the BK1-1. This review will be divided into a discussion of the following topics: 1) the underlying rationale of school readiness tests; 2) the outcomes that school readiness tests predict; 3) the skills assessed by school readiness tests; 4) how the predictive validity of school readiness tests is appropriately assessed; 5) predictive validities of selected readiness tests, and; 6) a comprehensive description of the BK1-1.

The Rationale of School Readiness Testing

In order to understand the purpose of the BK1-1, the rationale underlying school readiness tests in general must be discussed. School readiness testing refers to the practice of assessing whether or not a child has reached a level of functioning that is adequate for entering school or for beginning academic instruction in order to predict which children are likely to experience difficulty in school (Mercer et al., 1979a; Lichenstein & Ireton, 1984). School readiness can be assessed in a number of ways including the use of chronological age (Dietz & Wilson, 1985; DiPasquale, Moule & Flewelling, 1980; May & Welch, 1986; Randel, Fry & Ralls 1977; Sweetland & De

Simone, 1987); kindergarten teacher ratings, ranging from simple ranking of children in terms of the teacher's prediction of later academic success or failure, to structured rating scales (Bolig & Fletcher, 1973; Eaves, Kendall & Crichton, 1974; Feshbach, Adelman & Fuller, 1974, 1977; Glazzard, 1977; Hall & Keogh, 1978; Keogh & Smith, 1970; Lessler & Bridges, 1973; Lindsay & Wedell, 1982; Stevenson, Parker, Wilkinson, Hegion & Fish, 1976; Telegdy, 1975); and psychometric batteries which combine performance on numerous variables and tests related to cognitive, neurological, and perceptual- and sensory-motor functioning (Badian, 1982; Barnes, 1985; de Hirsch, Jansky & Langford, 1966; Eaves et al., 1974; Feshbach et al., 1974, 1977; Satz & Friel, 1974; Satz, Taylor, Friel & Fletcher, 1978). However, for the purpose of this investigation , it is the assessment of school readiness through the use of a single, multidimensional school readiness test, administered at the end of kindergarten or at the beginning of grade one, that is to be investigated.

The basic assumption underlying multidimensional school readiness tests is that a child's test performance is predictive of future academic difficulties and that these predicted difficulties may be minimized through early intervention strategies (Evans & Ferguson, 1974; Flynn & Flynn, 1978; Friedman et al., 1980; Gullo et al., 1984; Hase, 1977; Horn & Packard, 1985; Lindsay & Wedell, 1982; Mercer et al., 1979a, 1979b; Swanson et al., 1981; Telegdy, 1974a; Vacc et al., 1987). This assumption is related to the concept of secondary prevention. However, in order to understand secondary prevention, it must first be distinguished from primary prevention.

Primary prevention refers to strategies that are implemented before disorders appear whereas secondary prevention refers to intervention efforts implemented soon after indicators of a disorder are detected (Silver, 1978; Van Evra, 1983). Therefore, primary prevention necessarily requires the identification of the etiological factors underlying a disorder. Once these factors are determined, intervention programs are implemented to prevent the disorders from initially occurring. Two primary prevention strategies are possible. The first is the "high risk strategy" where certain groups of people known to be particularly at risk for developing certain disorders are targetted for intervention. For example, the availability of prenatal diagnosis may be targetted at women over the age of 35 who have an increased risk of having babies with birth defects. The second primary prevention strategy involves broad social and educational programs aimed at the general population. Several successful examples of this primary prevention strategy are evident in a medical context. These include widespread childhood immunization programs to prevent diseases like polio, the fluoridation of drinking water and dental health programs to prevent tooth decay, and prenatal parental education programs which promote a healthy pregnancy which leads to the prevention of birth disorders related to factors such as poor maternal nutrition.

Some attempts have been made to implement primary prevention techniques in the educational domain. Programs such as Head Start used a high risk primary prevention strategy in an attempt to alter the effects of inappropriate stimulation, associated with a low socioeconomic status home environment, upon young children in order to prevent school failure. However, attempting to apply a primary prevention model to an educational context is often problematic because, unlike medical disorders, the etiology of learning problems is not usually clear (Fedoruk, 1988). Therefore, in education we are generally limited to secondary prevention which seeks to identify the early indicators of learning difficulties and thereby detect, independent of cause, children who are likely to have difficulty in school and to implement intervention strategies before failure occurs (Silver, 1978).

The secondary prevention approach consists of three stages: screening, diagnosis, and intervention. The concept of screening originated within a medical or disease model (Frankenburg, 1985; Keogh & Becker, 1973) which assumes that the condition to be identified already exists in the child, that the diagnosis implies a specific treatment, and that the sooner a treatment is implemented, the more successful it will be. There is also the belief that early treatment will prevent the development of, or minimize problems secondary to the disorder, such as confounding emotional disturbances related to a primarily physical disorder (Keogh and Becker, 1973). Several examples of successful secondary prevention efforts exist within a medical context. These include the early identification and dietary treatment of phenylketonuria (PKU), and the early detection and treatment of visual defects and hearing impairment (Frankenburg, 1985; Keogh & Becker, 1973; Lindsay & Wedell, 1982).

Applying the assumptions of the medical model to education, it was reasoned that the early identification and treatment of children who might experience difficulty in school would prevent or minimize these problems and be more effective and economical than later remediation (Lichenstein & Ireton, 1984; Lindsay & Wedell, 1982; Satz & Fletcher, 1979; Wilson & Reichmuth, 1985). In addition, it was reasoned that the early identification and treatment of learning problems would also minimize the development of secondary emotional and behavioral problems associated with school failure (Horn and Packard, 1985; Raccioppi, 1982; Reinherz, 1977). Therefore, in an educational context, screening involves testing a large group of students with brief, simple, and cost-effective measures in order to identify students who are sufficiently different ("at risk"), in either a positive or negative sense, from their age mates that they require special attention (Lichtenstein & Ireton, 1984; Paget & Nagle, 1986; Reinherz, 1977; Salvia & Ysseldyke, 1981). Based on the results of screening, children diagnosed as being "at risk" are referred for further diagnostic assessment. Following this stage, decisions regarding appropriate placement and/or intervention strategies are made.

The focus of school readiness tests is solely on the first stage of the secondary prevention process, although screening is appropriately followed by diagnosis and intervention (Hall & Keogh, 1978; Salvia & Yssledyke, 1981). The emphasis is on identifying kindergarten children who, on the basis of the absence or presence of certain indicators assessed by a readiness test, are likely to experience academic difficulty and who would benefit from either repeating kindergarten or receiving specialized learning assistance early in school. In addition, grade one readiness screening can also be used to identify students who are advanced and might benefit from the provision of enrichment experiences in grade one. Simply stated, the basic purposes of school readiness tests are as follows: 1) to predict who is ready/not ready for formal academic instruction and; 2) to predict which children might profit from remedial (or enrichment) education programs (Salvia and Ysseldyke, 1981).

However, as the word "prediction" implies, the process of identifying a child as being "at risk" for school problems involves making a future judgement, an act of prophecy that may be accurate or inaccurate. Based on a child's ability to perform certain tasks on a school readiness test, we are hypothesizing about future learning problems, not confirming that they presently exist (Adelman, 1978; Keogh & Becker, 1973). When a child is identified as being "at risk", what is being referred to is a higher than average probability that a problem will occur, not that the problem will definitely occur (Keogh & Daley, 1983). As the word "probability" indicates, the concept of risk is related to expectancies based on population statistics and will not yield an exact prediction for each individual case (Frankenburg, 1985). Therefore, saying that a child is "at risk" for academic difficulty, based on the results of a school readiness test, means that the child has scored similarly to children in the population who have later experienced difficulty in school.

Thus far in this discussion, the term "academic difficulty" has been used to describe the risk condition predicted by school readiness tests. However, this term must be more precisely defined if one is to understand the academic outcomes which school readiness tests generally predict. Furthermore, a specification of these academic outcomes will provide a standard against which scores from the BK1-1 can be compared in order to evaluate its predictive effectiveness.

Academic Outcomes Predicted By School Readiness Tests

A critical consideration in determining the effectiveness of school readiness tests is the definition of outcome goals or criteria against which predictions are based (Keogh & Becker, 1973). Educational outcomes can be conceptualized in a number of ways, both academic and affective. Academically, educational success or failure can be defined by performance on norm-referenced achievement tests, achievement consistent with measured intellectual potential, mastery of subject matter, teacher-assigned marks or grade retention. Affectively, success or failure in school may relate to a child's self concept, popularity and personal fulfillment, and to parental satisfaction with a child's performance (Keogh & Becker, 1973). However, for the purpose of this study, discussion will be limited to academic educational outcomes because it is these outcomes that are traditionally the focus of early remedial and enrichment programs.

The majority of studies which have evaluated the predictive validity of specific tests of school readiness have used scores from standardized achievement tests administered in grades one to three as criterion or outcome variables (Flynn & Flynn, 1978; Gullo et al., 1984; Hase, 1977; Lichtenstein, 1981; Lindquist, 1982; Nagle, 1979; Nichta et al., 1982; Piersel & Kinsey, 1984; Randel et al., 1977; Rubin, Balow, Dorle & Rosen, 1978; Swanson et al., 1981; Telegdy, 1975; Vacc et al., 1987). It would appear that the educational outcome most often predicted by measures of school readiness is performance on a standardized achievement test. Therefore, returning to the concept of risk, a child who scores poorly on a readiness test is "at risk" for below-average performance on a standardized test of academic achievement. Certainly, this reflects only one type of academic outcome but to the extent that the results of such an achievement test relate to other academic outcomes such as teacher grades, mastery of subject matter, and grade retention, this would appear to be a valid criterion.

To evaluate the predictive validity of the BK1-1 in the present study, the outcome variable of academic achievement was defined primarily in terms of scores on standardized achievement tests. However, teacher-assigned grades were also used to provide an additional index of academic achievement. It is recognized that teacher-assigned grades reflect not only academic achievement but a number of other factors such as student behavior (Doherty & Conolly, 1985; Kornblau & Keogh, 1980; Leiter & Brown, 1985). However, the reality exists that teacher-assigned grades and report card marks are a widely used and important measure of academic achievement in schools. Teacher grades are often the only index of academic achievement seen by parents and children. Additionally, a teacher's assessment of a child may provide a better representation of actual day-to-day academic functioning than a score from a single, standardized achievement test. From this perspective, teacher-assigned grades possess social validity; that is, real-world significance (Wolf, 1978), and in conjunction with standardized measures, they provide another angle for viewing students' academic achievement. Teacher estimates of academic achievement have been found to correlate in the range of .55 -.70 with standardized measures of academic attainment (Doherty & Conolly, 1985; Leiter & Brown, 1985; Lessler & Bridges, 1973) which reflects a moderate level of agreement . Therefore, in the present study, it is assumed that teacher grades will function as an acceptable criterion measure of academic acheivement to which scores from the BK1-1 can be compared.

In summary, it is reasonable to assume that for a school readiness test to be valid, it must assess skills that are predictive of the academic outcomes previously discussed. Based on the degree of absence or presence of the skills assessed by a readiness test, it is predicted which students have a greater than average chance of success of failure in later schooling. Therefore, these skills, the indicators of future academic achievement, must be identified so that the rationale underlying the content of school readiness tests, such as the BK1-1, can be understood.

The Skills Assessed by School Readiness Tests

Several specific group and individual multidimensional tests of school readiness exist but they share marked similarities in content (Lindsay & Wedell, 1982). Multidimensional school readiness tests consist of a number of subtests, which assess the absence or presence of indicators that have been found to be the most predictive of early school achievement. A number of empirical studies have investigated the early predictive indicators of school success and failure. Generally, these studies have assessed language, cognitive, motor, perceptual/sensory-motor, and behavioral-emotional variables in kindergarten or early grade one and correlated them with measures of school achievement in the early grades (Horn & Packard, 1985; Keogh & Becker, 1973). Variables having the highest correlation with school achievement have then been defined as providing the best early prediction of future academic status (Horn & Packard, 1985). Different combinations of these variables define the content of most school readiness tests. Following a description of these variables, their specific relationship to the content of the BK1-1 will be presented.

Language Skills

Language variables include written expressive, oral expressive, and receptive language skills. Written expressive language tasks require children to produce written language through such tasks as printing numbers or their own name. Tasks assessing oral expressive language skills involve having children produce spoken language by correctly articulating words, saying the alphabet, and identifying pictures, letters, colours, or numbers with a verbal response. Finally, receptive language skills require children to demonstrate their understanding of language with a non-verbal response. Tasks of this nature might include having a child identify which one of four pictures corresponds to a vocabulary word spoken by the examiner. Several studies have investigated the relationship between language variables and school achievement (Badian, 1982; Barnes, 1985; Collis, Collis & Yore, 1986; Flynn & Flynn, 1978; Freebody & Rust, 1985; Lindquist, 1982; Schmidt & Perino, 1985). Generally, the correlations between language variables and academic achievement fall in the range of .40-.65 with measures of oral language being weaker predictors than measures of receptive or written expressive language (Horn and Packard, 1985).

Several tasks assessing language variables are evident in the content of most school readiness tests. For example, the content of the BK1-1 reflects a heavy emphasis on language skills. Of the thirteen skill areas assessed by the test, nine reflect language skills. Written expressive language is assessed by having children print numbers in sequence to seven and by having them print their first and last names. Oral expressive language is assessed by having children orally answer personal data questions (e.g., When is your birthday?), name colors, pictures, lower case letters; recite the alphabet, and count by rote to twenty. Receptive language skills are assessed by one skill area in which children must non-verbally demonstrate how many objects correspond to a printed numeral.

Cognitive Skills

Overall intellectual functioning or IQ is considered to be a strong predictor of academic achievement (Keogh & Becker, 1973). Many studies have investigated the relationship between general cognitive ability and school achievement using group or individually administered IQ tests (Badian, 1982; Book, 1974, 1980; Feshbach et al., 1974, 1977; Flynn & Flynn, 1978), human figure drawing tasks (Dunleavy, Hansen, Szasz & Baade, 1981; Flynn & Flynn, 1978; Schmidt & Perino, 1985; Simner, 1985), and Piagetian tasks (Collis et al., 1986). Generally, the correlations between these measures of intellectual ability and later school achievement are in the range of .35 -.55 with individual tests of intelligence, such as the WPPSI, yielding the best predictions and Piagetian tasks yielding the weakest prediction. The cognitive ability task most often found on tests of school readiness is human figure drawing. The BK1-1 includes such a Draw-A-Person task, which is scored for the presence of ten features.

Fine and Gross Motor Skills

Fine motor skills, such as pencil and scissor use, and gross motor skills, such as hopping and balancing on one foot, have also been correlated with school achievement. Correlations between these variables are generally in the range of .23-.30 (Badian, 1982; Horn & Packard, 1985; Lindquist, 1982). Although fine and gross motor tasks are not highly predictive of later school achievement, they are occasionally included in the content of school readiness tests. However, the BK1-1 does not include a formal assessment of these skills.

Perceptual and Sensory-Motor Skills

Perceptual and sensory-motor variables include visual-motor skills, auditory perception ability, visual perception skills, and sensory integration ability. Visual-motor tasks assess children's ability to copy or draw shapes and patterns. Auditory perception activities generally require children to discriminate between similar sounding words or sound patterns while tasks assessing visual perceptual ability involve having children match similar letters, words, or figures or choose which of several figures, words, or letters in a group is dissimilar. Sensory integration activities require children to integrate two sensory modalities such as identifying which pattern of printed dots and dashes corresponds to a series of short and long tones. Most of the correlations between visual-motor tasks and early school achievement are close to .40 (Horn & Packard, 1985; Lesiak, 1984). In general, correlations between visual perceptual ability and academic achievement in the early grades are in the range of .36-.45 (Barnes, 1985; Collis et al., 1986; Horn & Packard, 1985), while Horn and Packard (1985) report mean correlations of .37 and .24 respectively for studies investigating the relationship between auditory discrimination skills, sensory integration skills, and early school achievement.

A number of perceptual and sensory-motor tasks are evident in the content of most school readiness tests, especially tasks related to visual-motor, visual perceptual, and auditory perceptual skills. Following language ability, perceptual and sensory-motor skills are the most thoroughly assessed area on the BK1-1. Visual-motor ability is assessed by having children copy shapes while the visual perception task requires children to choose which of four letters or words is dissimilar. Auditory perception skills are assessed using an auditory discrimination activity similar to the Wepman Auditory Discrimination Test (Wepman, 1973), in which children state whether two similar sounding words are the same or different.

Behavioral and Emotional Variables

Finally, the relationship of early behavioral-emotional variables to early school achievement has also been investigated. These variables are usually assessed using a teacher rating scale or checklist and include measures of attention and distractibility, measures of externalizing behavior problems, such as conduct disorders; measures of internalizing behavior problems, such as anxiety; and assessments of self-help, and social skills. Horn and Packard (1985) have reported an overall mean correlation of .48 for studies investigating the relationship between behavioral-emotional variables and early school achievement, with measures of attention-distractibility (mean r=.63) and internalizing behavior problems (mean r=.59) being the strongest predictors. Despite the relatively high predictive power of behavioral-emotional variables, these skills are not often included in the content of multidimensional school readiness tests. The BK1-1 does not contain a formal assessment of behavioral-emotional variables.

In summary, a number of different variables are moderately predictive of school achievement in the early grades. However, although each variable contributes to the prediction of school achievement, no single indicator in isolation provides adequate prediction (Silver, 1978). According to Horn and Packard (1985), "even the best early predictors account for less than 36% of the variance in subsequent school achievement" (p. 604). Therefore, in order to improve prediction, a number of predictive indicators are pooled to form the content of multidimensional school readiness tests such as the BK1-1. A multidimensional readiness test generally yields a total or composite readiness score, subtest scores, and, in some cases, area scores which are composed of the results of two or more subtests assessing a similar skill area.

Given a general understanding of the content of multidimensional school readiness tests, the question then arises: How effectively do the scores from these tests predict academic outcomes? Before answering this question, one must first be familiar with how the predictive validity of school readiness tests is appropriately assessed. This familiarity will provide the understanding necessary for evaluating the predictive validity of the BK1-1 in the present study.

Evaluating the Predictive Validity of School Readiness Tests

The predictive validity of school readiness tests is generally evaluated using a correlation coefficient which determines how well the specific readiness test (predictor variable) predicts academic achievement in the future, usually at the end of grade one (criterion or outcome variable)(Mercer et al., 1979b). However, although correlations indicate the strength of the linear relationship between predictor and criterion scores, a correlation coefficient does not give a direct indication of how accurate the readiness test is for making decisions (e.g., refer for further assessment versus do not refer) (Lessler & Bridges, 1973; Lichenstein, 1981; Mercer et al., 1979b).

As previously stated, readiness tests are essentially screening devices whose aim is to select children who are at risk for academic difficulty in the early grades. The goal of screening tests is to maximize the correspondence between the group of children referred and the group of children who actually experience academic difficulty (Lichenstein & Ireton, 1984). Therefore, in addition to coefficients of predictive validity, the prospective users of a readiness test also require validity data indicating whether the test selects children as intended; that is, they need validity data in the form of a classificational analysis which indicates the number of accurate versus inaccurate predictions that the readiness test makes (Lichenstein, 1981; Lindsay & Wedell, 1982; Mercer et al., 1979a; Satz & Fletcher, 1979; Wilson & Reichmuth, 1985). Predictive accuracy is assessed using a prediction-performance matrix (see Figure 1). This matrix summarizes the relationship between the outcomes of a readiness test and the actual status of individuals on a criterion measure, such as performance on a standardized achievement test.

Four outcomes are possible based on this matrix, two reflecting accurate screening decisions and two reflecting screening errors. Accurate screening decisions occur when a child who is designated as being at risk on the readiness test later experiences academic difficulty (valid positive) and when a child who is designated as being "not at risk" on the readiness test does not experience academic difficulty (valid negative). Similarly, screening errors occur when a child who is designated as being at risk on the readiness test does not experience academic difficulty (false positive) and when a child who is designated as being "not at risk" on the readiness test does

Status on criterion measure of school achievement

		Child is experiencing difficulty	Child is not experiencing difficulty
Screening outcome positive) based on readiness test performance	Refer; high risk for academic difficulty (+)	A. accurate prediction (valid positive)	C. inaccurate prediction (false
	Do not refer; low risk for academic difficulty (-)	B. inaccurate prediction (false negative)	D. accurate prediction (valid negative)

Figure 1: Prediction-Performance Matrix
Adapted from Lichenstein & Ireton, 1984, p. 197

experience academic difficulty (false negative). Additionally, a single index that summarizes the overall predictive accuracy of the readiness test is referred to as the "overall hit rate" and is an expression of the proportion of accurate screening decisions out of the total number of possible screening decisions (Mercer et al., 1979b; Wilson & Reichmuth, 1985).

When evaluating the results of a classificational analysis, the overall hit rate must be interpreted with caution. In order for a readiness test to be valid from the perspective of classificational analysis, it must demonstrate that it is able to identify and refer those children who will later experience academic difficulty, a quality referred to as sensitivity; and that it accurately selects those children who will not experience later academic difficulties, the quality of specificity (Lichenstein & Ireton, 1984). An overall hit rate can be misleading because it does not directly reflect a test's sensitivity. For example, assume that 100 children were administered a readiness test. On the basis of the readiness test, one child was identified as being "at risk" for academic difficulty in grade one. However, on a measure of academic achievement administered at the end of grade one, ten children, including the one initially identified, were found to be experiencing academic difficulty. Therefore, the readiness test failed to identify 90%

(9/10) of the children who later experienced problems. This occurred even though the overall hit rate of the readiness test would have been 91% (i.e., 91 correct screening decisions out of 100). Situations similar to the one just described occur because the overall hit rate for a screening instrument is highly dependent on the criterion measure base rate; that is, the proportion of children in the criterion measure problem group. The smaller the criterion measure base rate, the more likely it is that a screening outcome of "not at risk" will be correct (Lichenstein & Ireton, 1984; Wilson & Reichmuth, 1985).

The results of a prediction-performance matrix should also be considered horizontally. A horizontal analysis is often overlooked but is important because it determines the levels of predictive accuracy associated with "at risk" vs. "not at risk" screening outcomes on a predictor variable (Frankenburg, 1985; Mercer et al., 1979a, 1979b; Wilson & Reichmuth, 1985). In other words, a horizontal analysis calculates children's chances of actually falling in the "at risk" group on the criterion variable if they received an "at risk" designation on the predictor variable. Similarly, it determines their chances of actually being in the "not at risk" group on the criterion variable if they received a "not at risk" screening outcome.

In summary, the results of a classificational analysis of predictive accuracy are not absolute. Rather, they reflect the results obtained with a specific readiness test and criterion measure using certain predetermined cut-off scores to specify risk versus no risk on the readiness test and academic difficulty versus no academic difficulty on the criterion measure. Different results would be obtained if the cut-off scores for either measure were changed or if different tests were used (Lichenstein & Ireton, 1984; Mercer et al., 1979a; Wilson & Reichmuth, 1985).

School readiness tests, a type of screening procedure, are not expected to yield perfectly accurate prediction (Adelman, 1978; Silver, 1978). Errors of over- and under-identification (false positives and false negatives) are expected and prospective test users must carefully consider the implications of each type of error for their particular situation (Mercer et al., 1979a). When assessing the predictive accuracy of a readiness test, the consequences of screening errors must be considered in terms of both cost efficiency and ethical implications.

Errors of overidentification (i.e., false positives) are troublesome from the perspective of cost-efficiency because the individualized diagnostic assessment which ideally follows a screening decision of "at risk", is costly in terms of professional time. If a readiness test results in many overidentification errors, there is an unnecessary expense in terms of time and money to a school district. Ethically, errors of overidentification are serious when the outcome is retention in kindergarten or remedial intervention. Overidentification errors are problematic because they may result in undue anxiety for parents. In addition, the erroneous application of an "at risk" label may negatively affect both a teacher's expectations and a child's self-expectations. Applying the concept of self-fulfilling prophecy, attaching labels of educational risk to children, when no risk actually exists, may result in a teacher expecting less of the children and the children expecting less of themselves and as a result, the children performing as predicted (Adelman, 1978; Feuerstein, 1979; Keogh & Becker, 1973).

Similarly, errors of underidentification (i.e., false negatives) are also costly in terms of professional time and money. Children who are truly at risk but not identified as such, and who do not receive early assistance, may require more intensive, long-term remedial intervention at a later point in time. Additionally, from an ethical perspective, placing children in learning environments for which they are unprepared may result in their experiencing needless failure and frustration and may also contribute to the development of secondary problems such as reduced self-concept and an aversion to school (Reinherz, 1977).

As previously stated, the use of school readiness tests will result in some identification errors. Each school district must carefully weigh the implications of each type of error for its situation. However, if school readiness tests are treated solely as screening tools whose decisions are to be followed up by individual, diagnostic assessment, the negative implications of identification errors are greatly reduced, especially from an ethical perspective. Lichenstein (1981) has asserted that errors of overidentification are less serious than underidentification errors because they can be corrected at the stage of individualized diagnostic assessment that appropriately follows a screening procedure. From this perspective, a screening test should ideally refer for in-depth assessment, all children who are truly at risk for academic difficulties as well as a few overreferrals. Therefore, the only identification errors would be those of overidentification, and these children would likely be picked up during the following diagnostic assessment. The issue of overidentification errors on a readiness test then becomes primarily an economic question for school officials; that is, is the cost of providing individual, diagnostic assessment to a greater number of false positives reasonable if all children actually requiring specialized assistance are identified ?

The process of establishing evidence for the predictive validity of a readiness test is complex and should ideally involve a consideration of both predictive validity correlation coefficients and classificational accuracy. In the present study, the predictive validity of the BK1-1 will be evaluated using both correlational and classificational procedures. With these concepts in mind, the question: How effectively do the scores from multidimensional school readiness tests predict academic outcomes?, can be answered.

The Predictive Validity of Selected School Readiness Tests

Evidence for the predictive validity of selected multidimensional school readiness tests is presented in this section in order to provide a standard against which to compare the predictive validity of the BK1-1 which was established in the present study. Following a brief description of each test, the procedure and results of selected predictive validation studies related to that test are summarized. At the end of this section, a discussion of methodological weaknesses common to the studies reviewed is presented along with a consideration of how these weaknesses are addressed in the present study.

The Metropolitan Readiness Test

In the past, the Metropolitan Readiness Test (MRT) (Hildreth, Griffiths & McGauvran, 1965) has been found to be the most commonly used measure of school readiness (Maitland et al., 1974; Reynolds, 1979). The 1976 edition of the MRT (Nurss & McGauvran, 1976) enjoys similar widespread use (Gullo et al., 1984). The MRT is "intended to assess several important skills needed for early school success" (Salvia & Ysseldyke, 1981, p. 467). Two levels of the test are available, one intended for use with children at the beginning or middle of kindergarten and one for use with children at the end of kindergarten or beginning of first grade. For the purposes of this investigation, discussion will be limited to the latter level of the MRT. The second level of the MRT assesses receptive language skills (School Language and Listening subtests), knowledge of grapheme-phoneme correspondence (Beginning Consonants and Sound-Letter Correspondence), visual-discrimination skills (Visual Matching and Finding Patterns subtests), and number skills (Quantitative Concepts and Quantitative Operations subtests).

A factor analysis of the 1965 edition of the MRT revealed that it is most strongly oriented to visual-perceptual-motor and language comprehension skills (Telegdy, 1974a). Reynolds (1979) also factor analyzed the 1965 edition of the MRT. He concluded that a single factor, General Readiness, best describes the test and that the use and interpretation of the total test score is most appropriate for predicting early school achievement. The 1976 edition of the MRT yields scores for Auditory, Visual, Language, and Quantitative skill areas. The Auditory, Visual, and Language skill area scores are further combined into a Pre-reading Skills Composite score. However, similar to Reynold's study (1979), a factor analysis of the MRT (1976) only supported a general readiness factor, reflective of the total battery score, and a second factor felt to represent a language dimension, reflective of the Language skill area score. The analysis neither supported a differentiation between Auditory, Visual, and Quantitative skill areas nor a factor construct supporting the existence of a Pre-reading Skills Composite score that differs from the Total Battery score (Watkins & Wiebe, 1984). Based on the results of the study, it would appear best to limit interpretations of the MRT (1976) to a general measure of school readiness ability.

Telegdy (1975) assessed the predictive validity of the MRT (1965) in addition to that of four other readiness measures. The MRT was administered to 56 children (28 boys and 28 girls) at the end of kindergarten. Scores from the MRT were correlated with scores from the Wide Range Achievement Test (WRAT) (Jastak, Bijou & Jastak, 1965) and the Gray Oral Reading Test (GORT) (Gray, 1963), which were administered at the end of grade one. The MRT total score was more highly predictive of later academic achievement than any of its subtest scores. Correlations ($p<.01$) between the MRT total score and the WRAT Reading, Spelling, and Math subtests were .70, .73, and .59 respectively. In addition, the MRT correlated ($p<.01$).58 with the GORT passage reading subtest and .67 with the GORT literal comprehension subtest. However, the choice of criterion measures in this study is questionable for the purpose of assessing academic achievement. For example, the Reading subtest of the WRAT only assesses skills in letter recognition, letter naming, and pronunciation of words in isolation - a very limited assessment of reading ability.

Randel et al. (1977) conducted a study to compare the predictive validities of the MRT (1965) and ABC Inventory (ABC) (Adair & Blesch, 1965). Using zero order correlations and a stepwise multiple regression procedure, they assessed the ability of MRT total scores, obtained in the spring of kindergarten, to predict a total raw score obtained from a combination of three reading subtests of the Stanford Achievement Test (SAT) (Kelley, Madden, Gardner & Rudman, 1963,1964) at the end of grade one (N=62) and grade three (N=65). The MRT was the first variable to enter the multiple regression equation in both cases and although statistically significant (p<.01), the correlations between the MRT total score and SAT reading raw scores were substantially lower than those obtained by Telegdy (i.e., .34 with grade one SAT reading and .51 with grade three SAT reading).

Flynn and Flynn (1978) included the MRT (1965) in their evaluation of the predictive validities of five screening measures. The MRT was administered to 81 kindergarten children and the scores were correlated with scores from the California Achievement Test (CAT) (CTB/McGraw-Hill, 1977) administered at the end of grade two. Both zero order correlations and a stepwise multiple regression procedure, with the MRT entered as the first variable, were used to analyze the data. The MRT total score correlated (p<.01) .34, .32, .30, and .28 with CAT Reading, Language, Math, and Total test scores respectively. These results are similar to those obtained by Randel, Fry and Ralls (1977).

The predictive validity of the 1976 edition of the MRT was investigated by Nagle (1979) using a sample of 176 children (73 girls and 103 boys). The MRT was administered during the first month of grade one and the raw scores for the four MRT skill areas were correlated with the Total Reading and Total Math grade level scores on the SAT, which was administered at the end of grade one. Correlations were computed for the total group and separately by gender to determine whether the predictive validity of the MRT would be significantly different for males and females. The only significant difference between males and females was found on the correlation between the MRT Pre-reading Skills Composite score and the SAT Total Reading score (Z=2.02, p<.05) which was .75 for males and .57 for females (both p<.01). The remaining correlations (p<.01) for the total group were .65, .50, and .50 respectively, between SAT Total Reading and the MRT Auditory, Visual, and Language area scores. In addition, a correlation of .70 (p<.01) was found between the MRT Quantitative area score and the SAT Total Mathematics score.

Swanson et al. (1981) conducted a comparative study of the predictive effectiveness of the MRT(1976) and The Meeting Street School Screening Test (MSSST) (Hainsworth & Siqueland, 1969). They correlated the MRT scores of two chronologically consecutive samples of students (N=72 and 64) with reading scores on the Metropolitan Achievement Test (MAT)(Durost, Bixler, Wrightstone, Prescott & Balow, 1976) administered at the end of grade one. Correlations (group 1 listed first) of .745/.822, .755/.664, .702/.507, and .819/.835 respectively, were found between the MRT Auditory, Visual, Language, and Pre-reading Composite Skills area scores and the MAT Total Reading score (significance stated but no confidence level reported). However, these impressively high correlations may be slightly inflated due to the fact that the criterion variable, the MAT, is very similar to the predictor variable, the MRT (Fedoruk, 1988).

In a similar study, Gullo et al. (1984) administered the MRT to a sample of 88 children (40 girls and 48 boys) at the end of kindergarten. They correlated the MRT area scores with Reading, Language, Math, and Total Test scores on the Scott, Foresman Achievement Test (SFAT) (Wick & Smith, 1981). Correlations (p<.001) of .79 and .75 were found between the MRT Pre-reading Skills Composite score and the Reading and Total Test scores of the SFAT. In addition, the MRT Quantitative area score correlated

(p<.001) .52 with SFAT Math and .71 with the SFAT Total Test score. These results are highly comparable to those obtained by Nagle (1979) and Swanson et al. (1981).

Based on the results of the previously cited studies it is clear that considerable evidence has been established to support the predictive validity of the MRT. In fact, the predictive validity of the MRT is considered superior relative to other school readiness tests (Salvia & Ysseldyke, 1981) and it is often used as the standard against which to establish concurrent validity evidence for new readiness measures (Chew & Morris, 1984; Tshuima et al., 1983). However, several other devices exist for the assessment of school readiness and evidence of their validity for predicting school success will be examined in order to provide a broader basis of comparison for the predictive validity of the BK1-1.

The First Grade Screening Test

The First Grade Screening Test (FGST)(Pate & Webb, 1966) was developed for the purpose of ". . . screening beginning or potential first grade students to identify the few who will not, without benefit of special assistance, make sufficient progress during the first grade to be ready for second grade the next year" (Pate & Webb, 1966, p. 5). The test yields a single score and consists of 27 items which can be grouped into the following areas: Picture Vocabulary, Drawing and Visual-Motor Functioning, Social and Practical Perception of Information, Memory for Pictures, and Following Directions. A factor analysis of the FGST revealed that it is most strongly oriented to language comprehension skills; that is, a child's ability to comprehend and store verbal concepts. To a lesser degree, the FGST was also felt to measure visual-perceptual-motor ability (Telegdy, 1974a).

Telegdy (1975) evaluated the predictive validity of the FGST in addition to that of the MRT. The FGST was administered at the end of kindergarten to 56 children (28 boys and 28 girls). Their scores were correlated with scores from both the WRAT and the

GORT administered at the end of grade one. Correlations ($p<.01$) between the FGST and the WRAT Reading, Spelling, and Math subtests were .50, .53, and .59 respectively. In addition, the FGST correlated ($p<.01$) .40 with the GORT passage reading subtest and .53 with the GORT literal comprehension subtest.

Hase (1977) assessed the ability of the FGST, administered prior to grade one to a sample of 212 children (112 boys and 100 girls), to predict teacher ratings of the children in December of grade one. The FGST yields a single score but for the purpose of this study, the author rationally grouped the 27 items of the test into four subscales (Visual-Motor, Self-Concept, Reasoning, and Picture Vocabulary), in addition to the composite score, which was designated as reflecting overall school performance. During a pilot study, Hase determined cut-off scores for each of the five scales so that children could be classified as "not at risk", "mild risk", or "moderate to high risk" for problems in coping with school requirements based on their performance on the FGST. The students in the 1977 study had their FGST scores classified in this manner. First grade teachers were asked to rate their students as "good", "fair", or "poor" on each of the five scales in December of the year studied. Comparisons were then made between the teacher ratings and the FGST ratings. The percentage of agreement between the two sets of ratings ranged from 64% for Self-Concept to 70% for Visual-Motor. Chi-square statistics were calculated for the agreement rates and found to be significant ($p<.01$). Following this stage, children with total FGST scores of 15 or less and all children with at least one subscale score in the "moderate to high risk" range were identified as being "at risk" for academic difficulty. All remaining children were classified as "not at risk". This prediction was compared to teachers' ratings of overall school performance, which were collapsed into two categories, "adequate" (good and fair ratings) and "poor". A classificational analysis of the data revealed an overall hit rate of 82.5% (Chi-square statistic significant at $p<.01$). The FGST screening procedure also correctly identified 89% of the children that teachers felt were doing well in school and 74% of the children identified by teachers as experiencing difficulty in school. False positive and false negative rates of 7.5% and 3% respectively were obtained. These levels of agreement are similar to those in other studies reporting a classificational analysis of data (e.g., Wilson & Reichmuth, 1985). However, the predictive accuracy of "at risk"/"not at risk" ratings on the FGST was not considered through a horizontal analysis of the data in this study.

Piersel and Kinsey (1984) used a sample of 67 children (36 girls and 31 boys) to evaluate the predictive validity of the FGST, administered during March of the kindergarten year. The total score from the FGST was correlated with scores from the Science Research Associates Achievement Series (SRAAS) (Naslund, Thorpe, & Lefever, 1978) administered during February of grade one. Correlations ($p<.01$) of .44, .41, and .49 were found between the FGST total score and the SRAAS Reading, Mathematics, and Total Battery scores. These results are similar to those reported by Telegdy (1975).

The Screening Test of Academic Readiness

The Screening Test of Academic Readiness (STAR)(Ahr, 1966) consists of eight subtests: Human Figure Drawing, Picture Vocabulary, Letters, Picture Completion, Copying, Picture Description, Relationships, and Numbers. The STAR yields a composite score in addition to eight subtest scores. The STAR was factor analyzed by Telegdy (1974a). This analysis revealed three distinct factors: visual-perceptual-motor ability, language comprehension, and abstraction of essential characteristics; that is, the ability to identify and isolate relevant details.

Telegdy (1975) investigated the predictive validity of the STAR in addition to that of the FGST and MRT. The STAR was administered to a sample of 56 children (28 boys and 28 girls) at the end of kindergarten. Scores from the STAR were correlated with scores from the WRAT and the GORT, administered at the end of grade one. The STAR total test score and the Letters subtest score correlated ($p < .01$, total test score listed first) .67/.76, .67/.76, and .66/.63 with scores on the WRAT Reading, Spelling and Mathematics subtests; and .48/.62 and .62/.65 with the GORT passage reading and literal comprehension subtests.

Nichta et al. (1982) used a sample of 28 children (15 boys and 13 girls) to evaluate the predictive validity of the STAR, administered prior to kindergarten . Scores from the STAR were correlated with scores from the Peabody Individual Achievement Test (PIAT)(Dunn & Markwardt, 1970), administered during the middle of grade one. Correlations ($p<.01$) of .71 and .60 were found between the STAR total test score and the age and grade equivalent total test scores from the PIAT. However, it must be noted that the relatively small sample size in this study may have resulted in spuriously high correlations on the basis of chance. Despite this limitation, these correlations are similar to those obtained by Telegdy (1975) which suggests that the predictive validity of the STAR is not limited by when it is administered; that is, at the beginning or end of kindergarten.

Other School Readiness Tests

The Developmental Indicators for the Assessment of Learning (DIAL)(Mardell & Goldenberg, 1975) is a norm-referenced, individually administered preschool screening instrument designed to identify those children with potential learning problems. The DIAL consists of four areas assessing gross motor, fine motor, conceptual, and communication skills. In addition to a total DIAL score, scores are available for each of the four areas.

Vacc et al.(1987) used a sample of 245 children to investigate the ability of the DIAL, administered prior to kindergarten, to predict scores on the CAT at the end of grade one. The results of a canonical correlational analysis revealed that the four subtests of the DIAL were significantly related to the four CAT subtests (R=.67, p<.0001), indicating that the DIAL is a valid predictor of school performance as measured by the CAT. A stepwise multiple regression analysis was used to determine the most significant predictors of performance on the CAT subtests. In each case, the DIAL Concepts subtest was the first variable entered into the equation and was the most significant predictor of CAT scores, correlating (p<.0001) .35 with CAT Reading and .38 with the total CAT score. The DIAL Concepts subtest assesses skills such as classifying objects on the basis of color, shape, and size; counting, and naming colors. However, the strength of the obtained correlations is low, and the authors overinterpret their results when they state that these correlations indicate that successful prediction of first grade performance is heavily dependent on a student's abilities in these areas. In addition, despite the fact that the predictive validity coefficients were not strong, the authors state that the DIAL Concepts subtest is a meaningful preschool screening procedure that will enable educators to identify children who need early intervention, further evaluation and/or treatment, and that it "permits, on a long-term basis, the monitoring of children found to be at risk" (p. 50). In light of the obtained results, these claims are not justified.

The *Denver Developmental Screening Test* (DDST)(Frankenburg, Dodds, Fandal, Kazuk & Cohrs, 1975) is an individually administered, norm-referenced device designed for the early identification of children, birth to six years of age, with developmental and behavioral problems. The DDST assesses skills in four general areas: personal-social skills, fine motor development, language ability, and gross motor skills.

Lindquist (1982) evaluated the ability of the DDST to predict reading achievement, as measured by the *Gates-MacGinitie Reading Tests* (GMRT)(MacGinitie, 1978), at the end of grades one, two, and three. Twenty-seven items of the DDST were selected as being the most appropriate for assessing grade one readiness and were administered to children prior to grade one in order to obtain a total score in addition to scores for each of the four areas. The final sample of 351 children was randomly selected from five elementary schools and stratified to include approximately equal numbers of students from each of the first, second, and third grades. DDST scores were found to be most predictive of grade one reading achievement. The DDST total score correlated .46 (p<.15) with the total score on the grade one GMRT. The DDST Fine-Motor-Adaptive area, including skills such as copying and human figure drawing, was the most highly predictive subtest, correlating .41(p<.15) with the GMRT at the end of grade one.

The ABC Inventory (ABC)(Adair & Blesch, 1965) consists of four sections: a) drawing a man; b) answering general information questions; c) answering scientific information questions; and d) counting four squares, folding a paper triangle, repeating four digits, and copying a square. In addition to investigating the predictive validity of the MRT, Randel, Fry and Ralls (1977) also examined that of the ABC. Total scores from the ABC, administered at the end of kindergarten, were correlated with raw scores from selected reading subtests on the SAT which was administered at the end of grades one (N = 62) and three (N = 65). Correlations between the ABC and both grade one and three reading performance on the SAT were .19 and .23 respectively (no significance level reported). Based on the results of this study, the ABC does not appear to be a valid predictor of early school reading achievement.

The Meeting Street School Screening Test (MSSST) (Hainsworth & Siqueland, 1969) consists of three subtests: Motor-Patterning, Visual-Perceptual-Motor, and Language. Swanson, Payne, and Jackson (1981) evaluated the predictive validity of the MSSST in addition to that of the MRT. MSSST scores from two chronologically consecutive samples (N = 72 and 64) were correlated with reading scores from the MAT, administered at the end of grade one. Overall, the MSSST was not as strong a predictor of scores on the MAT as the MRT. The total MSSST score correlated .773 (group 1) and .726 (group 2) with the Total Reading score on the MAT (significance stated but no confidence level reported). These results indicate that the MSSST is a good predictor of grade one reading achievement as measured by the MAT.

Common Methodological Weaknesses in Predictive Validation Studies

An inspection of the studies reviewed reveals several common methodological weaknesses which are addressed in the present study. First, the intellectual ability (IQ) levels of the samples used were only described in two studies (Flynn & Flynn, 1978; Randel et al., 1977). It is possible that school readiness tests are not equally predictive for groups with different mean IQ levels. Jansky (1978) has stated that it would be useful to know whether readiness tests are equally predictive for children who differ in intelligence. However, without a knowledge of the mean IQ level of the samples used in predictive validation studies, this question cannot be answered and it is difficult to generalize the results of the studies to other groups of children who may have markedly different mean intellectual ability levels. Because of the importance of this variable, the mean intellectual ability level of the sample in the present study is described so that the generalizability of the findings regarding the predictive validity of the BK1-1 will be enhanced.

Second, socioeconomic status (SES) level is a highly influential subject characteristic that could potentially confound the relationship between the readiness test (predictor variable) and the measure of academic achievement (criterion variable) (Telegdy, 1974b). Jansky (1978) has stated that SES affects the predictive effectiveness of readiness tests. Several of the studies reviewed failed to describe the SES level of their sample (Flynn & Flynn, 1978; Hase, 1977; Lindquist, 1982; Nagle, 1979; Vacc et al., 1987). This seriously limits the possibility of generalizing the results of these studies to other groups of children. Five studies used samples whose SES level was very homogeneous (Gullo et al., 1984; Nichta et al., 1982; Piersel & Kinsey, 1984; Randel et al., 1977; Swanson et al., 1981). In these studies the effects of SES were eliminated because SES was constant for all subjects. However, the results of these studies are only relevant for groups of children with similar SES levels. Finally, Telegdy (1975) used a sample that was representative of low, middle, and upper SES groups, which greatly increases the generalizability of the results from the study. However, interaction was not addressed; that is, the possibility that the predictive effectiveness of the readiness test was not the same for different SES groups was not tested. In the present study, the SES level of the sample was described using the occupational status of the primary wage earner in each child's household so that the results regarding the predictive validity of the BK1-1 could be appropriately generalized to other groups of children with a similar SES level.

Third, although the composition of the samples in terms of gender was usually described, the possibility of differential predictions for males and females was tested in only one study (Nagle, 1979). This is problematic because Badian and Serwer (1975) and Jansky (1978) have suggested that different predictive measures are often needed for girls and boys. Therefore, in the present study, the possibility that the BK1-1 could be differentially predictive for girls and boys was tested.

Fourth, the majority of the studies reviewed limited their evaluation of predictive validity to correlational statistics. Only one study reported a classificational analysis of the data (Hase, 1977). According to Lichenstein (1981) classificational analysis is very important to evaluating the predictive validity of school readiness tests. Although predictive validity correlation coefficients are necessary and indicate the strength of the relationship between two variables, classificational analysis reflects the actual agreement between the readiness test and the measure of academic achievement in making category distinctions. As stated previously in this chapter, classificational analysis summarizes the proportion of accurate and inaccurate predictions that would be made with a specific instrument using specified cut-off points. This type of analysis is highly relevant to the evaluation of an instrument whose main purpose is one of intial decision-making and the assignment of children to categories (ie. "at risk" vs. "not at risk"). In the present study, the predictive validity of the BK1-1 was evaluated using both correlational and classificational analysis.

Finally, none of the reviewed studies described the instructional programs or curricula to which children in their samples were exposed during the early grades. Several investigators have acknowledged that the predictive validity of school readiness tests may vary according to the type of instructional program a child receives (Abrahamson & Bell, 1979; Feshbach et al., 1974; Jansky, 1978; Silver, 1978). Salvia & Ysseldyke (1981) have stated that, although a readiness test may predict a child's progress under one set of instructional conditions, it may not predict that child's progress in another instructional program. When predictive validation studies fail to describe the types of curricula and instructional programs that intervened between the predictions of the readiness test and the measures of academic achievement, it is difficult to directly generalize the findings to other school settings which may be using highly different teaching methods and materials. Therefore, in the present study, a description of the reading and math programs that intervened between the administration of the BK1-1 and the administration of the criterion variables was provided so that the generalizability of the obtained results would be enhanced.

In summary, there is a wide range of predictive validity coefficients for school readiness tests. The majority of the correlations fall between .40 and .75 with the MRT and STAR emerging as the best predictors of early school achievement. Additionally, studies that have reported a classificational analysis of the data generally report overall accuracy rates of 46 - 86% (Mercer et al., 1979a, 1979b). An overall accuracy rate of 75% is most common for multidimensional readiness tests (Hase, 1977; Silver, 1978; Wilson & Reichmuth, 1985). These results suggests that, in general, single, multidimensional readiness tests do have adequate predictive validity for making initial screening decisions. Several readiness tests with acceptable levels of predictive validity for screening purposes currently exist. Therefore, to gain support as a viable alternative to existing readiness tests, the BK1-1 must demonstrate either a similar or superior level of predictive validity and accuracy. However, a thorough examination of the BK1-1 is necessary prior to an investigation of its predictive validity.

The BK1-1

As stated in Chapter 1, the BK1-1 is described as a criterion-referenced screening test of readiness skills considered necessary for success in grade one. The device is based on the rationale that readiness cannot be appropriately described by a norm-referenced score because it is an ongoing process resulting from normal development and from instruction in and exposure to a continuum of various skills and activities. Additionally, the test manual states that the results of the criterion-referenced screening assessment of readiness skills provided by the BK1-1 can be directly translated into the development of individualized instructional programs.

The BK1-1 was developed in response to requests from school personnel who were informally conducting kindergarten and grade one screening using selected items from the Brigance Diagnostic Inventory of Early Development (DIED) and the Brigance Diagnostic Inventory of Basic Skills (DIBS). They expressed a need for a single instrument comprised of assessments from these two devices that would be the most appropriate for kindergarten and grade one screening. During the 1980-81 school year, an initial field testing version of the BK1-1 was developed and piloted in 53 schools in 14 states. Field personnel, including special and regular education teachers, primary specialists, and speech pathologists, evaluated this edition in terms of the significance and appropriateness of each assessment, the acceptability of point values assigned to each skill, and revisions, additions, and corrections which they felt were needed. Skills included in the final version of the BK1-1 were those that could be assessed quickly with a high degree of validity and objectivity and "having the greatest predictive validity for success in . . . grade one" (Brigance, 1982, p. iii) as supported by a review of the literature, field testing, and user requirements.

The BK1-1 assesses 13 skill areas (see Appendix A). The device is easily administered in 10 to 20 minutes and does not require specialized training in test adminstration, which makes it attractive for use by teachers. Although it is usually individually administered, the manual also suggests small group and station approaches to testing. The BK1-1 consists of a coilbound binder which contains the technical manual, the examiner's directions and test stimuli, a student data sheet for recording answers, and optional forms for screening observations and teacher and parent ratings. The manual is easily understood and clearly presents the directions, scoring criteria, time limits, and cut-off points for each item. Examiners are able to view the procedures for each item on their side of the binder while the child views the corresponding test stimuli, which are generally clear and attractive. In addition, the manual contains suggestions for possible observations and ways in which to "test the limits" of a child's skills in an area using unscored supplemental assessments.

Items are marked objectively and the number of correct responses in each skill area is multiplied by a weighting factor to arrive at an area score. The area scores are added to obtain a composite score with a possible total of 100. Following the assessment, examiners can record their observations on an optional screening observation checklist which addresses such areas as vision, hearing, speech and emotional functioning. The BK1-1 attempts to provide triarchic assessment by supplementing the child's skill performance with optional teacher and parent rating scales covering areas such as beginning academic skills, social skills, and health.

The manual suggests that the scores of all the children in a testing group be rank ordered and then divided into "High", "Average", and "Low" groups. It is stated that the BK1-1 may be given a second time, as a post-test, to students obtaining low scores. The manual indicates that the lower 20% of a group and any child scoring under 60 or over 95 should receive a comprehensive assessment. The test developer suggests that each school program establish its own cutoff scores for different purposes and referrals. This statement provides support for establishing local norms and cutoff scores for the BK1-1 in a school that is presently using the device.

The BK1-1 appears to have potential as a useful device for school readiness screening. However, at the present time, the effectiveness of the BK1-1 for making accurate screening decisions is not supported by empirical evidence. Wright (1985) issued a harsh criticism when he stated that the author of the BK1-1 seems to have absolved himself of the responsibility for ensuring the psychometric adequacy of his instrument by stating that his test is "criterion-referenced". Robinson and Kovacevich (1985) have even challenged the BK1-1's claim to being criterion-referenced. However, arguing these criticisms is not productive because it implies that the term "criterion-referenced" is somehow synonymous with "lack of psychometric adequacy". This is not a valid assumption because, according to the guidelines proposed by Hambleton and Eignor (1978) for evaluating criterion-referenced tests and test manuals, evidence of validity for the intended uses of test scores is an essential technical feature of criterion-referenced devices.

It will be recalled from Chapter I that the BK1-1 test manual states that the results of this device can be used to identify students requiring a more comprehensive evaluation, to help determine appropriate placement and grouping of students, and to assist teachers with planning appropriate programs for students. The BK1-1 manual also suggests that the test can be given in the spring to identify problems to be worked on over the summer. In addition, the manual states that "students of low ability can be expected to produce low scores" (Brigance, 1982, p. ix). The items on the BK1-1 appear to assess common pre-academic skills similar to those assessed by existing readiness tests. However, the validity of these items for achieving the outlined objectives of the test is only implied by user opinion and a review of related research (Boehm, 1985; Helfeldt, 1984) and not supported by empirical evidence (Robinson & Kovacevich, 1985).

Presently, the validity of the items selected for inclusion on the BK1-1 appears to have been inferred from their similarity to items on existing school readiness tests. However, Helfeldt (1984) has stated that the existence of items on a test "should not be inferred merely because a test's items were adopted from more comprehensive batteries or some norm-referenced measure" (p. 823). In other words, although a particular subtest or type of test item came from or is similar to one from a valid and reliable instrument, its reliability and validity as part of a new test cannot be assumed. For example, on the BK1-1 test page assessing Recognition of Lower Case Letters, there is a citation from Bond and Dykstra (1967) referring to the fact that letter name knowledge assessed by the Murphy-Durrell Reading Readiness Test (1965) is a strong predictor of grade one reading performance. This citation implies that the BK1-1 letter recognition task is also a valid predictor of grade one reading. However, the two tasks are highly different. The Murphy-Durrell task requires the child to select which of four or five choices corresponds to a letter named by the examiner whereas the BK1-1 task requires the child to supply the names to all twenty-six lower case letters of the alphabet. Therefore, there is a clear need to establish the predictive validity of the various subtests and items on the BK1-1.

In addition, the process by which items were selected for the BK1-1 appears to have relied heavily on the judgement of the author and field personnel whose criteria for determining item acceptability and weighting factors are not clearly outlined. Furthermore, the grade one curricula in the schools in which the usefulness of items was judged by field personnel is not described. Therefore, the degree to which the skills assessed by the BK1-1 are valid and relevant to early school success is dependent on the curricula of the grade one programs that children enter. This is supported by Salvia and Ysseldyke (1981) who state that "predictive validity may vary according to curriculum" (p. 455). Clearly, this provides additional support for evaluating the predictive validity of the BK1-1 within a specific school setting.

The BK1-1 possesses positive factors for its use as a screening device; that is, its administration is uncomplicated and it is economical both in terms of cost and time. However, based on the preceding discussion, its usefulness for predicting academic outcomes must be empirically established before its results can be appropriately interpreted and used as part of an early identification program. Therefore, the following research questions and objectives were investigated: 1) How effectively does the BK1-1, administered at the end of kindergarten or beginning of grade one, predict indices of academic achievement at the end of grades one to three? 2) Does the BK1-1 have differential predictive validity for boys and girls? 3) What is the best combination of subtests from the BK1-1 to predict indices of academic achievement at the end of grades one to three? 4) If the BK1-1 was found to be an acceptable predictor of academic achievement, local norms and cut-off scores for screening and referral purposes were to be established for the school in which the research study was conducted. Additionally, in order to clarify the exact nature of the reference group used in the present study, the sample was described in terms of mean intellectual ability level, socioeconomic status level, and the nature of the reading and math programs to which the subjects were exposed in grades one through three.

CHAPTER III

Method and Procedures

The following chapter will examine the technical aspects of this research study. The subjects and instrumentation will be described. This description will be followed by an outline of the data collection procedures and data analysis techniques that were applied to answer the research questions and meet the objectives outlined in the previous chapter.

Subjects

The total sample of 149 subjects (80 boys and 69 girls) was drawn from a single urban-rural elementary school where the researcher had been employed as a resource room teacher from 1984 to1987. The subjects were those children who were enrolled in grades one to five during the 1988 - 1989 school year and who had been administered the BK1-1 either the end of their kindergarten year (n = 114) or during their first month of grade one at the school (n = 35). Although 154 children were originally identified as potential research subjects, parental consent was required to obtain the information necessary for the study and 5 parents declined to have their children included. However, the final sample of 149 children represents 97% of the target student population.

The subjects ranged in age from 63 to 85 months with an average age of 72 months at the time of BK1-1 administration and had a mean intellectual ability level of 110 (range = 50 - 143) as indicated by their standard age scores (M = 100, SD = 16) on the Canadian Cognitive Abilities Test administered at the end of grade one.

The subjects were from a small urban centre (population = 6,000) and surrounding rural areas. A survey of socioeconomic status (SES) (see pp. 54 - 55 for a description of the procedure used) based on the occupation of the primary wage earner in the home indicated that 6% of the sample fell in the professional category, 27.5% of the children were from homes where the primary wage earner was employed in a skilled (non-manual) occupation, 50.3% of the sample fell in the skilled (manual) occupational category, 14.1% of the subjects were from homes of unskilled laborers, and 2% of the sample was from homes where the primary wage earner was unemployed.

One hundred and forty of the 149 subjects attended the English language Early Childhood Services (E.C.S.) program at the participating school. The remaining 9 children attended E.C.S. programs at other schools but had transferred into grade one at the participating school over the summer. Of the 140 subjects who attended kindergarten at John Paul II School, 51 children (kindergarten classes of 1984 and 1985) received their kindergarten instruction from one teacher while the remaining 89 children (kindergarten classes of 1986, 1987, and 1988) received their kindergarten instruction from a second teacher. At the time of data collection, 36 students (23 boys and 13 girls) were in grade one, 25 (13 boys and 12 girls) were in grade two, 33 (18 boys and 15 girls) were in grade three, 31(12 boys and 19 girls) were in grade four, and 24 (14 boys and 10 girls) were in grade five.

Reading and math instruction in grades one to three at the participating school is consistent with the goals and objectives outlined in the curriculum guide and programs of studies developed by the Alberta Department of Education and there is reasonable degree of consistency between the instructional approaches of the various teachers at each grade level. In addition, the composition of the teaching staff at each grade level has remained quite constant over the past five years. The Holt Mathematics System textbook series is used for instructing math in grades one to three. Reading instruction in grades one to three was based on the "Mr. Mugs" reading series published by Ginn until the end of the 1986-87 school year. However, since the 1987-88 school year, the Holt

Impressions reading series has been used as the basis of reading instruction in the primary grades.

Instrumentation

Due to the fact that this study was retrospective in nature, the instruments chosen for inclusion as predictor and criterion measures were predetermined by the school's existing testing program. A description of each instrument including its reliability, validity, and methods of administration and scoring follows.

Predictor Variable

The Brigance K & 1 Screen for Kindergarten and First Grade - Grade One Form (BK1-1) (Brigance, 1982)

As stated in the previous chapter, the BK1-1 is a school readiness screening measure that assesses thirteen skill areas (refer to Appendix A for a description): a) Personal Data Response (PDR), b) Color Recognition (CR), c) Picture Vocabulary (PV), d) Visual-Discrimination (VD), e) Visual-Motor Skills (VMS), f) Draw-A-Person (DAP) g) Rote Counting (RC), h) Recites Alphabet (RA), i) Numeral Comprehension (NC), j) Letter Recognition (LR), k) Auditory Discrimination (AD), l)Prints Personal Data (PPD), and m) Numerals in Sequence (NS).

The BK1-1 is individually administered in 10 to 20 minutes and does not require specialized training in assessment. The test items are in an easel format with clear administration directions for the examiner provided on the opposite side of each student page. The majority of items are objectively scored as correct or incorrect. However, certain subtests require judgement on the part of the examiner. The VMS subtest requires judgement of the acceptability of design reproductions but examples of acceptable and unacceptable designs are provided on the examiner's page. The DAP subtest requires some degree of judgement in determining which body parts children have included in their drawings. On both the PPD and NS subtests the examiner notes reversals but does not count them as errors. The examiner must also determine whether or not the student's printed numbers on the NS subtest are sufficiently representative of the intended symbols to obtain credit. However, lenient scoring is suggested.

The number of correct responses in each skill area is summed and multiplied by a pre-determined point value to obtain 13 skill area scores. The area scores are then added to arrive at a total score (see Appendix A). For the purpose of this study, both the total BK1-1 score and the 13 subtest scores were used in the analysis. In cases where children had repeated kindergarten (n = 5), their second set of BK1-1 scores was used in the analysis because it was felt to be more reflective of their actual entry level status into grade one.

In terms of reliability, Wright (1985) has argued that because the BK1-1 allows considerable latitude in administration, not all children will have comparable experiences. Inter-rater reliability has not been established for the BK1-1 to address the possibility of independent examiners arriving at different scores for one child on items requiring judgement. This makes the direct comparison and ranking of scores obtained by different children difficult unless a single examiner is used. In the present study, the researcher, who completed the testing as part of her resource room teaching duties, had directly tested 85% of the subjects (n = 126) while a trained assistant tested 15% (n = 23) of the subjects. The researcher rechecked and rescored all protocols for those children whom she did not directly test.

Criterion Variables

The following section measures were used as criterion variables in this study: a) Canadian Cognitive Abilites Test (CCAT); b) Gates-MacGinitie Reading Tests, Canadian Edition (GMRTCE); c) Canadian Achievement Tests (CAT), and; d) Teacher-assigned grades.

Canadian Cognitive Abilities Test (CCAT) - Primary Battery, Level 1, Form 3 (Thorndike & Hagen, 1982)

The CCAT, Primary Battery, Level 1 (CCAT-PB1) is a single score, untimed, non-reading group test of general cognitive ability. The test authors state that the CCAT-PB1 is "designed to assess the development of cognitive abilities related to verbal, quantitative and non-verbal reasoning, and problem solving" (Thorndike & Hagen, 1982, p. 3). The CCAT is the Canadian version of the Cognitive Abilities Test (CAT) (Thorndike & Hagen, 1978). Certain minor modifications in content were made to make the CCAT more relevant for a Canadian population and Canadian national norms were derived (McInnes, 1985).

The CCAT-PB1 is comprised of four subtests: a) Relational Concepts; b) Multimental; c) Quantitative Concepts, and; d) Oral Vocabulary. This test is usually administered by the classroom teacher who follows the standardized directions provided in the examiner's manual. Each item requires the children to choose which of four pictorial choices best corresponds with the teacher's statement or question.

The test booklets are handscored using a pictorial strip key. The raw scores for each subtest are added to produce a total test raw score which is then converted into a universal scale score (USS) using the specified table. The USS can then be converted into a grade percentile rank (GPR) using the child's grade at testing and the appropriate norms (i.e., fall, winter, or spring). The USS, in combination with the child's chronological age (CA) at testing, is transformed into a Standard Age Score (SAS, M=100, SD =16) using norms tables that are divided into three month chronological age intervals from 5-0 to 9-6. The SAS is roughly equivalent to an IQ score and can be converted into both a percentile rank and stanine score. For the purpose of this study, each child's SAS was used in the analysis.

The internal consistency of the CCAT-PB1 was determined using the KR_{20} procedure and is reported to be .88 (Thorndike & Hagen, 1984). This suggests an acceptable level of internal consistency. Test-retest reliability coefficients are not reported in the technical manual for the CCAT-PB1. However, Randhawa, Hunt, and Rawlyk (1974) reported a test-retest reliability coefficient of .75 for an earlier edition of the CCAT-PB (Level 1, Form 1). Although the content validity of the CCAT-PB1 appears to be acceptable (McInnes, 1985), concurrent and predictive criterion related evidence is not presented in the technical manual. However, Randhawa, Hunt, and Rawlyk (1974) found a correlation of .63 between scores on an earlier edition of the CCAT-PB (Level 1, Form 1) and WISC-R IQ scores. Finally, the CCAT-PB was standardized in 1980 on 32,137 Canadian students, concurrently with the Canadian Test of Basic Skills (CTBS) (King et al., 1982). The group on which the CCAT-PB1 was standardized was a stratified random sample of students from English-speaking Canadian schools from all provinces and territories. Therefore, the norms are believed to be representative of the subjects in the present study.

In the participating school, the CCAT-PB1 is routinely administered by the classroom teacher to all students at the end (i.e., May or June) of grade one. The researcher handscored the test booklets for all children in this study and converted the raw scores into standard age scores using spring norms (i.e., April 1 to June). The SAS scores from the CCAT-PB1 were used for two purposes in this study. First, they were used to establish a description of the mean intellectual ability level of the children included in the study. Second, they were used in the analysis to determine the relationship between BK1-1 scores and intellectual ability and to substantiate the test manual's statement that students of lower ability can be expected to produce lower scores on the BK1-1.

<u>Gates-MacGinitie Reading Tests, Canadian Edition (GMRTCE), Levels A, B, and C (MacGinitie, 1980)</u>

The GMRTCE is an untimed, norm-referenced, group administered test of reading ability. Levels A, B, and C are intended for use with grades 1, 2, and 3 respectively. Each level of the test contains two subtests, vocabulary and comprehension. The vocabulary subtest at levels A and B primarily assesses decoding skills. This subtest consists of pictures paired with four graphically similar words from which the child must choose the one that corresponds to the picture. At level C, the vocabulary subtest involves having a child match a target word with its synonym to be selected out of four or five choices, primarily emphasizing knowledge of word meanings (Stahl, 1985).

The comprehension subtest at Levels A and B requires the child to read a "story", ranging in length from one sentence to a short paragraph, and then to indicate which of four pictures the story describes. At Level C, the child must read a short paragraph of three to five sentences in length, and then answer a series of multiple choice questions about that selection.

The GMRTCE is usually administered by the classroom teacher who follows the scripted instructions provided in the teacher's manual. The test is handscored and yields scores for each of the subtests in addition to a total test score. Raw scores can be converted into *T*-scores (*M* =50, *SD* =10), percentile ranks, stanines, grade equivalent scores, or extended scale scores using fall (October), midyear (February), or spring (May) norms. In this study, raw scores and, in certain cases, *T*-scores for each subtest and the total test score were used as criterion variables in the analysis. In cases where children had repeated a grade, their first set of GMRT scores was used in the analysis because these scores were felt to be most reflective of the academic status predicted by the BK1-1.

are not provided for the total test score. However, the total test score can be assumed to be reliable because it is a composite of the two subtests each of which has high reliability. The content validity of the GMRTCE was determined through an investigation of the skills and subject matter presented in typical school reading programs. However, the test manual does not provide statistical evidence of how well scores on the GMRTCE relate to other measures of reading ability; that is, no evidence of concurrent validity is presented. The GMRTCE was standardized on a representative sample of 46,000 Canadian students and the characteristics of the GMRTCE standardization sample appear to be comparable to those of the subjects in the present study.

The GMRTCE is routinely administered at the cooperating school, by classroom teachers, to their students at some point during the school year, usually in May or June. Scores from the GMRTCE were used in this study to provide a standardized measure of reading achievement in grades one to three, to which results of the BK1-1 were compared. In addition, the GMRTCE (Level A) is the only standardized achievement measure routinely administered at the grade one level in the participating school.

Canadian Achievement Tests (CAT), Levels 12, and 13 Form A (CTC/McGraw-Hill, 1981)

The CAT is a timed, multiple-choice, group-administered, standardized test of academic achievement which assesses skills in four content areas: reading, spelling, language, and mathematics. Level 12 of the test is intended for use with children in grades one and two while Level 13 is used with grade three.

At Levels 12 and 13 the Reading area consists of four subtests assessing skills in phonic analysis (e.g., recognizing beginning consonant sounds), structural analysis (e.g., forming compound words), reading vocabulary (i.e., identifying synonyms and antonyms), and reading comprehension (i.e., short passages followed by multiple choice questions assessing literal and inferential comprehension). The Spelling area consists of a single subtest in which children must determine whether or not an underlined word in a sentence is spelled correctly. The Language area includes two subtests, Language Mechanics (i.e., capitalization and punctuation) and Language Expression (e.g., usage). The Mathematics area consists of two subtests, Mathematics Computation (e.g., addition and subraction) and Mathematics Concepts and Applications (i.e., applying mathematical operations to a variety of contexts such as measurement).

The CAT is usually administered by the classroom teacher who follows the instructions and time limits outlined in the examiner's manual. The CAT is computer-scored and yields raw scores, percentile ranks, grade equivalents, and stanines for each subtest, area (i.e., Reading, Spelling, Language, Math), and the total test score. In this study, the raw scores for each subtest, area, and the total test score were used as criterion variables in the analysis.

KR_{20} estimates of the internal consistency of the CAT indicate coefficients ranging from .69 for spelling at Level 12 to .97 for the Total Battery at Level 13. Internal reliability coefficients for the Total Reading, Total Math, and Total Battery scores at both levels of the CAT are equal to or greater than .91. Test-retest reliability data are not presented in the technical manual but the majority of test-retest reliability coefficients for the California Achievement Test (CAT) (CTB/McGraw-Hill, 1977, 1978), from which the CAT is developed, are reported to be in the range of .65-.75 (Salvia & Ysseldyke, 1981).

The content validity of the CAT was primarily established through the formation of category objectives which represent a category of skills. These category objectives were developed by reviewing provincial and school district curriculum guides, major textbooks, and the objectives of several criterion-referenced testing programs (CTC/McGraw-Hill, 1983). Items were then developed to measure these category objectives. However, the test developers assert that the ultimate evaluation of the CAT's validity is the degree to which it corresponds to a particular school district's objectives and curriculum.

The CAT was standardized on a total sample of 76,485 English-speaking Canadian students in grades one to twelve. The standardization involved a stratified random sampling of school districts by geogaphic area and degree of urbanization and random sampling of schools within districts. The socioeconomic status level of the sample was ascertained by questionnaire and is thoroughly described in the technical manual. The CAT norms are believed to be appropriate for use with the subjects in the present study.

The participating school began using the CAT with grade two in 1988 and with grade three in 1989. The test is routinely administered, by the classroom teacher, during May of each school year. Scores from the CAT were used in this study as a standardized measure of achievement at the grade two and three levels to which scores from the BK1-1 were compared.

Teacher-Assigned Grades

Year-end teacher-assigned marks for grades one to three from the 1987-1988 school year provided an additional index of academic achievement in this study. Report card marks for reading, written expression, printing, and mathematics from the fourth reporting period (end of June) were used. The participating school uses a letter grade system for grades one to three which is based on the following scale:

A - Outstanding (Excellent)

B - Above Average (Very Good)

C - Average (Good)

D - Fair

U - Unsatisfactory

This scale is based on teacher judgement and is not equated with percentages. In addition, teachers also utilize + and - ratings with the letter grades (i.e., A+). For the purpose of this study, the letter grades were converted into the following numerical scale:

A+ = 5.33	A = 5	A- = 4.66
B+ = 4.33	B = 4	B- = 3.66
C+ = 3.33	C = 3	C- = 2.66
D+ = 2.33	D = 2	D- = 1.66
	U = 1	

The choice of this continuous five-point scale was arbitrary and simply intended to express the qualitative letter grades in a numerical form which could be subjected to data analysis.

Data Collection

The data for this research study already existed in the cumulative records of the subjects (N = 149) at the participating school. However, the collection of data proceeded in two phases. The first phase involved students from the kindergarten classes of 1984 to 1987 (total possible n = 115, final n = 113) while the second phase involved children from the kindergarten class of 1988 (total possible n = 39, final n = 36). For the first phase of data collection, a written proposal (see covering letter in Appendix C) outlining the pertinent details of this research study was submitted to the superintendent of the participating school on January 6, 1989 in order to obtain permission to access data in the cumulative records of students who were registered in grades two to five at the school (i.e., cumulative records for children from the kindergarten classes of 1984, 1985, 1986, 1987). Permission was granted by the school board on January 9, 1989, subject to administrative guidelines which included clarifying whether or not parental consent would be required for accessing the information in the student records. Subsequently, on February 8, 1989, it was determined that it would be necessary to obtain written permission from the parents of the potential research subjects before accessing the information in their cumulative records (see correspondence in Appendix C).

In order to comply with these guidelines, a consent form, a letter describing the research study, and a covering letter from the superintendent (see Appendix C) were sent to the parents/ guardians of 115 potential subjects enrolled in grades two through five at the participating school on February 16, 1989. March 1, 1989 was specified as the deadline date for the return of completed consent forms. Following this deadline date, the researcher conducted a telephone follow-up of consents that had not been returned. Upon completion of this follow-up, 113 consents were returned while 2 parents declined to have their children included in the study. This represents a return rate of 98%.

Beginning on March 13, 1989, the cumulative records of these 113 subjects were reviewed. Children were assigned ID numbers starting with 001 and this number, rather than their name, was used on the data form in order to protect the confidentiality and anonymity of the information. All relevant data (e.g., demographic characteristics, test scores) were recorded onto a prepared data form (see Appendix B), transferred to a data matrix, keypunched, and then entered into the the University of Alberta computer system. In addition, the researcher rechecked the scoring of all BK1-1 test protocols for accuracy at this time.

The data collection procedure also involved obtaining information about each child's socioeconomic status level. To facilitate this process, the parents/guardians of each child were asked to indicate the occupation of the primary wage earner in their household on the consent form which was returned to the researcher. This information was recorded onto each child's data form using a five point occupational classification system which was operationally defined as follows:

5 Professional - any occupation for which university level education is essential and/or any legally recognized profession (i.e., recognized under Alberta provincial statute). (e.g., lawyer, doctor, nurse, teacher, engineer)

4 Skilled (non-manual) - primary non-manual occupations requiring specialized training and/or certification/licensing. (e.g., certified accountant, realtor, insurance broker, business manager)

3 Skilled (manual) - primarily manual occupations requiring specialized training and/or certification/licensing. (e.g., plumber, electrician, welder, hairdresser)

2 Unskilled - manual or non-manual occupations for which specialized training is not required or "on-the-job" training is provided (e.g., janitor, sales clerk, waitress)

1 Unemployed

Note: In cases where individuals are retired, they are considered under the occupational category in which they were last employed.

The researcher used these definitions to rate the occupational status of the primary wage earner in each subject's family and this numerical classification was used in the data analysis. The reliability of this socioeconomic status classification system was verified by having two additional individuals use the definitions to complete independent ratings of the subjects. An interrater reliability alpha coefficient (Cronbach, 1971) of .963 and a percentage rate of agreement of 81% between the three raters were obtained to confirm the accuracy of this classification system.

During the course of data collection, it was noted that report card marks were not recorded in the student records. The school principal was able to provide an administrative copy of report card marks for the 1987-1988 school year but it was determined that similar administrative copies of teacher-assigned marks for the 1984 - 1987 school years were not available. Therefore, this study was limited to the use of report card marks for the end of the 1988 school year only.

A preliminary inspection of the data collected during this first phase indicated that the number of subjects available for computing correlations between the BK1-1 and certain criterion variables was not as large as desired. Therefore, it was decided to expand the research sample to include children enrolled in grade one (i.e., kindergarten class of 1988) so that an additional set of scores from the BK1-1, grade one GMRTCE, and CCAT could be obtained. In addition, it was decided to extend the data collection period to the end of May, 1989 in order to obtain an additional set of scores from the GMRTCE at the ends of grades two and three and the CCAT at the grade two and three levels.

The decision to expand the sample to include those students enrolled in grade one was verbally approved by the superintendent and on April 3,1989 consent forms, a letter describing the research study, and a covering letter from the superintendent were sent to the parents/ guardians of 39 potential subjects (see Appendix C). April 10, 1989 was specified as the deadline date for the return of consent forms and, following this deadline, the researcher again conducted a telephone follow-up of those consents that had not been returned. Upon completion of this follow-up, 36 of the consents were returned while 3 parents declined to have their children included. This represents a 92% return rate at the grade one level and raised the total number of subjects to 149. Pertinent data from the cumulative records of the grade one students were then entered onto the prepared data form and transferred to a data matrix. The researcher then manually entered these data into the University of Alberta computer system.

At the beginning of May, 1989, the CAT was administered to students in grades two and three by their classroom teacher. The test protocols were computer-scored and when the students' results were available on June 2, 1989, the researcher transferred the relevant scores onto the data matrix and manually entered them into the University of Alberta computer system. Similarly, following the administration of the GMRTCE by the classroom teachers in grades one, two and three during the last two weeks of May, 1989, the researcher hand-scored all the test booklets, transferred the relevant scores to the data matrix, and manually entered them into the computer.

Data Analysis

This section will outline the statistical techniques that were applied to the data in order to answer the research questions and meet the objectives outlined in the previous chapter.

Descriptive statistics (i.e., N, M, SD, and range) were computed for all predictor and criterion variables using the SPSS^X statistical package. The descriptive statistics were computed for both the total group and seperately for boys and girls. Significant differences between means for boys and girls were calculated using a two-tailed t-test for independent groups.

Research Questions

1) How effectively does the BK1-1, administered at the end of kindergarten or beginning of grade one, predict indices of academic achievement at the end of grades one, two, and three?

This research question was examined in two ways:

a) Pearson product moment correlations (alpha=.05) were calculated between the total test score of the predictor variable, BK1-1, and scores on all criterion variables, (i.e., the CCAT, subtest and total test scores from the GMRTC and CAT; and report card marks) using the SPSS^X statistical package. It must be noted that different numbers of subjects were involved in calculating the various correlations . The correlations were computed for the total group and, where numbers were sufficient, separately for boys and girls.

The following descriptive classification system, proposed by Garrett (1954, p. 176), was used to discuss the strength of correlation coefficients in this study:

<.20	-	slight correlation
.20 to .40	-	low/weak correlation
.40 to .70	-	moderate correlation
.70 to .90	-	high correlation
.90 to .100	-	very high correlation

b) The predictive effectiveness of the BK1-1 was also investigated from the perspective of classificational analysis using the CROSSTABS function from the SPSSx statistical program. This produced a series of prediction-performance matrices similar to the one presented below.

Status on criterion measure of school achievement

		Difficulty	No Difficulty
Screening outcome based on BK1-1 performance	At risk for for academic difficulty (+)	A. accurate prediction (valid positive)	C. inaccurate prediction (false positive)
	Not at risk risk for academic difficulty (-)	B. inaccurate prediction (false negative)	D. accurate prediction (valid negative)

Figure 2: Prediction-Performance Matrix
Adapted from Lichenstein & Ireton, 1984, p. 197

Each matrix summarized the relationship between the outcomes of the BK1-1 total test score and the actual status of individuals on selected criterion measures at each grade level. The following five indices of predictive efficiency were calculated and reported for each matrix:

a) Overall Hit Rate $\left(\frac{A + D}{A + B + C + D}\right)$ - the proportion of accurate screening decisions out of the total number of possible screening decisions;

b) False Positive Rate $\left(\frac{C}{A + B + C + D}\right)$ - the proportion of individuals who were designated as "at risk on the BK1-1 but who did not experience later academic difficulty;

c) False Negative Rate $\left(\frac{B}{A + B + C + D}\right)$ - the proportion of individuals who were designated as "not at risk" on the BK1-1 but who later experienced academic difficulty;

d) Specificity Rate $\left(\frac{D}{C + D}\right)$ - the BK1-1's accuracy in selecting out children who did not experience academic difficulty at a later point in time;

e) Sensitivity Rate $\left(\frac{A}{A + B}\right)$ - the BK1-1's accuracy in selecting out children who later experienced academic difficulty.

Classificational analysis requires the establishment of cut-off scores to indicate "at risk" and "not at risk" categories on both the predictor (i.e., BK1-1 total test score) and criterion variables. No definitive or generally accepted cut-off point for indicating "at risk" status on an academic achievement outcome variable was evident in a review of the literature in this area. Therefore, for the purpose of this study, children were identified as experiencing academic difficulty (i.e., "at risk") if their score on a measure of academic achievement was less than or equal to the score falling at a point one standard deviation below the sample mean on that variable.

A similar procedure was used to determine "at risk" status on the predictor variable, the BK1-1. However, the predictive effectiveness of the BK1-1 was investigated using three different cut-off scores. The first cut-off score was one standard deviation below the sample mean on the total test score of the BK1-1 (i.e., BKIT1). A second cut-off score was designated as being one-half standard deviation below the BKIT sample mean (i.e., BKIT2) and a third cut-off score was the BKIT sample mean itself (i.e., BKIT3).

2) Does the BK1-1 have differential predictive validity for boys and girls?

In order to answer this research question, correlations for boys and girls between the BK1-1 total test score and selected criterion variables were tested for significant differences using a Z-test for comparing independent correlations (Glass & Hopkins, 1984, pp. 307 - 309). The following formula was used to calculate this statistic:

$$Z = \frac{Zr\ boys - Zr\ girls}{\left(\frac{1}{N\ boys - 1}\right) + \left(\frac{1}{N\ girls - 1}\right)} \qquad (\text{alpha} = .05,\ Z = 1.96)$$

3) What is the best combination of subtests from the BK1-1 to predict indices of academic achievement at the end of grades one to three?

This research question was addressed using the stepwise multiple regression analysis function of the SPSSx statistical package. This analysis proceeded in incremental steps. In the first step, the BK1-1 subtest which best predicted a specified criterion variable was chosen and a correlation was calculated. In the next step, the next most powerful predictor of the criterion variable was chosen based on the relationship between the predictor and the criterion, statistically controlling for the relationship between the criterion and the predictor chosen in the first step. These steps were continued until no additional variables were able to make a statistically significant (alpha=.05) contribution to the predictable variance of the criterion. At each step, an incremental R^2 was calculated to determine whether the variable entered at that step added to the predictable variance of the criterion over and above the predictors entered in previous steps. A significance test (*F*) was performed on each incremental R^2 to determine whether, in the population, that variable was likely to contribute significantly to the predictability of the criterion (Smith & Glass, 1987, pp. 215 - 216).

The analysis was conducted between the BK1-1 subtests and selected criterion variables at grades one, two, and three. Only criterion variables with sample sizes of seventy or more subjects were selected for inclusion in this analysis because, according to Smith and Glass (1987), when small sample sizes are used and when the ratio of the number of predictor variables (i.e., BK1-1 subtests) to the size of the sample is high, the results become less stable due to capitalization on chance (p. 216). To correct for possible inflated values of R^2, a correction for shrinkage was applied to the R^2. This correction basically reduced the magnitude of R^2 in accordance with the sample size and ratio of predictors to sample size (Smith & Glass, 1987, p. 216). Although this procedure is not a substitute for large sample sizes, it does help to limit misleading interpretations of the results.

Research Objective

To achieve the objective of establishing norms and appropriate cut-off scores for screening and referral purposes for the participating school it was determined that norms and cut-off points would only be developed if, on the basis of correlational and classificational analyses, the BK1-1 was found to possess an acceptable level of predictive validity for screening purposes. However, following an examination of the results obtained in response to the first research question, it was determined that it was not appropriate to pursue the development of local norms.

Although local norms for the BK1-1 were not developed, the following questions which were supplementary to the proposed norming process were addressed:

a) Do significant gender differences exist for the total test scores of the BK1-1?

The mean and standard deviation of the BK1-1 total test score was computed for both the total group (*N* = 149) and separately for boys and girls. A *t*-test for comparing independent groups was calculated (alpha=.05) to determine if significant gender differences existed.

b) Is BK1-1 test performance influenced by a child's chronological age at the time of test administration?

After surveying the chronological ages of the students at the time of BK1-1 administration, it was decided to subdivide the total group into five different age groups based on four month age intervals. A one-way factorial ANOVA was calculated to determine whether significant age differences existed in BK1-1 scores.

c) Does the time of BK1-1 test administration (i.e., end of kindergarten vs. beginning of grade one) affect BK1-1 test performance?

The mean of BK1-1 total test scores was calculated separately for students who were administered the BK1-1 at the end of kindergarten and those who were administered the BK1-1 at the beginning of grade one. A *t*-test for independent groups (alpha = .05) calculated to determine whether the time of BK1-1 administration had a significant effect on obtained scores.

CHAPTER IV

RESULTS

This chapter presents the results of the statistical procedures described in Chapter 3. An abbreviated listing of the predictor and criterion variables used in this study is included in Appendix D. The abbreviated form of each variable is used in the following presentation.

The first section of this chapter summarizes the descriptive statistics in this study. Following this, results pertaining to each of the three research questions and the research objective are presented.

Descriptive Statistics

A summary of the descriptive statistics (i.e., N, M, SD, and range) is presented in Table D.1. Means and standard deviations were calculated for the total group and separately by gender. Gender differences between means on all variables were calculated and are reported in Table D.2.

Significant gender differences between means were evident on 9 of the 69 variables and are reported in Table 1. An inspection of this table indicates that girls were significantly younger than boys at the time of BK1-1 administration. This can likely be attributed to the fact that all five children in the sample who repeated kindergarten were boys. As previously stated (see p. 46), in cases where children repeated kindergarten, their second set of BK1-1 scores was used in the analysis and this clearly would have elevated the mean age of the boys in the sample.

In terms of overall grade one readiness as measured by the total test score of the BK1-1 (BKIT), girls performed significantly better than boys. With respect to specific skills, girls outperformed boys on two tasks of alphabet knowledge; that is, reciting the alphabet (RA) and naming all twenty-six lower case letters (RLL). They also exhibited

Table 1

Significant Gender Differences Between Variable Means

Variable	Boys N	Boys M	Girls N	Girls M	T Value	p Value
CA	80	72.85	69	70.84	3.03	.003
PDR	80	3.66	69	4.03	-2.51	.013
RA	80	3.64	69	4.41	-3.03	.003
RLL	80	9.18	69	10.71	-3.39	.001
PPD	80	5.94	69	7.61	-4.27	<.001
BKIT	80	81.72	69	87.80	-3.67	<.001
GRC1	66	24.14	56	30.00	-4.54	<.001
GRT1	66	54.95	56	63.07	-3.26	.001
CATRLM2	30	13.33	27	15.56	-2.36	.022

significantly higher mean scores than boys on subtests that required them to print their first and last names (PPD) and to orally provide personal data (PDR) such as their birthdate and address.

In terms of grade one reading achievement as assessed by the Gates-MacGinite Reading Test (Level A, Form 1), girls performed significantly better on measures of both reading comprehension (GRC1) and overall reading ability (GRT1). A final significant difference between girls and boys was at the grade two level on the Language Mechanics subtest of the Canadian Achievement Test (CATRLM2) where girls had a higher mean score than boys.

Of note is the fact that none of the differences between boys and girls with respect to either grade one readiness or early academic achievement could be attributed to differential intellectual ability levels because a comparison of standard age scores from the Canadian Cognitive Abilities Test (CCAT) failed to reveal a significant difference between the two groups. It is also interesting to note that significant differences in academic achievement between boys and girls were virtually nonexistent past the grade one level.

Research Questions

1. How effectively does the BK1-1, administered at the end of kindergarten or beginning of grade one, predict indices of academic ability and achievement at the end of grades one, two, and three?

a) Correlational Analysis

Pearson product moment correlations between all predictor and criterion variables for both the total group and, where numbers were sufficient, separately by gender were calculated and are presented in Tables D.3 - D.11 in Appendix D. Correlations between the total test score of the BK1-1 (i.e., BKIT) and all criterion variables are reported in Tables 2 - 6 for the total group and, where appropriate, separately by gender. Only those correlations achieving a .05 level of significance or

Table 2

Pearson Correlation Coefficients: BK1-1 Total Test Score (BKIT) with Canadian Cognitive Abilities Test (CCAT) for Total Group, Boys, and Girls

	CCAT Total Group (N = 149)	CCAT Males (n = 80)	CCAT Females (n = 69)
BKIT	.523***	.578***	.439***

***$p \leq .001$

better will be discussed in this section. Specific significance levels for the correlations are supplied in the accompanying tables.

Table 2 presents correlations between the BKIT and standard age scores on the CCAT. The highest correlation between the BKIT and CCAT was .578 for boys. This was followed by correlations of .523 for the total group and .439 for girls. On the basis of these results, the BKIT appears to be moderately related to intellectual ability at the end of grade one as measured by the CCAT.

Correlations between the BKIT and scores on the Gates-MacGinitie Reading Test (GMRT) for the total group are reported in Table 3. It will be noted that correlations between the BKIT and both raw and *T*-scores are reported for the Grade 2 and 3 levels of the GMRT. The use of *T*-scores was necessary because several students had received the GMRT at different times in the school year in grades 2 and 3 which precluded a direct comparison of raw scores. Therefore, these students' *T*-scores were used because *T*-scores adjust raw scores, on the basis of time of test administration, to an equivalent scale for comparison. In addition, at the Grade 2 level of the GMRT it will be noted that there are twenty fewer subjects involved in the calculation of correlations for the comprehension and vocabulary subtests than for the total test score. This occurred because one teacher recorded only total test scores in the cumulative records for one class of grade 2 students.

An inspection of Table 3 for the total group indicates that the strongest correlations were found at the grade one level where BKIT correlated .610, .608, and .515 with GRC1, GRT1, and GRV1 respectively. This indicates that the BKIT was a moderate predictor of grade one reading achievement for the total group as measured by the GMRT. At the grade two level, correlations between BKIT and GMRT were not as strong as at the grade one level. However, BKIT remained a moderate predictor of reading achievement correlating .508 and .555 with GRC2 and GRT2 respectively. The downward trend in terms of the strength of correlations between BKIT and GMRT continued at the grade three level where none of the correlations exceeded .400. This indicates that that the BKIT was a weak predictor of grade three reading achievement for the total group.

Table 3 also presents the correlations between BKIT and scores on the GMRT for boys. Similar to the trend noted for the total group, the strongest correlations were found at the grade one level where BKIT was a moderate predictor of grade one reading achievement for boys correlating .552 with GRC1 and .552 with GRT1. Also consistent with the trend noted for the total group, there was a marked decline in the overall strength of correlations at the grade two and three levels. None of the correlations at either of these levels exceeded .350 and thus, none were statistically significant.

Correlations between the BKIT and scores from the GMRT for girls are also reported in Table 3. When compared with both boys and the total group, the overall frequency of moderate, statistically significant correlations is noticeably greater for girls. Also, unlike the trends noted for both the total group and boys, both the strength and number of statistically significant correlations remained relatively constant over all three grade levels. At the grade one level, the BKIT was a moderate predictor of reading achievement correlating .601, .569, and .549 with GRT1, GRV1, and GRC1 respectively. At the grade two level, the BKIT remained a moderate predictor of reading achievement with all correlations exceeding .500 and the highest correlation occurring with GRT2 (r = .643). Similarly, all correlations at the grade three level also exceeded .500 with the strongest relationship being between BKIT and GTT3 (r = .622).

In summary, the correlations between the BK1-1 total test score and the GMRT indicate that the BKIT is a moderate predictor of reading achievement at the grade one level for the total group as well as for boys and girls. The strength of this prediction decreased slightly at the grade two level for the total group, dropped substantially for boys, but remained relatively unchanged for girls. At the grade three level, the BKIT

Table 3

Pearson Correlation Coefficients: BK1-1 Total Test Score (BKIT) with Gates-MacGinitie Reading Test Scores for Total Group, Boys, and Girls

	BKIT (Total Group)		BKIT (Boys)		BKIT (Girls)	
	N	r	n	r	n	r
GRV1+	122	.515***	66	.473***	56	.569***
GRC1	122	.610***	66	.552***	56	.549***
GRT1	122	.608***	66	.551***	56	.601***
GRV2	33	.400**	16	.317	17	.571**
GRC2	33	.508***	16	.341	17	.594**
GRT2	53	.555***	23	.304	30	.643***
GRV3	57	.148	32	-.056	25	.518**
GRC3	57	.220*	32	-.0004	25	.549**
GRT3	57	.200	32	-.033	25	.557**
GTV2	58	.275*	31	.163	27	.525**
GTC2	58	.364**	31	.216	27	.512**
GTT2	78	.450***	38	.224	40	.595***
GTV3	86	.322***	44	.045	42	.555***
GTC3	86	.347**	44	.067	42	.577***
GTT3	86	.373***	44	.072	42	.622***

* $p \leq .05$
** $p \leq .01$
*** $p \leq .001$

+See Appendix D for a description of abbreviations.

remained a moderate predictor of reading achievement for girls but had limited predictive utility for both the total group and boys.

Table 4 presents the correlations between the BKIT and subtest, area, and total test scores on the Canadian Achievement Test (CAT) at the grade two level. An inspection of Table 4 for the total group indicates that the strongest correlations were .507 between BKIT and CATRP2 and .408 between BKIT and CATRR2. Similar to previously reported correlations with the GMRT (see Table 3), these results indicate that the BKIT is a moderate predictor of phonetic skills and overall reading achievement at the grade two level. However, the BKIT did not emerge as a very substantial predictor of spelling, language, math, or overall academic achievement at the end of grade two for the total group.

Table 4 also presents correlations between the BKIT and grade two CAT scores for boys. As with the total group, the strongest relationship was between BKIT and CATRP2 (r = .380). However, none of the remaining coefficients exceeded .271 thus failing to achieve statistical significance. Therefore, the BKIT was weak predictor of phonetic skills and an insignificant predictor of reading, spelling, language, math, and overall academic achievement for boys at the end of grade two.

Correlations between the BKIT and scores from the grade two CAT for girls are also presented in Table 4. Similar to the trend noted for both boys and the total group, the highest correlation was between BKIT and CATRP2 (r = .649). This was followed by a correlation of .511 between BKIT and CATRR2 and a correlation of .486 between BKIT and CATRB2. In addition, BKIT correlated .429 and .409 with CATRSP2 and CATRCO2 respectively. These results indicate that the BKIT is a moderate predictor of phonetic skills, overall reading achievement, spelling, computational skills, and general academic achievement at the end of grade two for girls.

In summary, the results of correlations between the BK1-1 total test score and scores from the CAT administered at the end of grade two indicate that the BKIT is a

Table 4

Pearson Correlation Coefficients: BK1-1 Total Test Score (BKIT) with Grade 2 Canadian Achievement Test (CAT) Scores for Total Group, Boys, and Girls

	BKIT - Total Group (N = 57)	BKIT - Boys (n = 30)	BKIT - Girls (n = 27)
CATRP2+	.507***	.380*	.649***
CATRS2	.241*	.112	.314
CATRV2	.290*	.170	.382*
CATRC2	.111	.007	.208
CATRR2	.408***	.269	.511**
CATRSP2	.291*	.243	.429*
CATRLM2	.234*	.085	.1[illegible]7
CATRLE2	.134	-.050	.347*
CATRL2	.239*	.036	.323*
CATRCO2	.323**	.271	.409*
CATRA2	.236**	.200	.309
CATRM2	.291*	.247	.380*
CATRB2	.387***	.253	.486**

* $p \leq .05$
** $p \leq .01$
*** $p \leq .001$

+See Appendix D for a description of abbreviations.

moderate predictor of phonetic skills and overall reading achievement for both the total group and girls but only a weak predictor of phonetic skills for boys. These results are generally consistent with the trends observed between the BKIT and GMRT where, although BKIT was a moderate predictor of reading achievement for the total group and girls, it was only a weak predictor for boys at the end of grade two. Overall, the strongest correlations between the BKIT and grade two CAT scores were for girls. Again, this is similar to the trend noted between the BKIT and GMRT where the strongest correlations occurred in the female sample. In addition, unlike the results for boys and the total group, the BKIT was also a moderate predictor of grade two CAT scores in spelling, mathematical computation, and overall academic achievement for girls.

Table 5 presents correlations between the BKIT and subtest, area, and total test scores on the CAT at the grade three level. The total number of boys (n=18) and girls (n=15) at this level was too small to obtain reliable, stable correlation coefficents because extremely small sample sizes tend to artificially inflate coefficients (Smith & Glass, 1987, p. 216). Therefore, only results for the total group will be discussed in this section.

An inspection of Table 5 indicates that the strongest correlations were .496, .482, and .466 between the BKIT and CATRR3, CATRB3, and CATRP3 respectively. Similar to correlations between the BKIT and CAT at the grade two level (see Table 3), these results indicate that the BKIT was also a moderate predictor of phonetic skills and overall reading achievement at the end of grade three. However, unlike the results at the grade two level, BKIT was also moderately predictive of spelling, language expression, and overall academic achievement at the end of grade three for the total group.

Table 6 presents correlations between the BKIT and teacher-assigned report card marks in grades one through three for the total group. Results are not reported separately by gender because of the extremely small sample sizes involved. At the grade one level, the BKIT emerged as a moderate predictor of both reading (r = .400) and math

Table 5

Pearson Correlation Coefficients: BK1-1 Total Test Score (BKIT) with Grade 3 Canadian Achievement Test (CAT) Scores for Total Group

	BKIT - Total Group (N = 32)
CATRP3+	.466**
CATRS3	.278
CATRV3	.273
CATRC3	.320*
CATRR3	.496**
CATRSP3	.437**
CATRLM3	.107
CATRLE3	.433**
CATRL3	.390**
CATRCO3	.061
CATRA3	.397**
CATRM3	.250
CATRB3	.482**

* $p \leq .05$
** $p \leq .01$
*** $p \leq .001$

+See Appendix D for a description of abbreviations.

Table 6

Pearson Correlation Coefficients: BK1-1 Total Test Score (BKIT) with Report Card Marks by Grade for Total Group

	BKIT - Grade One (N = 24)	BKIT - Grade Two (N = 33)	BKIT - Grade Three (N = 30)
Reading	.400*	.286	.521**
Written Expression	.307	.327*	.536***
Printing	.161	.060	.519**
Math	.537**	.290*	.427**

* $p \leq .05$
** $p \leq .01$
*** $p \leq .001$

(r = .537) achievement. However, this predictive strength was markedly less at the grade two level where the BKIT was only a weak predictor of written expression (r = .327) and math skills (r = .290). The relationship between the BKIT and teacher-assigned marks improved at grade three. At this level, the BKIT correlated .536 , .521, .519, and .427 with written expression, reading, printing, and math respectively. These results indicate that the BKIT was a moderate predictor of academic achievement, as determined by the classroom teacher, at the grade three level.

b) Classificational Analysis

A series of prediction-performance matrices was used to investigate the relationship between the outcomes of the BK1-1 and the actual status of children on selected indices of academic ability and achievement. However, unlike correlational analysis where coefficients approach stable values with sample sizes of 30 to 40 subjects, relatively larger numbers are necessary to establish a stable pattern of results using classificational analysis (Lichenstein, 1981). Therefore, although classificational analyses were calculated for several criterion measures in this study, only results for those with sample sizes of fifty or more subjects are discussed in detail because they are felt to reflect the most stable and reliable patterns. These results are presented in Tables 7 - 19. Additional classificational analyses for smaller sample sizes (i.e., 33 - 49 subjects) are presented as supplementary information in Tables D.12 - D.22.

Table 7 presents the relationship between outcomes of the BK1-1 total test score (i.e., BKIT) and the general intellectual ability level of children as measured by the Canadian Cognitive Abilities Test (CCAT) at the end of grade one. Section (a) of Table 7 presents the results using a cutoff point of one standard deviation below the sample mean on the BKIT (i.e., BKIT1). This matrix indicates impressively low false positive (8%) and false negative (9%)rates and impressively high specificity (91%), and overall hit (83%) rates. However, out of the 21 children who were actually below-average in

Table 7

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade One Canadian Cognitive Abilities Test (CCAT) Standard Age Scores for the Total Group (N = 149)

a)

		CCAT		
		Below Average Ability (< 94)	Average to Above Average Ability (> 95)	Row Total
BKIT1	At risk (≤ 74.5)	7	12	19
	Not at risk (≥ 75)	14	116	130
	Column Total	21	128	149

Overall Hit Rate: 123/149 = 83% False Positive Rate: 12/149 = 8% False Negative Rate: 14/149 = 9%
Specificity Rate: 116/128 = 91% Sensitivity Rate: 7/21 = 33%

b)

		CCAT		
		Below Average Ability (< 94)	Average to Above Average Ability (> 95)	Row Total
BKIT2	At risk (≤ 79.5)	10	23	33
	Not at risk (≥ 80)	11	105	116
	Column Total	21	128	149

Overall Hit Rate: 115/149 = 77% False Positive Rate: 23/149 = 15% False Negative Rate: 11/149 = 7%
Specificity Rate: 105/128 = 82% Sensitivity Rate: 10/21 = 48%

c)

		CCAT		
		Below Average Ability (< 94)	Average to Above Average Ability (> 95)	Row Total
BKIT3	At risk (≤ 84.5)	15	50	65
	Not at risk (≥ 85)	6	78	84
	Column Total	21	128	149

Overall Hit Rate: 93/149 = 62% False Positive Rate: 50/149 = 34% False Negative Rate: 6/149 = 4%
Specificity Rate: 78/128 = 61% Sensitivity Rate: 15/21 = 71%

Note: All percentages rounded to the nearest whole number.

terms of intellectual ability, only 7 were identified by the BK1-1 (i.e., Sensitivity rate = 33%) using this cut-off point. Section (b) of Table 7 presents the results of the classificational analysis using the second BKIT cut-off score which was one-half standard deviation below the sample mean (i.e., BKIT2). Use of this cut-off point reduced the overall hit rate to 77% but resulted in correct identification of 48% of the children who later fell below-average on the CCAT. Section (c) of Table 7 reports classificational analysis results using the third cut-off point which was equal to the BKIT sample mean (i.e., BKIT3). Using BKIT3 as a cut-off score resulted in a relatively impressive sensitivity rate of 71%; that is, of the 21 children who later fell below-average in terms of intellectual functioning on the CCAT, 15 were identified by the BK1-1. However, in order to achieve this level of sensitivity, almost half (i.e, 65/149 = 44%) of the subjects had to be identified as being "at risk" on the BK1-1. The information in Table 7(c) can also be considered horizontally; that is, of the 65 children that the BK1-1 identified as being "at risk", only 15 (23%) were actually below-average on the CCAT while 50 (77%) fell in the average to above-average range on the CCAT. Therefore, the efficiency of the "at risk" screening outcome was only 23%. In contrast, of the 84 children identified as being "not at risk" on the BK1-1, 78 actually fell in the average to above-average range on the CCAT. This respresents a predictive accuracy rate of 93% for the "not at risk" screening outcome using BKIT3 as a cutoff point.

It will be recalled from the discussion of the descriptive statistics in this study that there was a significant gender difference on the BK1-1 total test score. Therefore, to determine whether the level of predictive accuracy observed for the total group could be improved, it was decided to recalculate the classificational analysis between the BKIT and CCAT separately for boys and girls.

Table 8 presents the relationship between the BKIT and the CCAT for boys. Generally, the pattern of results is quite similar to that observed for the total group.

Table 8

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade One Canadian Cognitive Abilities Test (CCAT) Standard Age Scores for Boys (n = 80)

a) CCAT

BKIT1		Below Average Ability (≤ 94)	Average to Above Average Ability (≥ 95)	Row Total
	At risk (≤ 71.5)	5	7	12
	Not at risk (≥ 72)	9	59	68
	Column Total	14	66	80

Overall Hit Rate: 64/80 = 80%
Specificity Rate: 59/66 = 89%
False Positive Rate: 7/80 = 9%
Sensitivity Rate: 5/14 = 36%
False Negative Rate: 9/80 = 11%

b) CCAT

BKIT2		Below Average Ability (≤ 94)	Average to Above Average Ability (≥ 95)	Row Total
	At risk (≤ 76.5)	6	12	18
	Not at risk (≥ 77)	8	54	62
	Column Total	14	66	80

Overall Hit Rate: 60/80 = 75%
Specificity Rate: 54/66 = 82%
False Positive Rate: 12/80 = 15%
Sensitivity Rate: 6/14 = 43%
False Negative Rate: 8/80 = 10%

c) CCAT

BKIT3		Below Average Ability (≤ 94)	Average to Above Average Ability (≥ 95)	Row Total
	At risk (≤ 82.5)	10	23	33
	Not at risk (≥ 83)	4	43	47
	Column Total	14	66	80

Overall Hit Rate: 53/80 = 66%
Specificity Rate: 43/66 = 65%
False Positive Rate: 23/80 = 29%
Sensitivity Rate: 10/14 = 71%
False Negative Rate: 4/80 = 5%

Note: All percentages rounded to the nearest whole number.

That is, using the highest cutoff score (i.e., BKIT3), 10 out of the 14 (71%) boys who later fell below-average on the CCAT were correctly identified by the BK1-1. However, to achieve this level of sensitivity, 41% of the sample had to be initially identified as being "at risk" on the BK1-1. A horizontal consideration of the data in Table 8(c) indicates that a screening outcome of "at risk" on the BK1-1 was only accurate in 30% of the cases (i.e., 10/33) whereas a screening outcome of "not at risk" using BKIT3 was 92% (i.e., 43/47) accurate.

Table 9 presents the results of the classificational analysis between the BK1-1 and CCAT for girls. The optimal cut-off point for girls was different than that for both the total group and boys; that is, one-half rather than one full standard deviation below the sample mean. The use of this cutoff point (i.e., BKIT2) resulted in an overall hit rate of 81%, a false negative rate of 4%, and a sensitivity rate of 57%. In addition, only 20% (i.e., 14/69) of the sample had to be identified as being "at risk" to achieve this sensitivity rate. A horizontal analysis of the data in Table 9(b) indicates that a screening outcome of "at risk" was only accurate for 4 out of 14 cases (29%). However, similar to both boys and the total group, the screening outcome of "not at risk" was highly efficient - accurate in 52 out of 55 cases (95%).

In summary, considering the results of the classificational analysis between the BK1-1 and CCAT separately by gender did not yield a marked improvement over the results obtained for the total group. Although the overall accuracy of an "at risk" screening outcome was slightly better for boys (30%) and girls (29%) than for the total group (23%), none of these accuracy rates reflect a very efficient screening procedure. In other words, the BK1-1 is not a very accurate predictor of which children will actually fall in the below-average classification of intellectual ability as measured by the CCAT at the end of grade one. However, it must also be noted that in all three situations, a screening outcome of "not at risk" was highly accurate (i.e., 92 - 95%). Therefore, if children are identified as being "not at risk" on the BK1-1, their

Table 9

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade One Canadian Cognitive Abilities Test (CCAT) Standard Age Scores for Girls (n = 69)

a)

		CCAT		
		Below Average Ability (< 94)	Average to Above Average Ability (> 95)	Row Total
BKIT1	At risk (≤ 78.5)	2	4	6
	Not at risk (≥ 79)	5	58	63
	Column Total	7	62	69

Overall Hit Rate: 60/69 = 87% False Positive Rate: 4/69 = 6% False Negative Rate: 5/69 = 7%
Specificity Rate: 58/62 = 94% Sensitivity Rate: 2/7 = 29%

b)

		CCAT		
		Below Average Ability (< 94)	Average to Above Average Ability (> 95)	Row Total
BKIT2	At risk (≤ 82.5)	4	10	14
	Not at risk (≥ 83)	3	52	55
	Column Total	7	62	69

Overall Hit Rate: 56/69 = 81% False Positive Rate: 10/69 = 15% False Negative Rate: 3/69 = 4%
Specificity Rate: 52/62 = 84% Sensitivity Rate: 4/7 = 57%

c)

		CCAT		
		Below Average Ability (< 94)	Average to Above Average Ability (> 95)	Row Total
BKIT3	At risk (≤ 87.5)	4	21	25
	Not at risk (≥ 88)	3	41	44
	Column Total	7	62	69

Overall Hit Rate: 5/69 = 65% False Positive Rate: 21/69 = 30% False Negative Rate: 3/69 = 4%
Specificity Rate: 41/62 = 66% Sensitivity Rate: 4/7 = 57%

Note: All percentages rounded to the nearest whole number.

chances of performing in the average to above-average range of intellectual ability on the CCAT are approximately 9 in 10.

Table 10 presents the results of the classificational analysis between the BK1-1 and the total test raw score of the Gates-MacGinitie Reading Test (GMRT) administered at the end of grade one for the total group. Section (a) of Table 10 presents the results using the lowest cutoff score. Use of this cut-off point yielded acceptable false positive (5%), specificity (94%), and overall hit (82%) rates. However, using BKIT1 as a cutoff score also resulted in a low sensitivity rate (i.e., 27%) which translated into a failure to identify 73% of the children who actually experienced reading difficulty at the end of grade one as measured by the GMRT. Using the next highest cutoff score (i.e., BKIT2) improved the sensitivity rate to 46% but still resulted in a failure to identify more than half of the students who had actual reading difficulty. Using the highest cutoff point (i.e., BKIT3), resulted in the correct identification of 82% of the children who later experienced reading difficulty. However, in order to achieve this impressive sensitivity rate, 40% of the children (i.e., 49/122) had to be initially designated as being "at risk" on the BK1-1 which means that the accuracy rate for a screening outcome of "at risk" was only 37% (i.e., 18/49). In contrast, a screening outcome of "not at risk" was accurate for 95% of the cases (i.e., 69/73).

As with the CCAT, the classificational analysis between the BK1-1 and grade one GMRT was recalculated separately for boys and girls to determine whether the predictive accuracy of the BK1-1 could be improved by taking gender differences on the BKIT into account. The results of this second analysis are presented in Table 11 for boys and Table 12 for girls. Table 11 indicates that the accuracy of the BKIT for predicting reading difficulty vs. no reading difficulty on the GMRT at the end of grade one was slightly improved for boys by using BKIT cutoff scores based on the mean and standard deviation for the male sample. Considering all the indices of predictive effectiveness in combination, the optimal cutoff point for boys would appear to be BKIT2 which resulted

Table 10

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade One Gates-MacGinitie Reading Test (GMRT) Total Test Raw Scores for the Total Group (N = 122)

a)

		GMRT		
		Difficulty (< 44)	No Difficulty (> 45)	Row Total
BKIT1	At risk (≤ 74.5)	6	6	12
	Not at risk (≥ 75)	16	94	110
	Column Total	22	100	122

Overall Hit Rate: 100/122 = 82
Specificity Rate: 94/100 = 94%
False Positive Rate: 6/122 = 5%
Sensitivity Rate: 6/22 = 27%
False Negative Rate: 16/122 = 13%

b)

		GMRT		
		Difficulty (< 44)	No Difficulty (> 45)	Row Total
BKIT2	At risk (≤ 79.5)	10	15	25
	Not at risk (≥ 78)	12	85	97
	Column Total	22	100	122

Overall Hit Rate: 5/122 = 78%
Specificity Rate: 85/100 = 85%
False Positive Rate: 15/122 = 12%
Sensitivity Rate: 10/22 = 46%
False Negative Rate: 12/122 = 10%

c)

		GMRT		
		Difficulty (< 44)	No Difficulty (>45)	Row Total
BKIT3	At risk (≤ 84.5)	18	31	49
	Not at risk (≥ 85)	4	69	73
	Column Total	22	100	122

Overall Hit Rate: 87/122 = 71%
Specificity Rate: 69/100 = 69%
False Positive Rate: 31/122 = 25%
Sensitivity Rate: 18/22 = 82%
False Negative Rate: 4/122 = 3%

Note: All percentages rounded to the nearest whole number.

in an overall hit rate of 82%, a false positive rate of 8%, a false negative rate of 11%, and a sensitivity rate of 56%. To achieve this moderate sensitivity rate, only 21% of the total sample intitially had to be designated as being "at risk" on the BK1-1. In addition, a screening outcome of "not at risk" was 87% accurate while a screening outcome of "at risk" using the BKIT2 cutoff was 64% accurate - substantially greater than the 37% accuracy rate reported for the total group using BKIT3 as a cutoff score.

Table 12 indicates that using the BK1-1 to predict a grade one girl's status on the GMRT is improved by using BKIT cutoff scores based on the mean and standard deviation for the female sample. Clearly, the lowest cutoff score (i.e., BKIT1) is totally unacceptable. Even though the overall hit rate using BKIT1 is 86%, the sensitivity rate is 0%. In other words, none of the girls who experienced reading difficulty at the end of grade one on the GMRT were identified by the BK1-1. This is an excellent example of how potentially misleading an overall hit rate can be. Raising the cutoff score by half a standard deviation (i.e., BKIT2) retained the high overall hit rate of 86% and resulted in equally impressive false positive (11%) and false negative (4%) rates. In addition, 67% of the girls who actually experienced reading difficulty were identified even though only 18% of the sample had to be initially designated "at risk". This resulted in a 40% accuracy rate for a screening outcome of "at risk". Finally, 100% of the girls who later experienced reading difficulty were correctly identifed by using the ...est cutoff point (i.e., BKIT3) while still maintaining an acceptable overall hit rate of 77%. However, to achieve this remarkably high sensitivity rate, 32% of the sample had to be designated as being "at risk" on the BK1-1 which further translated into a weak 32% accuracy rate for an "at risk" screening outcome. By contrast, a screening outcome of "not at risk" using the BKIT3 cutoff point was 100% accurate. In other words, in this study, if a girl achieved a total score of 88 or better on the BK1-1 at the end of kindergarten, she had no chance of experiencing reading difficulty as measured by the GMRT at the end of grade one.

Table 11

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade One Gates-MacGinitie Reading Test (GMRT) Total Test Raw Scores for Boys (n = 66)

a)

		GMRT*		
		Difficulty (< 44)	No Difficulty (> 45)	Row Total
BKIT1	At risk (≤ 71.5)	6	3	9
	Not at risk (≥ 72)	10	47	57
	Column Total	16	50	66

Overall Hit Rate: 53/66 = 80% False Positive Rate: 3/66 = 5% False Negative Rate: 10/66 = 15%
Specificity Rate: 47/50 = 94% Sensitivity Rate 6/16 = 38%

b)

		GMRT*		
		Difficulty (< 44)	No Difficulty (> 45)	Row Total
BKIT2	At risk (≤ 76.5)	9	5	14
	Not at risk (≥ 77)	7	45	52
	Column Total	16	50	66

Overall Hit Rate: 54/66 = 82% False Positive Rate: 5/66 = 8% False Negative Rate: 7/66 = 11%
Specificity Rate: 45/50 = 90% Sensitivity Rate: 9/16 = 56%

c)

		GMRT*		
		Difficulty (< 44)	No Difficulty (>45)	Row Total
BKIT3	At risk (≤ 82.5)	10	15	25
	Not at risk (≥ 83)	6	35	41
	Column Total	16	50	66

Overall Hit Rate: 45/66 = 68% False Positive Rate: 15/66 = 23% False Negative Rate: 6/66 = 9%
Specificity Rate: 35/50 = 70% Sensitivity Rate: 10/16 = 63%

*GMRT cut-off scores are based on the mean and standard deviation for the total group.

Note: All percentages rounded to the nearest whole number.

Table 12

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade One Gates-MacGinitie Reading Test (GMRT) Total Test Raw Scores for Girls (n = 56)

a)

GMRT*

		Difficulty (< 44)	No Difficulty (> 45)	Row Total
BKIT1	At risk (≤ 78.5)	0	2	2
	Not at risk (≥ 79)	6	48	54
	Column Total	6	50	56

Overall Hit Rate: 48/56 = 86%
Specificity Rate: 48/50 = 96%
False Positive Rate: 2/56 = 4%
Sensitivity Rate: 0/6 = 0%
False Negative Rate: 6/56 = 11%

b)

GMRT*

		Difficulty (< 44)	No Difficulty (> 45)	Row Total
BKIT2	At risk (≤ 82.5)	4	6	10
	Not at risk (≥ 83)	2	44	46
	Column Total	6	50	56

Overall Hit Rate: 48/56 = 86%
Specificity Rate: 44/50 = 88%
False Positive Rate: 6/56 = 11%
Sensitivity Rate: 4/6 = 67%
False Negative Rate: 2/56 = 4%

c)

GMRT*

		Difficulty (< 44)	No Difficulty (>45)	Row Total
BKIT3	At risk (≤ 87.5)	6	13	19
	Not at risk (≥ 88)	0	37	37
	Column Total	6	50	56

Overall Hit Rate: 43/56 = 77%
Specificity Rate: 37/50 = 74%
False Positive Rate: 13/56 = 23%
Sensitivity Rate: 6/6 = 100%
False Negative Rate: 0/56 = 0%

*GMRT cut-off scores are based on the mean and standard deviation for the total group.

Note: All percentages rounded to the nearest whole number.

As will be recalled from the discussion of descriptive statistics at the beginning of this chapter, there was also a significant gender difference on the total test score of the grade one GMRT. Therefore, the results of the classificational analysis between the BK1-1 and GMRT for boys and girls were recalculated a second time using GMRT cutoff scores adjusted to reflect the differential means and standard deviations of the male and female samples. The results of this analysis are reported in Table 13 for males and Table 14 for females.

Similar to the results presented in Table 11, Table 13 indicates that the optimal cut-off point for boys is BKIT2 although using the adjusted GMRT cut-off scores resulted in both a higher overall hit rate (i.e. 85% vs. 82%) and a higher sensitivity rate (i.e., 75% vs. 56%) than when unadjusted GMRT cut-off scores were used. However, the 19% increase in the sensitivity rate did not increase the proportion of the sample that had to be initially identified as being "at risk" on the BK1-1. This proportion stayed constant at 21%. A screening outcome of "at risk" in this situation was accurate for 43% of the cases while an outcome of "not at risk" was 96% accurate.

The results in Table 14 indicate that BKIT3 was the optimal cut-off point for girls when adjusted GMRT cut-off scores were used. The use of BKIT3 resulted in acceptable false negative (5%), sensitivity (77%), and overall hit rates (79%). In order to achieve these results, 34% of the sample had to be designated as being "at risk" on the BK1-1. A screening outcome of "at risk" was accurate in 53% of the cases while a "not at risk" outcome was 92% accurate.

In summary, the BK1-1 was not highly accurate in identifying those children from the total group in kindergarten who would actually exhibit reading difficulty at the end of grade one as measured by the GMRT. The accuracy of prediction was somewhat improved when the analysis was recalculated separately by gender using adjusted cut-off scores on both the BK1-1 and GMRT. However, with the exception of a 64% accuracy rate for boys using the BKIT2 cut-off and unadjusted GMRT scores, an "at risk"

Table 13

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade One Gates-MacGinitie Reading Test (GMRT) Total Test Raw Scores for Boys (n = 66)

a)

		GMRT*		
		Difficulty (< 44)	No Difficulty (> 45)	Row Total
BKIT1	At risk (≤ 71.5)	4	5	9
	Not at risk (≥ 72)	4	53	57
	Column Total	8	58	66

Overall Hit Rate: 57/66 = 86% False Positive Rate: 5/66 = 8 False Negative Rate: 4/66 = 6%
Specificity Rate: 53/58 = 91% Sensitivity Rate: 4/8 = 50%

b)

		GMRT*		
		Difficulty (< 44)	No Difficulty (> 45)	Row Total
BKIT2	At risk (≤ 76.5)	6	8	14
	Not at risk (≥ 77)	2	50	52
	Column Total	8	58	66

Overall Hit Rate: 56/66 = 85% False Positive Rate: 8/66 = 12% False Negative Rate: 2/66 = 3%
Specificity Rate: 50/58 = 86% Sensitivity Rate: 6/8 = 75%

c)

		GMRT*		
		Difficulty (< 44)	No Difficulty (>45)	Row Total
BKIT3	At risk (≤ 82.5)	6	19	25
	Not at risk (≥ 83)	2	39	41
	Column Total	8	58	66

Overall Hit Rate: 45/66 = 68% False Positive Rate: 19/66 = 29% False Negative Rate: 2/66 = 3%
Specificity Rate: 39/58 = 67% Sensitivity Rate: 6/8 = 75%

*GMRT cut-off scores are based on the mean and standard deviation for the male sample.

Note: All percentages rounded to the nearest whole number.

Table 14

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade One Gates-MacGinitie Reading Test (GMRT) Total Test Raw Scores for Girls (n = 56)

a)

		GMRT*		
		Difficulty (< 49)	No Difficulty (> 50)	Row Total
BKIT1	At risk (≤ 78.5)	1	1	2
	Not at risk (≥ 79)	12	42	54
	Column Total	13	43	56

Overall Hit Rate: 43/56 = 77%
Specificity Rate: 42/43 = 98%
False Positive Rate: 1/56 = 2%
Sensitivity Rate: 1/13 = 8%
False Negative Rate: 12/56 = 21%

b)

		GMRT*		
		Difficulty (< 49)	No Difficulty (> 50)	Row Total
BKIT2	At risk (≤ 82.5)	6	4	10
	Not at risk (≥ 83)	7	39	46
	Column Total	13	43	56

Overall Hit Rate: 45/56 = 80%
Specificity Rate: 39/43 = 91%
False Positive Rate: 4/56 = 7%
Sensitivity Rate: 6/13 = 46%
False Negative Rate: 7/56 = 13%

c)

		GMRT*		
		Difficulty (< 49)	No Difficulty (>50)	Row Total
BKIT3	At risk (≤ 87.5)	10	9	19
	Not at risk (≥ 88)	3	34	37
	Column Total	13	43	56

Overall Hit Rate: 44/56 = 79%
Specificity Rate: 34/43 = 79%
False Positive Rate: 9/56 = 16%
Sensitivity Rate: 10/13 = 77%
False Negative Rate: 3/56 = 5%

*GMRT cut-off scores are based on the mean and standard deviation for the female sample.

Note: All percentages rounded to the nearest whole number.

screening outcome for boys, girls, and the total group resulted in more identification errors than it did in correct identifications. In contrast, and similar to the trend noted for the CCAT, a screening outcome of "not at risk" was highly accurate (i.e., ranging from 87 - 100%) for boys, girls, and the total group. Clearly, considerably more confidence can be placed in a BK1-1 screening outcome of "not at risk" than one of "at risk".

Table 15 presents the results of the classificational analysis between the BK1-1 and total test T-scores on the Gates-MacGinitie Reading Test (GMRT) at the grade two level for the total group (see p. 67 for a discussion of why T-scores were necessary). A classificational analysis was also conducted for total test raw scores on the grade two GMRT and is reported in Table D.14 in Appendix D. However, the number of subjects involved in the T-score analysis was greater than the number of subjects included in the raw score analysis and because larger numbers of subjects yield more stable results for classificational analysis (Lichenstein, 1981), the T-scores results are reported for discussion in this chapter.

Section (a) of Table 15 presents the results of the classificational analysis using the BKIT1 as a cutoff point. This cutoff point was not adequate because it failed to identify 86% (i.e., 12/14) of the students who later experienced reading difficulty in grade two as measured by their performance on the GMRT. Table 15(b) indicates that raising the cutoff point to an intermediate level (i.e., BKIT2) resulted in acceptable overall hit (82%), false positive (8%), and false negative (10%) rates. However, a sensitivity rate of only 43% was achieved using BKIT2 as a cut-off which means that 57% of the children who later experienced reading difficulty were not identified as being "at risk" on the BK1-1. Section(c) of Table 15 presents the results using the highest cut-off point (i.e., BKIT3). Although the overall hit rate dropped to 73% using this cut-off score, a sensitivity rate of 79% was achieved which means that only 21% of the

Table 15

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade Two Gates-MacGinitie Reading Test (GMRT) Total Test T-Scores for the Total Group (N = 78)

a)

		GMRT Difficulty (< 45)	GMRT No Difficulty (> 46)	Row Total
BKIT1	At risk (≤ 74.5)	2	4	6
	Not at risk (≥ 75)	12	60	72
	Column Total	14	64	78

Overall Hit Rate: 62/78 = 79% False Positive Rate: 4/78 = 5% False Negative Rate: 12/78 = 15%
Specificity Rate: 60/64 = 94% Sensitivity Rate: 2/14 = 14%

b)

		GMRT Difficulty (< 45)	GMRT No Difficulty (> 46)	Row Total
BKIT2	At risk (≤ 79.5)	6	6	12
	Not at risk (≥ 80)	8	58	66
	Column Total	14	64	78

Overall Hit Rate: 64/78 = 82% False Positive Rate: 6/78 = 8% False Negative Rate: 8/78 = 10%
Specificity Rate: 58/64 = 91% Sensitivity Rate: 6/14 = 43%

c)

		GMRT* Difficulty (< 45)	GMRT* No Difficulty (>46)	Row Total
BKIT3	At risk (≤ 84.5)	11	18	29
	Not at risk (≥ 85)	3	46	49
	Column Total	14	64	78

Overall Hit Rate: 57/78 = 73% False Positive Rate: 18/78 = 23% False Negative Rate: 3/78 = 4%
Specificity Rate: 46/64 = 72% Sensitivity Rate: 11/14 = 79%

Note: All percentages rounded to the nearest whole number.

children who actually experienced reading difficulty in grade 2 were not identified by the procedure. However, to achieve this sensitivity rate, 29 of the 78 children (37%) had to be designated as "at risk" on the BK1-1 initially. In addition, although a screening outcome of "not at risk" was highly accurate (i.e., 96%), a screening outcome of "at risk" was only accurate in 38% of the cases.

Although the results of Table 15 were recalculated separately for boys and girls, the small numbers of subjects involved in the analysis are not felt to reflect highly stable results. Therefore, these additional results are reported as information in Tables D.12 and D.13 in Appendix D. On the basis of an informal analysis, Table D.12 seems to indicate that the pattern of prediction between the BK1-1 and grade two GMRT T-scores was not improved by considering boys as a separate group. However, Table D.13 appears to show a trend towards better predictive accuracy for girls with estimated overall hit rates ranging from 80 - 90%, an estimated sensitivity rate of 83% and an estimated 66% accuracy rate for an "at risk" screening outome using the BKIT3 cut-off point.

In summary, the results of the classificational analysis between the total test score of the BK1-1 (i.e., BKIT) and T-scores from the grade two GMRT are similar to those observed between the grade one GMRT and the BK1-1. In other words, the BKIT was not highly accurate in identifying which children would ultimately experience reading difficulty in grade two as assessed by the GMRT. Also similar to trends noted for both the CCAT and grade one GMRT, although a screening designation of "not at risk" was highly accurate, a screening outcome of "at risk" actually resulted in more erroneous than correct predictions.

The results of a classificational analysis between the BK1-1 and total test T-scores from the grade three Gates-MacGinitie Reading Test (GMRT) for the total group are reported in Table 16. Classificational analyses based on raw scores were also calculated. However, as previously discussed with respect to Table 15, because the sample size available for the T-score analysis was larger than that available for the raw score analysis, only the results of the *T*-score analysis are reported for discussion in this chapter. The results of the classificational analysis using raw scores are reported in Table D.17.

Table 16(a) presents the results of the classificational analysis using BKIT1 as a cut-off point. Despite the relatively high overall hit rate of 83% and the low false positive rate of 4%, a sensitivity rate of only 8% was achieved. In other words, 92% of the children who later experienced reading difficulty in grade three, as measured by the GMRT, were not identified as being "at risk" by the BK1-1. Raising the cut-off score by one-half standard deviation (i.e., BKIT2) only improved the sensitivity rate to 23% which means that there was still a failure to identify 77% of the students who experienced reading difficulty in grade three. Section (c) of Table 15 summarizes the results using the highest cut-off point (i.e., BKIT3). Use of this cut-off score resulted in correct identification of 77% of the children who actually experienced reading difficulty at a later date. However, to achieve this sensitivity rate, 37% of the total sample had to initially be identifed as being "at risk" on the BK1-1. In addition, although a screening outcome of "not at risk" was 94% accurate using this cut-off, an "at risk" designation on the BK1-1 was only accurate in 31% of the cases.

As with the grade two GMRT (see Table 15), the results from Table 16 were recalculated separately for boys and girls but, due to the small number of subjects involved in the analysis, were not felt to represent a highly stable pattern of results. These additional results are presented in Tables D.15 and D.16 in Appendix D. An informal consideration of these tables appears to indicate that the predictive accuracy of the BK1-1 is not improved by considering boys as a separate group (i.e., Table D.15). In fact, the predictive accuracy for boys actually appears to be worse than that reported for the total group with highest sensitivity rate being 14%. In contrast, Table D.16 appears to indicate that the predictive accuracy of the BK1-1 for girls is somewhat better than that noted for the total group. The optimal cutoff point for girls, BKIT3,

Table 16

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade Three Gates-MacGinitie Reading Test (GMRT) Total Test T-Scores for the Total Group (N = 86)

a)

		GMRT Difficulty (< 47)	No Difficulty (> 48)	Row Total
BKIT1	At risk (≤ 74.5)	1	3	4
	Not at risk (≥ 75)	12	70	82
	Column Total	13	73	86

Overall Hit Rate: 71/86 = 83%
False Positive Rate: 3/86 = 4%
False Negative Rate: 12/86 = 14%
Specificity Rate: 70/73 = 96%
Sensitivity Rate: 1/13 = 8%

b)

		GMRT Difficulty (< 47)	No Difficulty (> 48)	Row Total
BKIT2	At risk (≤ 79.5)	3	11	14
	Not at risk (≥ 80)	10	62	72
	Column Total	13	73	86

Overall Hit Rate: 65/86 = 76%
False Positive Rate: 11/86 = 13%
False Negative Rate: 10/86 = 12%
Specificity Rate: 62/73 = 85%
Sensitivity Rate: 3/13 = 23%

c)

		GMRT Difficulty (< 47)	No Difficulty (>48)	Row Total
BKIT3	At risk (≤ 84.5)	10	22	32
	Not at risk (≥ 85)	3	51	54
	Column Total	13	73	86

Overall Hit Rate: 61/86 = 71%
False Positive Rate: 22/86 = 26%
False Negative Rate: 3/86 = 23%
Specificity Rate: 51/73 = 70%
Sensitivity Rate: 10/13 = 77%

Note: All percentages rounded to the nearest whole number.

resulted in both acceptable overall hit (81%) and sensitivity (83%) rates. However, similar to the trend noted for the total group, although a screening outcome of "not at risk" was highly accurate (i.e., 97%), a screening result of "at risk" using the BK1-1 was only accurate in 42% of the cases.

In summary, similar to the relationship between the BKIT and the GMRT at both the grade one and two levels, the BKIT was not highly accurate or efficient in its ability to identify which kindergarten children would actually experience reading difficulty in grade three as measured by the GMRT. Again, although considerable confidence can be placed in a screening outcome of "not at risk", very little certainty can be attached to an "at risk" screening outcome based on the total test score of the BK1-1.

Tables 17 - 19 present the results of the classificational analyses between the BK1-1 and raw scores for the total reading area, total math area, and total battery of the grade two level of the Canadian Achievement Test (CAT) for the total group. These results were not recalculated separately for boys and girls because of the extremely small sample sizes that would have been involved.

Table 17(a) reports results between the CAT total reading area and the BKIT using the lowest cut-off score (i.e., BKIT1). This cut-off point results in a failure to identify 70% of the children who experienced reading difficulty as measured by the CAT at the end of grade two. Raising the cutoff point (i.e., BKIT2) improved the sensitivity rate by 10% but at the same time resulted in a situation where the screening outcome, "at risk", was only accurate in 29% of the cases. The highest cut-off scores (i.e., BKIT3) resulted in an impressive sensitivity rate of 80%. However, to achieve this high sensitivity rate, over 50% of the sample had to be initially designated as being "at risk" - not much of an improvement over chance alone. In addition, the accuracy of being identified as "at risk" on the BK1-1 was only 28%. In other words, out of every 100 cases designated as being "at risk", only 28 would actually experience academic

Table 17

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade Two Canadian Achievement Test (CAT) Total Reading Area Raw Scores for the Total Group (N = 57)

a)

BKIT1		CAT: Difficulty (< 48)	CAT: No Difficulty (> 49)	Row Total
BKIT1	At risk (≤ 74.5)	3	5	8
	Not at risk (≥ 75)	7	42	49
	Column Total	10	47	57

Overall Hit Rate: 45/57 = 79%
Specificity Rate: 42/47 = 89%
False Positive Rate: 5/57 = 9%
Sensitivity Rate: 3/10 = 30%
False Negative Rate: 7/57 = 12%

b)

BKIT2		CAT: Difficulty (< 48)	CAT: No Difficulty (> 49)	Row Total
BKIT2	At risk (≤ 79.5)	4	10	14
	Not at risk (≥ 80)	6	37	43
	Column Total	10	47	57

Overall Hit Rate: 41/57 = 72%
Specificity Rate: 37/47 = 79%
False Positive Rate: 10/57 = 18%
Sensitivity Rate: 4/10 = 40%
False Negative Rate: 6/57 = 11%

c)

BKIT3		CAT: Difficulty (< 48)	CAT: No Difficulty (>49)	Row Total
BKIT3	At risk (≤ 84.5)	8	21	29
	Not at risk (≥ 85)	2	26	28
	Column Total	10	47	57

Overall Hit Rate: 34/57 = 60%
Specificity Rate: 26/47 = 55%
False Positive Rate: 21/57 = 37%
Sensitivity Rate: 8/10 = 80%
False Negative Rate: 2/57 = 4%

Note: All percentages rounded to the nearest whole number.

difficulty at a later date. By contrast, being labelled "not at risk" on the BK1-1 was accurate in 93% of the cases.

Table 18 presents results between the CAT total math area and the BK1-1. Section (a) of this table indicates than an overall hit rate of 81% and a sensitivity rate of 33% are achieved using the lowest cut-off score (i.e., BKIT1). Raising the cut-off by one-half standard deviation (BKIT2) lowered the overall hit rate to 70% but was not accompanied by an increase in the sensitivity rate, which remained at 33%. This sensitivity rate means that 67% of the children who later experienced difficulty in math were not identified as being "at risk" by the BK1-1. Results using BKIT3 as the cut-off score are displayed in Section (c) of Table 18. These results indicate that the sensitivity rate was only 56% despite the fact that over 50% of the sample had to be initially designated as being "at risk" on the BK1-1. Furthermore, a screening outcome of "at risk" using the BKIT3 cut-off was only accurate for 17% of the cases which represents a highly unacceptable level of predictive accuracy. Additionally, even though a screening outcome of "not at risk" was considerably more accurate at 86%, this level is less than the "not at risk" predictive accuracy values observed for other outcome variables (e.g., 93% for the CAT total reading area).

Table 19 presents the results of the classificational analysis between the BKIT and raw scores on the total battery of the CAT at the grade two level. Similar to the trend noted for the total math area of the CAT, raising the cut-off score on the BK1-1 (i.e., from BKIT1 to BKIT2) did not improve the sensitivity rate of 38%. By using the highest cut-off score (i.e., BKIT3) the sensitivity rate was improved to 63%; that is, correct identification of 63% of the children who later exhibited general academic difficulty at the end of grade two as measured by their total battery scores on the CAT. However, in order to achieve this sensitivity rate, over half of the total sample had to be identified as "at risk" on the BK1-1. In fact, if children received an "at risk" designation on the BK1-1 at the end of kindergarten using BKIT3 as the cut-off score, there was only a

Table 18

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade Two Canadian Achievement Test (CAT) Total Math Area Raw Scores for the Total Group (N = 57)

a)

		CAT		
		Difficulty (< 44)	No Difficulty (> 45)	Row Total
BKIT1	At risk (≤ 74.5)	3	5	8
	Not at risk (≥ 75)	6	43	49
	Column Total	9	48	57

Overall Hit Rate: 46/57 = 81% False Positive Rate: 5/57 = 9% False Negative Rate: 6/57 = 11%
Specificity Rate: 43/48 = 90% Sensitivity Rate: 3/9 = 33%

b)

		CAT		
		Difficulty (< 44)	No Difficulty (> 45)	Row Total
BKIT2	At risk (≤ 79.5)	3	11	14
	Not at risk (≥ 80)	6	37	43
	Column Total	9	48	57

Overall Hit Rate: 40/57 = 70% False Positive Rate: 11/57 = 19% False Negative Rate: 6/57 = 11%
Specificity Rate: 37/48 = 77% Sensitivity Rate: 3/9 = 33%

c)

		CAT		
		Difficulty (< 44)	No Difficulty (>45)	Row Total
BKIT3	At risk (≤ 84.5)	5	24	29
	Not at risk (≥ 85)	4	24	28
	Column Total	9	48	57

Overall Hit Rate: 29/57 = 51% False Positive Rate: 24/57 = 42% False Negative Rate: 4/57 = 7%
Specificity Rate: 24/48 = 50% Sensitivity Rate: 5/9 = 56%

Note: All percentages rounded to the nearest whole number.

Table 19

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade Two Canadian Achievement Test (CAT) Total Battery Raw Scores for the Total Group (N = 57)

a)

		CAT		
		Difficulty (< 134)	No Difficulty (> 135)	Row Total
BKIT1	At risk (≤ 74.5)	3	5	8
	Not at risk (≥ 75)	5	44	49
	Column Total	8	49	57

Overall Hit Rate: 47/57 = 82% False Positive Rate: 5/57 = 9% False Negative Rate: 5/57 = 9%
Specificity Rate: 44/49 = 90% Sensitivity Rate: 3/8 = 38%

b)

		CAT		
		Difficulty (< 134)	No Difficulty (> 135)	Row Total
BKIT2	At risk (≤ 79.5)	3	11	14
	Not at risk (≥ 80)	5	38	43
	Column Total	8	49	57

Overall Hit Rate: 41/57 = 72% False Positive Rate: 11/57 = 19% False Negative Rate: 5/57 = 9%
Specificity Rate: 38/49 = 78% Sensitivity Rate: 3/8 = 38%

c)

		CAT		
		Difficulty (< 134)	No Difficulty (>135)	Row Total
BKIT3	At risk (≤ 84.5)	5	24	29
	Not at risk (≥ 85)	3	25	28
	Column Total	8	49	57

Overall Hit Rate: 30/57 = 53% False Positive Rate: 24/57 = 42% False Negative Rate: 3/57 = 5%
Specificity Rate: 25/49 = 51% Sensitivity Rate: 5/8 = 63%

Note: All percentages rounded to the nearest whole number.

17% chance that they would acutally experience academic difficulty at the end of grade two. However, a screening outcome of "not at risk" was accurate for 89% of the cases.

In summary, an analysis of the prediction-performance matrices depicting the relationship between the outcomes of the BK1-1 and raw scores on the CAT administered at the end of grade two generally indicate a low level of predictive accuracy and efficiency. Similar to its relationship with both the CCAT and GMRT, the BKIT was not highly effective in identifying which kindergarten children would actually experience reading, mathematical, or general academic difficulty at the end of grade two. Also, consistent with the trends observed for both the CCAT and GMRT, the accuracy of an "at risk" designation on the BK1-1 was very weak although considerable certainty could be placed in a "not at risk" screening outcome.

Classificational analyses were also calculated to investigate the relationship between the BK1-1 and total reading area, total math area, and total battery raw score on the CAT administered at the end of grade three. However, the sample size for this analysis was judged to be too small to obtain a highly stable and reliable pattern of results. Therefore, these results are reported as information in Tables D.18 - D.20 in Appendix D. An informal inspection of these tables appears to indicate results that are quite consistent with the trends noted for the grade two level of the CAT.

2. <u>Does the BK1-1 have differential predictive validity for boys and girls?</u>

All correlations between the BK1-1 total test score (i.e., BKIT) and the criterion variables (Tables 2- 4) for boys and girls were tested for significant differences. Table 20 presents comparisons between correlations for boys and girls which were significantly different. An inspection of these comparisons reveals that the BKIT was equally predictive for boys and girls for all variables at all grade levels with the exception of the Gates-MacGinitie Reading Test (GMRT) at the grade three level. These

Table 20

Significant Gender Differences Between Correlations

Criterion Variable	Correlation for Boys	Correlation for Girls	Z value	p level
GRC3	.0004	.549	2.38	<.05
GRT3	.033	.557	2.11	<.05
GTV3	.045	.555	2.59	<.01
GTC3	.067	.577	2.64	<.01
GTT3	.072	.622	2.93	<.01

results indicate that the BKIT was a significantly better predictor of grade three reading achievement as measured by the GMRT for girls than boys.

3. <u>What is the best combination of subtests from the BK1-1 to predict indices of academic ability and achievement at the ends of grades one to three?</u>

Tables 21 - 30 present the results of the stepwise multiple regression analysis between the thirteen subtests of the BK1-1 and selected academic outcome variables. As previously stated in Chapter 3, only criterion variables with sample sizes of seventy or more subjects were selected for inclusion in this analysis.

Table 21 shows the stepwise regression procedure for standard age scores on the Canadian Cognitive Abilities Test (CCAT) administered at the end of grade one for the total group. An inspection of this table reveals that only three of the thirteen BK1-1 subtests contributed significantly to the regression equation. The first variable to enter the equation was NC (Number Comprehension) accounting for 21.5% of the variance. In combination, the remaining two variables, RLL (Recognition of Lowercase Letters) and AD (Auditory Discrimination), added an additional 8% to the total variance accounted for, raising it to 29.5%. Of note is the fact that the final multiple correlation of .556 between the CCAT and a combination of three BK1-1 subtests is higher than the simple correlation of .523 between the total test score of the BK1-1 (i.e., BKIT) and the CCAT reported in Table 2.

Tables 22 and 23 present the results of the stepwise regression procedure between the CCAT and BK1-1 subtests carried out separately for boys and girls. In both cases, only two variables contributed significantly to the equation. However, different variables were involved in each situation. For boys (Table 22), RLL was the first variable to enter the equation and accounted for 28.89% of the total variance. RC (Rote Counting) raised the total amount of explained variance by 4.3% to 33.19%. Similar to the results noted for the total group, the multiple correlation of .591 based on two

Table 21

Stepwise Multiple Regression Analysis: Thirteen BK1-1 Subtests with Grade One Canadian Cognitive Abilities Test (CCAT) Standard Age Scores for the Total Group (N = 149)

Step Number	Variable	F Value	p level	Multiple R	%age of Variance Accounted for*	%age of Variance Added
1	NC	41.608	<.001	.470	21.50%	--
2	RLL	27.938	<.001	.526	26.69%	5.19%
3	AD	21.650	<.001	.556	29.50%	2.81%

*adjusted for shrinkage

Table 22

Stepwise Multiple Regression Analysis: Thirteen BK1-1 Subtests with Grade One Canadian Cognitive Abilities Test (CCAT) Standard Age Scores for Boys (n = 80)

Step Number	Variable	F Value	p level	Multiple R	%age of Variance Accounted for*	%age of Variance Added
1	RLL	33.099	<.001	.546	28.89%	--
2	RC	20.618	<.001	.591	33.19%	4.3%

*adjusted for shrinkage

Table 23

Stepwise Multiple Regression Analysis: Thirteen BK1-1 Subtests with Grade One Canadian Cognitive Abilities Test (CCAT) Standard Age Scores for Girls (n = 69)

Step Number	Variable	F Value	p level	Multiple R	%age of Variance Accounted for*	%age of Variance Added
1	NC	17.202	<.001	.452	20.43%	--
2	AD	11.214	<.001	.504	23.10%	2.67%

*adjusted for shrinkage

subtests was higher than the correlation of .578 between the CCAT and BKIT. Table 23 indicates that NC was most highly predictive of performance on the CCAT for girls, accounting for 20.43% of the variance. AD entered the equation on the second step adding 2.67% and raising the explained variance to 23.10%. Consistent with the trend noted for both boys and the total group, the final multiple correlation of .504, based on two subtests, was higher than the correlation of .439 between the CCAT and BKIT (see Table 2).

In summary, for boys, girls, and the total group, a small number of BK1-1 subtests was more effective than the total test score of the BK1-1 (i.e., BKIT) for predicting grade one CCAT scores. However, in all three cases these subtest combinations only accounted for between 23 and 33% of the total variance. In other words, 67 to 77% of the variance in grade one CCAT performance was not accounted for by performance on the subtests of the BK1-1.

Tables 24 to 26 present the results of the stepwise multiple regression analysis of the thirteen BK1-1 subtests with vocabulary, comprehension, and total test raw scores for the total group on the grade one Gates-MacGinitie Reading Test (GMRT). In all three cases, RLL was the first variable to enter the equation accounting for 25.63%, 35.90%, and 35.72% of the total variance for the vocabulary (Table 24), comprehension (Table 25), and total test (Table 26) raw scores respectively. Table 24 indicates that NS (Number in Sequence) was the second variable to enter the regression equation for the GMRT Vocabulary subtest. NS contributed 6.26% of the variance and increased the total amount of variance accounted for to 31.89%. Together, RLL and NS accounted for more variance in grade one reading vocabulary performance (i.e., multiple $\underline{R}$ = .575) than the BKIT (i.e., Table 3, $\underline{r}$ = .515).

In terms of the GMRT Comprehension subtest (Table 25), PPD (Prints Personal Data) followed RLL as the second variable in the regression equation. PPD added 2.0% to the variance which raised the total explained variance to 37.90%. Whereas the simple

Table 24

Stepwise Multiple Regression Analysis: Thirteen BK1-1 Subtests with Grade One Gates- MacGinitie Reading Test (GMRT) Vocabulary Subtest Raw Scores for the Total Group (N = 122)

Step Number	Variable	F Value	p level	Multiple R	%age of Variance Accounted for*	%age ofVariance Added
1	RLL	42.698	<.001	.512	25.63%	- -
2	NS	29.322	<.001	.575	31.89%	6.26%

*adjusted for shrinkage

Table 25

Stepwise Multiple Regression Analysis: Thirteen BK1-1 Subtests with Grade One Gates- MacGinitie Reading Test (GMRT) Comprehension Subtest Raw Scores for the Total Group(N = 122)

Step Number	Variable	F Value	p level	Multiple R	%age of Variance Accounted for*	%age ofVariance Added
1	RLL	68.761	<.001	.604	35.90%	- -
2	PPD	37.925	<.001	.624	37.90%	2.0%

*adjusted for shrinkage

Table 26

Stepwise Multiple Regression Analysis: Thirteen BK1-1 Subtests with Grade One Gates- MacGinitie Reading Test (GMRT) Total Test Raw Scores for the Total Group(N = 122)

Step Number	Variable	F Value	p level	Multiple R	%age of Variance Accounted for*	%age ofVariance Added
1	RLL	68.249	<.001	.602	35.72%	- -
2	NS	40.666	<.001	.637	39.60%	3.88%

*adjusted for shrinkage

correlation between the BKIT and the GMRT Comprehension subtest was .610 (Table 3), the RLL and PPD subtests in combination resulted in a multiple correlation of .624.

Similar to the regression equation for the GMRT Vocabulary subtest, Table 26 indicates that NS was also the second variable to enter the regression equation for the GMRT total test score. NS accounted for 3.88% of the variance and raised the total amount of variance accounted for to 39.60%. Consistent with the trend noted for both the Vocabulary and Comprehension subtests of the GMRT, two BK1-1 subtests in combination resulted in a stronger relationship (i.e., multiple R = .367) with the total test score of the GMRT than the BKIT (Table 3, r = .608).

In summary, a small number of BK1-1 subtests in combination were more effective than the total test score of the BK1-1 (i.e., BKIT) for predicting grade one reading performance as measured by the GMRT. The ability to recognize lower case letters emerged as the single most effective predictor for all three criterion measures. The total amount of variance accounted for by the various BK1-1 subtest combinations ranged from 31.89% for the GMRT Vocabulary subtest to 39.60% for the total test score of the GMRT. Although this represents moderate predictive power, it must be emphasized that between 60 and 68% of the variance in grade one reading performance is still not accounted for by the BK1-1 subtests.

Table 27 shows the results of the stepwise multiple regression analysis between the thirteen subtests of the BK1-1 and total test T-scores on the grade two GMRT for the total group. Similar to the trend noted for the grade one GMRT total test score, the first variable to enter the equation was RLL, accounting for 15.14% of the total variance. The second and final variable to add significantly to the regression equation was VD (Visual Discrimination) which contributed 6.28% to the total variance raising it to 21.42%. This combination of two subtests resulted in a stronger relationship (multiple R = .484) with the grade two GMRT than the total test score of the BK1-1 (Table 3, r = .450). However, although there was a moderate relationship between the best

Table 27

Stepwise Multiple Regression Analysis: Thirteen BK1-1 Subtests with Grade Two Gates- MacGinitie Reading Test T-Scores for the Total Group(N = 78)

Step Number	Variable	F Value	p level	Multiple R	%age of Variance Accounted for*	%age of Variance Added
1	RLL	14.732	<.001	.403	15.14%	--
2	VD	11.492	<.001	.484	21.42%	6.28%

*adjusted for shrinkage

combination of BK1-1 subtests and the grade two GMRT, the strength of this relationship was noticeably less than that observed at the grade one level (i.e., .484 vs. .637).

The results of the stepwise multiple regression analysis for the vocabulary, comprehension, and total test *T*-scores of the grade three GMRT for the total group are reported in Tables 28 to 30. Consistent with the trend observed for the grade one GMRT, the RLL subtest was the first variable to enter the equation in all three situations. However, the total amount of variance accounted for by RLL was considerably less at the grade three level than at the grade one level.

Table 28 indicates that RLL accounted for 10.15% of the variance for the GMRT Vocabulary subtest. Following RLL, the second variable to enter the equation was CR (Color Recognition) which raised the total variance to 17.04% by adding 6.89%. The third and final variable to contribute significantly to the equation was NC which increased the total amount of variance accounted for to 21.70%. The final multiple correlation of .495 was noticeably stronger than the correlation of .322 between the sum of all thirteen BK1-1 subtests (i.e., BKIT) and the GMRT Vocabulary subtest (Table 3).

An inspection of Table 29 indicates that RLL accounted for 15.26% of the total variance in the regression equation for the GMRT Comprehension subtest. This was followed by RA which raised the total amount of explained variance to 18.75%. Whereas the simple correlation between BKIT and the GMRT Comprehension subtest was .347 (Table 3), the RLL and RA subtests resulted in a multiple correlation of .455.

Table 30 indicates that both the combination of variables and their order of entry into the regression equation for the GMRT total test score was identical to that observed for the GMRT Vocabulary subtest. RLL entered the equation first and accounted for 15.25% of the variance. This was followed by CR and NC which accounted for 5.04% and 5.36% of the variance respectively and raised the total amount of explained variance to 25.65%. Similar to the trends noted for both the Vocabulary and Comprehension

Table 28

Stepwise Multiple Regression Analysis: Thirteen BK1-1 Subtests with Grade Three Gates- MacGinitie Reading Test · Vocabulary Subtest T-Scores for the Total Group(N = 86)

Step Number	Variable	F Value	p level	Multiple R	%age of Variance Accounted for*	%age ofVariance Added
1	RLL	10.599	≤.01	.335	10.15%	- -
2	CR	9.727	<.001	.436	17.04%	6.89%
3	NC	8.853	<.001	.495	21.70%	4.66%

*adjusted for shrinkage

Table 29

Stepwise Multiple Regression Analysis: Thirteen BK1-1 Subtests with Grade Three Gates- MacGinitie Reading Test Comprehension Subtest T-Scores for the Total Group(N = 86)

Step Number	Variable	F Value	p level	Multiple R	%age of Variance Accounted for*	%age ofVariance Added
1	RLL	16.301	<.001	.403	15.26%	- -
2	RA	10.809	<.001	.455	18.75%	3.49%

*adjusted for shrinkage

Table 30

Stepwise Multiple Regression Analysis: Thirteen BK1-1 Subtests with Grade Three Gates- MacGinitie Reading Test Test T-Scores for the Total Group(N = 86)

Step Number	Variable	F Value	p level	Multiple R	%age of Variance Accounted for*	%age ofVariance Added
1	RLL	16.290	<.001	.403	15.25%	- -
2	CR	11.817	<.001	.471	20.29%	5.04%
3	NC	10.775	<.001	.532	25.65%	5.36%

*adjusted for shrinkage

subtests, this combination of three BK1-1 subtests resulted in a stronger relationship (i.e., multiple $\underline{R}$ = .532) with the grade three GMRT total test score than the BKIT (Table 3, $\underline{r}$= .373).

In summary, consistent with the results of stepwise multiple regression analyses at both the grade one and two levels of the GMRT, a very small number of subtests accounted for more variance in grade three reading performance as measured by the GMRT than the total test score of the BK1-1. Also consistent with previously noted trends, the RLL subtest emerged as the single best predictor in each of the three equations. However, the total amount of variance in grade three GMRT performance which was accounted for by the various combinations of BK1-1 subtests was actually quite small, ranging from 18.75% for the Comprehension subtest to 25.65% for the total test score. This leaves 76 - 81% of the variance in grade three reading performance unaccounted for by the BK1-1. In addition, similar to the trend noted for the GMRT at the grade two level, the relationship between the most highly predictive combination of BK1-1 subtests and the GMRT at the grade three level is not as strong as that observed at the grade one level.

<u>Research Objective</u>

At the outset of this study it was proposed that local norms would be developed for the school in which the research study was conducted if, on the basis of the results of the correlational and classificational analysis, the BK1-1 was found to possess an acceptable level of predictive validity for screening purposes. Although the BK1-1 was found to correlate moderately with academic ability and achievement at the end of grade one, the results of classificational analysis indicate that the BK1-1 is not highly accurate or efficient in its ability to make screening decisions. On the basis of this information it was determined that the development of norms and cut-off points for this instrument was not appropriate.

However, as outlined in Chapter 3, four questions were supplementary to the proposed norming process and the results pertaining to these questions will be presented.

a) Do significant gender differences exist for the total test score of the BK1-1?

An inspection of Table 1(p. 64) indicates that there was a significant difference between boys (M = 81.72) and girls (M = 87.80) with respect to the total test score of the BK1-1 (t = -3.67, p < .001). The results indicate that girls scored significantly higher on the BK1-1 than boys.

b) Is BK1-1 test performance influenced by a child's chronological age at the time of test administration?

The results of a one-way factorial ANOVA using chronological age (CA) intervals of 4 months, are reported in Table 31. These results indicate that CA had a significant effect on overall test performance. An inspection of the univariate comparisons of the five CA groups indicates that there was one significant comparison; that is, between the oldest age group (i.e., ≥ 79 months, M = 74.25) and the third oldest age group (i.e., 71 - 74 months, M = 87.77). However, as noted in the section on descriptive statistics where children repeated kindergarten, their second set of BK1-1 scores was used in the analysis. Clearly, the children who repeated kindergarten would have been in the oldest age group and it is likely that the children who repeated kindergarten represent a below-average group in terms of skill development.

c) Does the time of BK1-1 test administration (i.e., end of kindergarten vs. beginning of grade one) affect BK1-1 test performance?

A t-test for independent groups was used to compare the mean score of children who were administered the BK1-1 at the end of kindergarten and those who were

Table 31

One-Way ANOVA: The Effect of Chronological Age (4 month intervals) on BK1-1 Total Test Scores (N = 149)

Degrees of Source	Sum of Freedom	Mean Squares	Squares	F Ratio	F Probability
Between Groups	4	1557.656	389.414	3.800	.0057
Within Groups	144	14755.676	102.470		
Total	148	16313.332			

administered the BK1-1 at the beginning of grade one. The results of this test indicated that there was no significant difference between these two groups. Therefore, it can be reasonably concluded that children do not make significant gains over July and August in the types of skills assessed by the BK1-1.

CHAPTER V

DISCUSSION OF RESULTS

This chapter will highlight the major findings from the results presented in the previous chapter. The implications of these findings and, where appropriate, their relationship to pertinent literature in the area of school readiness testing will be discussed.

Gender Differences in School Readiness and Early Academic Achievement

A comparison of the mean scores achieved by girls and boys in this study revealed that girls scored significantly higher than boys both in terms of grade one readiness and grade one reading achievement. These findings occurred for reasons other than differential intellectual ability level and are highly consistent with literature in the area of school readiness and early school achievement. For example, May and Welch (1986) found that girls scored significantly higher than boys on the Gesell School Readiness Test while Feshbach et al. (1977) obtained results which indicated that girls had significantly higher scores than boys with respect to grade one reading achievement. Similarly, Burns and Roe (1980), in their summary of relevant research, concluded that girls generally tend to score higher "than boys on both readiness measures and in reading achievement at the end of grade one" (p. 35).

Explanations for gender differences in school readiness and early school achievement have been explored by several researchers. It is generally accepted that girls are more mature in physical and linguistic development than boys (Burns & Roe, 1980; DiPasquale et al., 1980; Gredler, 1978; Ilg & Ames, 1965). On the basis of this explanation, girls are simply more developmentally ready than boys to deal successfully with early school-related activities that emphasize fine motor control (e.g., pencil and paper tasks) and language skills (e.g., sound-letter relationships). In support of this developmental hypothesis in the present study is the fact that significant gender differences in academic achievement were virtually nonexistent past grade one. This suggests that as the boys matured physically and linguistically, their mean level of academic achievement rose to match that of the girls; that is, the significance of poor performance by boys on the tasks of the BK1-1 decreased in importance with time. Therefore, it would seem ill-advised to make any type of long term prediction for boys on the basis of their low performance on the BK1-1.

Apart from this developmental explanation, it has also been suggested that gender differences in school readiness and early school achievement may be more a result of cultural influences and sex-role expectations than biological maturation (Burns & Roe, 1980; Nadon, 1983). From this perspective it is suggested that girls in our culture are encouraged from a much earlier age than boys to practice the type of skills (e.g., printing, coloring, letter recognition) and demonstrate the type of behaviors (e.g., sitting quietly for extended periods of time) that are associated with early school success. On the basis of this explanation, girls simply have more practice than boys with the types of skills and behaviors that facilitate a smooth adjustment to the demands of a traditional school setting. As previously stated, it was found that significant gender differences in academic achievement in the present study disappeared after grade one and it was suggested that this was related to biological development. However, it is also possible that this finding is reflecting the fact that the academic performance of the boys improved as they practiced and became more proficient with the types of skills and behaviors necessary for successful school adjustment.

However, regardless of the existence of gender differences or the reasons why girls outperform boys on measures of school readiness and early school achievement, it must be emphasized that the variation within each gender group with respect to these two factors is far greater than the average differences between girls and boys (Burns & Roe, 1980, p. 36). In other words, some boys will be more ready for grade one than some girls and vice versa. Therefore, the primary message for educators is that each child should be responded to as an individual rather than on the basis of gender expectations (Burns & Roe, 1980, p. 36).

The Predictive Effectiveness of the BK1-1

The correlational and classificational analyses used to examine the predictive effectiveness of the BK1-1 revealed that the total test score of the BK1-1 was moderately predictive (i.e., r = .439 - .578) of intellectual ability as measured by the Canadian Cognitive Abilities Test (CCAT) administered at the end of grade one. It will be recalled from Chapter 2 (see p. 40) that the BK1-1 test manual states that "students of low ability can be expected to produce low scores" (Brigance, 1982, p. ix). However, on the basis of these results this claim is misleading because although the correlation between the BK1-1 and CCAT for the total group was .523, this relationship only accounts for 25% of the variance in grade one intellectual ability scores. Furthermore, a classificational analysis of the data indicated a high level of predictive inaccuracy, especially for children who achieved relatively lower scores than their peers on the BK1-1. On the basis of the classificational analysis it was concluded that children who were designated as being "not at risk" on the BK1-1 generally scored in the average to above-average range of intellectual ability. However, the opposite relationship was not true; that is, if children were identified as being "at risk" on the BK1-1, they did not necessarily obtain below-average scores on the CCAT. This type of situation is quite common in early identification literature (e.g., Badian, 1982; Mercer et al., 1979b) and is succinctly summarized by Nadon (1983) who stated that ". . . success may indicate an ability but failure does not necessarily indicate lack of ability" (p. 88).

Children could obtain low scores on the BK1-1 for a variety of reasons. For example, as shown in the previous section on descriptive statistics, boys obtained significantly lower BK1-1 scores than girls. Their lower scores were not a result of lower intellectual ability but were likely due to factors related to maturation and a lack of task familiarity. Therefore, the test manual's claim of a relationship between BK1-1 test scores and intellectual ability is not supported.

The correlational analysis of the data in this study also indicated that the total test score of the BK1-1 was moderately predictive (i.e., r = .408 - .608) of academic achievement in the early grades. This relationship was strongest at the grade one level and decreased in strength in grades two and three. Certainly, it is recognized that the BK1-1 test manual makes no claim to predict academic achievement in grades two and three. However, it was judged that tracing longitudinal trends in predictive effectiveness would provide useful information from a theoretical perspective. Table 32 provides an overview of the predictive validity correlation coefficients from the research literature on other school readiness tests as compared to those obtained for the BK1-1. An examination of this table indicates that, with the exception of the Metropolitan Readiness Test (MRT), the predictive validity coefficients for the BK1-1 are comparable to those obtained for other school readiness tests.

As previously stated, the strength of the predictive validity coefficients for the BK1-1 decreased in grades two and three. Only two of the studies reviewed reported results past grade one which made it difficult to compare this finding with those from similar studies. However, this downward trend in correlational strength is intuitively logical when the degree of similarity between the content of the BK1-1 and the

Table 32

Comparative Predictive Validity Correlation Coefficients for Selected School Readiness Tests

Study	N	Criterion Variable (CV)	Time of CV Administration	Correlation
BK1-1 (Total Test Score)				
Shillington (1989)	122	Gates-MacGinitie Reading Test (GMRT)	end of gr. 1	.608
"	78	GMRT	gr. 2	.450
"	86	GMRT	gr. 3	.373
"	57	Canadian Achievement Test (Total Reading)	end of gr. 2	.408
"	57	Canadian Achievement Test (Total Battery)	"	.387
"	32	Canadian Achievement Test (Total Reading)	end of gr. 3	.496
"	32	Canadian Achiovment Test (Total Battery)	"	.482
Metropolitan Readiness Test (Total Test Score)				
Telegdy (1975)	56	Wide Range Achievement Test (WRAT) (Reading)	end of gr. 1	.700
"	56	Gray Oral Reading Test (GORT)(oral reading)	"	.580
"	56	GORT (literal comprehension)	"	.670
Randel, Fry & Ralls (1977)	62	Stanford Achievement Test (SAT) (Total Reading)	"	.340
"	65	"	end of gr. 3	.510
Flynn & Flynn (1978)	81	California Achievement Test (CAT)(Total Reading)	end of gr. 2	.340
"	81	CAT (Total Battery)	"	.280

Table 32 (continued)

Study	N	Criterion Variable (CV)	Time of CV Administration	Correlation
Metropolitan Readiness Test (Pre-reading Skills Composite)				
Nagle (1979)	176	SAT (Total Reading)	end of gr. 1	.670
Swanson, Payne & Jackson (1981)	72	Metropolitan Achievement Test (MAT)(Total Reading)	"	.819
	64	"	"	.835
Gullo, Clements & Robertson (1984)	88	Scott-Foresman Achievement Test (SFAT) (Total Reading)	"	.790
"	88	SFAT (Total Battery)	"	.750
First Grade Screening Test (Total Test Score)				
Telegdy (1975)	56	WRAT (Reading)	"	.500
"	56	GORT (oral reading)	"	.400
"	56	GORT (literal comprehension)	"	.530
Piersel & Kinsey (1984)	67	Science Research Associates Achievement Series (SRAAS) (Reading)	mid gr. 1	.440
"	67	SRAAS (Total Battery)	"	.490
Screening Test of Academic Readiness (Total Score)				
Telegdy (1975)	56	WRAT (Reading)	end of gr. 1	.670
"	56	GORT (oral reading)	"	.480
"	56	GORT (literal comprehension)	"	.620
Nichta, Federici & Schueroger (1982)[1]	28	Peabody Individual Achievement Test (PIAT) Total Test grade equivalent score)	mid gr. 1	.600

Table 32 (continued)

Study	N	Criterion Variable (CV)	Time of CV Administration	Correlation
Develomental Indicators for the Assessment of Learning (Concepts Subtest Score)				
Vacc, Vacc & Fogelman (1987)[1]	245	CAT (Reading)	end of gr. 1	.350
"	245	CAT (Total Battery)	"	.380
Denver Developmental Screening Test (Total Score)				
Lindquist (1982)	351	GMRT (Total Test Score)	*	.460
ABC Inventory				
Randel, Fry & Ralls (1977)	62	SAT (Total Reading)	*	.190
"	65	"	end of gr. 3	.230
Meeting Street School Screening Test (Total Score)				
Swanson, Payne & Jackson	72	MAT (Total Reading)	end of gr. 1	.773
"	69	"	"	.726

*administered either at the end of kindergarten or beginning of grade one unless otherwise noted.

[1]administered at the beginning of kindergarten

curricular content at each grade level is considered. That is, the predictive validity coefficients were likely strongest at the grade one level because the skills that are emphasized in grade one are more similar to the tasks on the BK1-1 than the type of skills emphasized at later grade levels. This line of reasoning is supported by Feshbach al. (1977) and Lindsay and Wedell (1982) who have asserted that the types of abilities and processes influencing reading achievement in the early grades are much different from those in later grades.

Finally, a classificational analysis of the data indicated that even though the BK1-1 was moderately correlated with indices of academic ability and achievement, it had an extremely low level of predictive accuracy for making screening decisions about individual children. This finding is an excellent example of why correlations are not useful for making generalizations about individual cases (Eaves et al., 1974; Keogh & Becker, 1973). Furthermore, it serves to show how a test that possesses moderate predictive validity from a correlational perspective may not necessarily be educationally valid or relevant. This adds support to Badian's (1982) assertion that: "a significant correlation between preschool test scores and later school achievement does not mean that the preschool instrument is valid for identifying children at risk" (p. 309). Therefore, consistent with the recommendations of several writers (Lessler & Bridges, 1973; Lichenstein, 1981; Lindsay & Wedell, 1982; Mercer et al., 1979a, 1979b; Satz & Fletcher, 1979; Wilson & Reichmuth, 1985), this finding reemphasizes the importance of evaluating screening devices from the perspective of classificational analysis. Unfortunately, as previously discussed in Chapter 2 (see pp. 35-36), the predictive validity of school readiness tests is rarely considered in this manner.

The results of the classificational analysis in this study also demonstrated that the results of such an analysis are never absolute and that cut-off points can be manipulated to obtain different outcomes (Lichenstein & Ireton, 1984; Mercer et al., 1979a; Wilson & Reichmuth, 1985). For example, using a low cut-off point on the BK1-1 in this study produced impressive false positive, false negative, and overall hit rates but failed to identify the majority of children who later experienced academic difficulty. This finding indicates that a high overall hit rate does not necessarily mean that a test is adequate for making screening decisions - a concern that has been expressed by several writers (Lichenstein & Ireton, 1984; Lindsay & Wedell, 1982; Wilson & Reichmuth, 1985). By raising the cut-off score on the BK1-1, the majority of the children who later experienced academic difficulty were correctly identified. However, raising the cut-off score to this level usually meant that nearly half of the children in the sample had to be initially designated as being "at risk" on the BK1-1, which resulted in an exceptionally high false positive rate. This paradoxical situation supports Adelman's (1978) observation that there are very few screening procedures that can "identify a large number of problems without making many false positive errors" (p. 156).

A final observation from the classificational analysis of the data in this study is that it demonstrated the importance of considering the results of a prediction-performance matrix horizontally (Frankenburg, 1985; Mercer et al., 1979b; Wilson & Reichmuth, 1985). Similar to the situation previously noted for the CCAT, a horizontal consideration of the results of the classificational analyses indicated that very little confidence could be placed in an "at risk" designation on the BK1-1 although considerable trust could be attached to a "not at risk" classification. This relationship is highly consistent with the findings of other researchers such as Barnes (1985), who concluded that the predictive value of an "at risk" outcome on the Jansky Screening Index was extremely low. Keogh and Smith (1970) commented on this general trend by stating that, "more predictive information is gained from the fact that a child does perform well on a school-related task than from the fact that he does not perform well" (p. 289). Similarly, Rubin et al. (1978) have concluded that, "far greater reliance can be placed on the use of high preschool readiness scores as predictors of essentially normal or better academic performance than on use of low preschool readiness scores as predictors of poor performance" (p. 63).

Gender Differences in the Predictive Validity of the BK1-1

A comparison of the correlations obtained for boys and girls in this study revealed that although the BK1-1 was moderately predictive of academic achievement for boys at the grade one level, none of the correlations for boys at the grade two and three levels exceeded .40 and only one was statistically significant. In contrast, the BK1-1 was a moderate, significant predictor of academic achievement for girls at all three grade levels. Additionally, the BK1-1 was found to be a significantly better predictor of reading achievement for girls than boys at the end of grade three.

As discussed in Chapter 2 (see p. 35), only one of the studies reviewed provided separate correlations for girls and boys so that the issue of differential predictive validity could be addressed. Furthermore, although some studies (e.g., Bolig & Fletcher, 1973) have reported that certain readiness tests are more predictive for girls than boys, they have not statistically tested the significance of the differences between the correlations obtained by the two groups. Therefore, it is difficult to compare the present findings to those from previous investigations. However, the results of this study are consistent with both the developmental and task familiarity hypotheses forwarded to explain why girls scored significantly higher than boys in terms of both grade one readiness and grade one reading achievement.

Using these hypotheses it can be argued that either because of advanced biological maturation or greater task familiarity, girls are more stable and firmly established relative to one another at the end of kindergarten than boys with respect to the types of skills assessed by the BK1-1. Therefore, their relative rank order tends to persist throughout the early grades which means that the level of correlational strength observed at grade one would tend to remain consistent over time. Conversely, because boys mature more slowly and /or gain familiarity with school-related tasks and behaviors later than girls, they do not establish themselves into a relatively stable rank ordering by the end of kindergarten. This means that although end-of-kindergarten standing for boys is moderately predictive on a short term basis (i.e., end of grade one), it is not valid for long term predictions which explains the drop in correlational strength observed for the male sample in grades two and three. On the basis of this explanation it can be further hypothesized that a low score on the BK1-1 at the end of kindergarten for a girl is more indicative of possible academic problems in the future than for a boy.

The Best Predictive Combination of BK1-1 Subtests

A stepwise multiple regression procedure was used to determine the best combination of BK1-1 subtests for predicting indices of academic ability and achievement in the early grades. The results of this procedure are summarized in Table 33 and indicate that the RLL (Recognition of Lowercase Letters) subtest emerged as the best single predictor of reading achievement at all three grade levels. This finding is highly consistent with the research literature where it has been stated that the ability to identify letters is a "hardy perennial" that consistently emerges as a good predictor of reading ability (Satz et al., 1978). In support of this claim, Badian (1982) found that a child's ability to name thirteen uppercase letters on the Holbrook Screening Battery

Table 33

Summary of Stepwise Multiple Regression Analysis of the Best Combination of BK1-1 Subtests for Predicting Each Criterion Measure

Criteria*	Best Combination of BK1-1 Subtests**	% of Variance Accounted for by Best Combination
CCAT	NC, RLL, AD	29.50%
CCAT (boys)	RLL, RC	33.10%
CCAT (girls)	NC, AD	23.10%
GRV1	RLL, NS	31.89%
GRC1	RLL, PPD	37.90%
GRT1	RLL, NS	39.60%
GTT2	RLL, VD	21.42%
GTV3	RLL, CR, NC	21.70%
GTC3	RLL, RA	18.75%
GTT3	RLL, CR, NC	25.65%

Note: *Based on total group unless otherwise noted.

**All subtests in combination contributed significantly (i.e. $p \leq .05$) and are listed in order of their contribution to the model.

six months prior to kindergarten was the best single predictor (i.e., r = .608) of reading achievement at the end of grade one. In addition, Telegdy (1975) found that the Letters subtest of the STAR was the best single predictor (i.e., r = .760) of grade one reading achievement while Klein (1977) found the same subtest to be the second strongest predictor of reading achievement (i.e., r = .505) at the grade two level. Also, Freebody and Rust (1985) and Horn and O'Donnell (1984) have stated that letter knowledge is a major predictor of grade one reading achievement. Similarly, Barnes (1985), through his work with the Jansky Screening Index (JSI), concluded that the "only consistent predictive screening measure for later reading disability was the letter naming subtest [on the JSI]" (p. 190). Mann (1984) has supported this conclusion with her observation from her study that "future poor readers were slower at naming the letters and made more errors than the future good readers did" (p. 125).

Clearly, early familiarity and proficiency with the code of the English language facilitates the development of reading skills. However, it is also evident from Table 33 that the percentage of variance accounted for by the combination of RLL and one or two additional subtests is noticeably greater in grade one than at the second and third grade levels. This trend is consistent with the hypothesis forwarded to explain why the strength of correlations between the total test score of the BK1-1 and indices of academic achievement decreased in grades two and three (see p. 116). That is, although letter recognition skills are likely emphasized and important for success in the grade one reading program, they are not highly relevant to reading achievement in later grades. Furthermore, even the best combination of subtests only accounted for 40% of the variance in grade one reading achievement. Therefore, it must be emphasized that many skills and processes not tapped by the BK1-1 have a highly significant influence on early reading achievement.

Table 33 also indicates that the majority of the subtests included in the most highly predictive subtest combinations were language-related; that is, RLL, RA, and CR are oral expressive language tasks, NS and PPD assess written expressive language skills, and NC involves receptive language processes. This finding coincides with Horn and Packard's (1985) assertion that language variables consistently show a strong relationship to reading achievement. Also evident from Table 33 and the results presented in Chapter 4 (see pp. 101 - 109) is that a very small number of BK1-1 subtests, usually two or three, accounted for more variance in academic abiltiy and achievement than the total test score of the BK1-1. This likely results from a high degree of overlap between the various subtests on the BK1-1. Intercorrelations between the thirteen subtests of the BK1-1 are presented in Table D.21 and do indicate that many of the subtests are related to one another. Therefore, despite having different names, the different subtests on the BK1-1 are not necessarily measuring different abilities. This is consistent with Silver's (1978) statement that "what is measured is not necessarily what the label on the test says" (p. 368) and supports his observation that redundancy within instruments is characteristic of many first grade readiness tests.

CHAPTER VI

SUMMARY AND CONCLUSIONS

This chapter will summarize the purpose, procedures, and major findings of the present study. Following this summary, both the limitations of this investigation and conclusions regarding the usefulness of the BK1-1 as a grade one readiness screening device will be discussed. Finally, recommendations for further practice and research will be presented within the context of this discussion.

Summary and Major Findings

A review of the literature indicates that there is a great deal of interest in and support for screening children prior to grade one in order to predict which students are likely to experience academic difficulty in the early grades (e.g., Adelman, 1978; Keogh & Becker, 1973; Maitland et al., 1974; Racciopi, 1982; Satz & Fletcher, 1979). This interest and support rests on the assumption that if children with potential academic problems are identified early enough, they can be referred for further assessment and/or intervention so that potential problems can be prevented or minimized (e.g., Evans & Ferguson, 1974; Friedman et al., 1980; Horn & Packard, 1985; Lindsay & Wedell, 1982; Vacc et al., 1987). However, in order for an early identification program to be maximally effective, the screening tests that are used to identify children requiring further assessment or intervention must be valid and accurate predictors of future academic performance (Evans & Ferguson, 1974; Piersel & Kinsey, 1984; Telegdy, 1975).

School readiness tests are routinely used as screening devices in most school districts (Flynn & Flynn, 1978; Lindsay & Wedell, 1982; Maitland et al., 1974). Several school readiness tests exist for which predictive validity evidence has been established (e.g., Gullo et al., 1984; Klein, 1977; Nagle, 1979; Nichta et al., 1982;

Piersel & Kinsey, 1984; Vacc et al., 1987). However, the validity of many new school readiness tests for making predictions about future academic performance has not been established (Adelman, 1978; Lewis, 1980; Lindsay & Wedell, 1982; Lindquist, 1982). One such test is the recently published Brigance K and 1 Screen for Kindergarten and First Grade (BK1) (Brigance, 1982). Although the BK1-1 is easily administered and cost-effective, which makes it a potentially useful screening device, its effectiveness for predicting school success and identifying children with potential learning problems has not been examined. Therefore, the purpose of this study was to systematically evaluate the the predictive validity of the BK1-1 on a longitudinal basis using a retrospective design.

The BK1-1 scores for a total sample of 149 students (80 boys and 69 girls) were used to predict indices of academic ability (i.e., CCAT scores) and achievement (i.e., GMRT, CAT, and teacher-assigned marks) at the end of grades one, two, and three. Following the calculation of descriptive statistics, these data were subjected to both correlational and classificational analysis to establish the predictive effectiveness of the BK1-1. In addition, a stepwise multiple regression analysis was used to determine which of the thirteen BK1-1 subtests were maximally effective for predicting academic ability and achievement.

On the the basis of the descriptive statistics, it was established that girls scored significantly higher than boys both in terms of their total test score on the BK1-1 and grade one reading achievement. Gender differences also emerged from the correlational analysis which indicated that correlations were generally stronger for girls than boys and that the BK1-1 was a significantly better predictor of grade three reading achievement for girls than boys. In terms of predictive effectiveness, the correlational analysis indicated that the BK1-1 was moderately predictive of academic ability and achievement for the total group in the early grades with correlations comparable to those for many existing school readiness tests (see Table 32). However, despite these moderate correlations, the BK1-1 was found to possess an extremely low level of predictive accuracy for making screening decisions about individual students - a finding which casts considerable doubt on the instrument's practical value for educators. Finally, of the thirteen BK1-1 subtests only one, Recognition of Lowercase Letters (RLL), consistently emerged as a potentially valuable predictor of later academic achievement. Both the implications of these findings and their relationship to existing literature were elaborated in Chapter 5.

Limitations and Recommendations for Future Research

Certainly the present study had limitations which must be considered when interpreting the results. However, prior to presenting specific limitations, two general limitations that are relevant to the majority of studies that have attempted to establish the predictive validity of school readiness tests will be discussed. First, stability is necessary to obtain high estimates of predictive validity; that is, children must be in roughly the same position on a criterion variable as they were on the predictor variable for high levels of predictive accuracy to be attained. However, children in the five to seven year old range represent anything but a stable group (Weerdenburg, 1983). The rapid developmental changes, behavioral fluctuations, and wide variability in early experiential background that are associated with this age group have a direct impact on both the reliability and predictive validity of school readiness screening instruments (Paget & Nagle, 1986). Lichenstein (1982) has summarized this concern by stating that:

> Individual differences in the rate and nature of developmental changes place limitations on the extent to which early deficits are indicative of the need for special intervention. Furthermore, measurement error is introduced by

characteristics of young children (e.g., attention span, motivation, unfamiliarity with the testing situation)(p. 70).

Therefore, the problem of accurately predicting later school achievement from performance on a readiness test may lie in the nature of the child rather than in the instrument (Flynn & Flynn, 1978).

A second general limitation is related to the basic nature of school readiness tests. It is important to emphasize that the assumptions underlying the practice of school readiness testing were borrowed from a medical or physical disease model of early identification (see pp. 9-10) which assumes that the condition to be identified already exists within an individual (Frankenburg, 1985; Racciopi, 1982). By applying the medical model of early identification to education, learning problems are equated with illness and are considered to be the result of deficiencies within the child - the child-deficit model (Adelman, 1978; Fedoruk, 1988; Hall & Keogh, 1978; Keogh & Becker, 1973). This rationale is reflected in the content of school readiness tests such as the BK1-1 which focus solely on the assessment of skills that are child-specific. However, a one-to-one correspondence between performance on a readiness test and later academic achievement cannot be assumed because many influences external to the child, which are not assessed by readiness tests, also contribute to academic success and failure in the early grades and likely account for many of the observed prediction errors made on the basis of readiness tests scores (Adelman, 1978; Feshbach et al., 1974; Freebody & Rust, 1985; Hall & Keogh, 1978; Jansky, 1978; Keogh & Becker, 1973; Paget & Nagle, 1986). Silver (1978) has succinctly summarized this ecological point of view by stating that "risk is not simply a status or condition of the child but also a function of the school, the classroom atmosphere, the teacher, and the nature of the instructional program" (p. 370). Therefore, a fruitful area for future research in the area of school readiness testing would be to incorporate some of these intervening variables into the equation. In fact, some research efforts of this nature are already in progress (e.g., Fedoruk, 1988).

In addition to these two general issues, specific limitations of the present study must also be considered. First, the restrospective nature of the study meant that many variables were not under the direct control of the researcher. For example, it was not possible to control variables related to the effectiveness of the teachers to whom the students were exposed in kindergarten through grade three. In addition, administrative factors such the time of year when achievement tests were administered (see p. 67 for an explanation of this point), how test results were recorded (see p. 67), and the availability of teacher-assigned marks (see p. 56) were primarily a result of the retrospective design. However, although these factors likely affected the purity of the results, they also reflect the day-to-day realities of the educational system which does not operate under highly controlled or experimentally "sterile" conditions.

A second limitation of this study was that only two broad groups (i.e., "at risk" vs. "not at risk") were used in the classificational analyses. This likely lowered the predictive accuracy estimates for the BK1-1 because according to both Mercer et al. (1979b) and Jansky (1978), screening tests generally predict best at the extremes of ability. Because the detection of mild academic difficulties is very difficult (Adelman, 1978; Paget & Nagle, 1986) it may have been preferable to have used more than two risk categories (e.g., severe risk, moderate/mild risk, no risk) or a comparison of the top and bottom quartiles in the group to calculate the classificational analyses. However, such a procedure would have required larger sample sizes. In addition, although the cut-off scores could have been adjusted so that only the most extreme cases were identified, Adelman (1978) has stated that these children would likely be selected by the teacher without the assistance of a test.

A third limitation of this study is that it treated all children who experienced academic difficulty in the early grades as a homogeneous group. However, according to Hall and Keogh (1978), "high risk" is not a unitary condition - children experience difficulty in school for a variety of reasons. Horn and O'Donnell (1984) have distinguished between academic difficulties that are expected (e.g., due to low intellectual ability) and those that are unexpected (e.g., due to a learning disability). No attempt was made to differentiate between these two groups in the present study and, as suggested by Horn and Packard (1985), it is unreasonable to expect a single screening test to be sensitive to all types of potential academic difficulty. Perhaps future research efforts can attempt to make this differentiation when evaluating the predictive effectiveneess of instruments.

Fourth, students who had received remedial assistance (i.e., resource room), those who had repeated a grade, and those who had been included in the school's 1986-87 enrichment program were not excluded from the sample. Therefore, it was not possible to compare children who were designated as "at risk"/"not at risk" on the BK1-1 and who received special assistance with a comparable sample of "at risk"/"not at risk" children who did not receive specialized attention. However, this limitation is essentially unavoidable in any predictive validation study because it would be unethical to conduct a prospective experimental study in which below average and superior students were not provided with appropriate intervention.

A fifth limitation was that although the relationship of the BK1-1 to intellectual ability was addressed, intellectual ability level was not considered when the relationship of the BK1-1 to indices of academic achievement was examined. It is commonly accepted that intellectual ability is strongly related to early school success (Fedoruk, 1988; Horn & Packard, 1985). Including intellectual ability as a predictor in the stepwise multiple regression analysis would have established the independent relationship of the

BK1-1 to indices of academic achievement. Additionally, Jansky (1978) has stated that it would be useful to know whether readiness tests are equally predictive for children who differ in intelligence. Therefore, in future research, the interaction of intellectual ability with a readiness test for predicting academic outcomes could be tested using a *Z*-test for testing differences among several independent correlation coefficients (Glass & Hopkins, 1984, p. 309). The result of this procedure would indicate whether the predictive effectiveness of a readiness test is different for different ability levels (e.g., low, average, high).

A final limitation of the present study is that the sample sizes at the grade two and three levels were relatively small. Larger sample sizes would have been useful for more clearly establishing trends related to predictive effectiveness and gender differences. Additionally, larger sample sizes would have allowed for more extensive stepwise multiple regression analyses.

Conclusions and Recommendations for Practice

Despite the previously described limitations, the results of the present study indicate a clear need to reconsider the educational utility of the BK1-1 as a grade one readiness screening device from both practical and psychometric perspectives. To achieve this objective, several statements from the BK1-1 test manual will be evaluated within the context of the present results.

The BK1-1 manual states that the skills included in the test are those "having the greatest predictive validity for success in . . . grade one" (Brigance, 1982, p. iii). However, on the basis of the present results, this statement is not supported because only one skill, the ability to name lowercase letters, had a moderate relationship to grade one reading achievement. Therefore, although the names of the subtests on the BK1-1 correspond to skills that have been found to be moderately predictive of early school success (see Horn & Packard, 1985), the skills as specifically measured by the

BK1-1 do not have validity for predicting academic achievement in grades one to three. This supports Silver's (1978) claim that subtest names are not always indicative of the skills they purport to measure. In addition, as discussed in Chapter 2 (see p. 40), the validity of the skills selected for inclusion on the BK1-1 was basically inferred from their similarity to items on existing school readiness tests. The present findings highlight the danger of making such inferences and supports Helfeldt's (1984) claim that the validity of test items cannot be assumed simply on the basis of their similarity to items on existing instruments.

To summarize, a basic assumption underlying school readiness screening instruments is that test performance is a valid predictor of future academic performance (Evans & Ferguson, 1974; Piersel & Kinsey, 1984; Telegdy, 1975). However, on the basis of the present information, the BK1-1 does not fulfill this basic requirement. Therefore, it is inadvisable to make predictive inferences about academic success on the basis of scores that children attain on the BK1-1.

The BK1-1 manual also suggests that the test can be given in the spring to identify skill areas than can be worked on over the summer and that the results of the BK1-1 can be used to assist teachers with planning appropriate educational programs for children. Both these statements presuppose that the skills included on the BK1-1 are relevant to early school success - an assumption that the previously discussed results does not support. This is problematic because, according to Flynn and Flynn (1978), if teachers are basing programs on diagnostic data from a certain test, they must be certain that the test's scores are actually related to future achievement.

These two statements from the BK1-1 manual also imply that training children in the specific skills of the BK1-1 will increase their chances of academic success. This is not an educationally sound implication because it suggests a causal relationship between mastery of the skills on the BK1-1 and later academic achievement. Several researchers have emphasized the fact that there is no simple casual chain between readiness test performance and reading achievement (Eaves et al., 1974; Jansky, 1978; Keogh & Becker, 1973; Lindsay & Wedell, 1982). In fact, Telegdy (1975) has cited evidence to demonstrate that although training children daily in letter recognition skills improved their ability to quickly identify these symbols, there was no significant effect on their reading level. Telegdy concluded that there is a "very imminent danger of training children on readiness skills. Such training may lead only to test sophistication without any transfer to new learning situations" (p. 9). Similarly, Mann (1984) has asserted that remedying observable symptoms of reading difficulty such as poor letter recognition skills will not necessarily improve reading ability. Clearly, training children on the specific skills of any readiness test, including the BK1-1, is not educationally useful. In addition, it cannot be assumed that the skills included on the BK1-1 are compatible with all grade one curricula. This is important to consider because according to both Freebody and Rust (1985) and Salvia and Ysseldyke (1981), the validity of school readiness tests for predicting early academic success depends on the nature of the instructional program to which a child is exposed. From this perspective it is logical to assume that although the content of the BK1-1 may be compatible with a grade one program that emphasizes the mastery of specific sub-skills, it would not likely be compatible with a grade one program that emphasizes a more holistic approach. Therefore, in light of the current holistic emphasis of the Alberta Department of Education in the primary grades (i.e., program continuity, whole language approach to language arts instruction), the use of the BK1-1 as a screening tool is of questionable value. An interesting area for future research would be to examine whether various school readiness tests do in fact have differential predictive validity for top-down vs. bottom-up approaches (see Otto, 1982) to grade one reading instruction.

Finally, the BK1-1 manual also suggests that scores from the test can be used to identify children requiring a more comprehensive evaluation and to help determine appropriate placement and grouping of students. Neither of these suggested uses of the test scores is warranted on the basis of the results obtained from the classificational analysis of the data in this study. The accuracy of an "at risk" designation on the BK1-1 was much too low to consider using the test to identify children requiring a more comprehensive assessment let alone determining appropriate student grouping and placements. The high level of inaccuracy associated with an "at risk" designation would result in a large number of unnecessary referrals which would basically cancel out any cost-benefits associated with the instrument. Furthermore, many children would be inappropriately grouped or placed on the basis of their BK1-1 scores.

In summary, the BK1-1 is a brief, low-cost screening device but the results of the present study have clearly indicated that it is not effective for predicting future academic outcomes - a necessary feature for an educationally useful screening device. Frankenburg (1985) has commented that a major problem with screening programs is the failure to confine screening to tests having a high degree of predictive validity and accuracy. Therefore, it would be advisable for schools currently using the BK1-1 to carefully reconsider the reasons why they have chosen to use this instrument.

More specifically, schools should reevaluate their use of the BK1-1. First, it must be determined whether the information gained from the BK1-1 is of practical use to and consistent with the instructional approaches of the kindergarten and grade one teachers in the schools. This is an important issue to consider because according to Frankenburg (1985) another major problem with screening programs is the failure to ensure that the information supplied by the screening procedure will be considered appropriate by those who give the treatment - in this case, the grade one teachers. It may be useful for the grade one teachers to meet and develop a list of skills that they feel are necessary to success in their programs and then compare this list of skills to those assessed by the BK1-1.

If the primary concern of teachers and administration is one of prediction and referring individual children for further assessment, it is recommended that an instrument with a high degree of predictive validity such as the Metropolitan Readiness Test (Nurss & McGauvran, 1976) be used. However, if the concern is mainly one of determining a child's proficiency with specific skills, a school-based checklist of skills developed by the teachers would likely be more useful than the BK1-1 because the closer an evaluation is to a specific learning situation, the more accurate it will be (Paget & Nagle, 1986).

In addition, Horn and Packard (1985) have found that teacher ratings of behavioral and socio-emotional indicators (e.g., attention/distractibility) are one of the stronger predictors of early school sucess. However, despite the potential predictive strength of these indicators, the BK1-1 does not include a formal assessment of these variables. Several studies have also indicated that a kindergarten teacher's ratings can predict academic success in grade one at least as well as a school readiness test (e.g., Barnes, 1985; Feschbach et al., 1977; Lindsay & Wedell, 1982). Therefore, a kindergarten teacher rating scale (e.g, Myklebust Pupil Rating Scale, Myklebust, 1971) which includes an assessment of behavioral and socio-emotional skills would be a valuable addition to a school readiness screening program.

Finally, teachers and administration must determine why they actually use the BK1-1, or any school readiness instrument. They must assess the benefit of the information gained from the test and how that information will be practically used to help children. The question must not simply be one of "How do we screen children?" but "Why do we screen children?" (Hall & Keogh, 1978). Establishing a rationale for the use of school readiness tests is necessary because without a clear purpose and statement of expected outcomes, school readiness testing can simply become a routine procedure that lacks educational relevance (Badian, 1982). The entire issue of predictive accuracy in school readiness screening is meaningless unless the predictions based on the test results lead to better educational opportunities for students (Wilson & Reichmuth, 1985). Therefore, school readiness testing is only meaningful in the context of a total early identification program in which the results of screening lead to further assessment and ultimately, to appropriate intervention (Adelman, 1978; Badian, 1982; Book, 1980; Eaves et al., 1974; Keogh & Becker, 1973; Paget & Nagle, 1986; Racciopi, 1982; Silver, 1978; Weerdenburg, 1983).

REFERENCES

Abrahamson, D., & Bell, A. (1979). Assessment of the school readiness section of the Early Detection Inventory: Preschool prediction across situational factors. Journal of School Psychology, 17 (2), 162 - 171.

Adair, N. & Blesch, G. (1965). The ABC Inventory: To determine kindergarten and school readiness. Muskegon, Mich.: Educational Studies and Development.

Adelman, H. S. (1978). Predicting psychoeducational problems in childhood. Behavior Disorders, 3, 148 - 159.

Ahr, E. (1966). Screening Test of Academic Readiness. Skokie, Ill.: Priority Innovations.

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

Badian, N. A. (1982). The prediction of good and poor reading before kindergarten entry: a four-year follow-up. The Jounral of Special Education, 16 (3), 309 - 318.

Badian, N. & Serwer, B. (1975). The identification of high risk children: A retrospective look at selection criteria. Journal of Learning Disabilities, 8 (5), 27 - 31.

Barnes, K. (1985). The Jansky Screening Index: A seven-year predictive evaluation and comparative study. In W. K. Frankenburg, R. N. Emde, & J. W. Sullivan (Eds.), Early identification of children at risk: An international perspective (pp. 185 - 191). New York: Plenum Press.

Boehm, A. (1985). Review of Brigance K and 1 Screen for Kindergarten and First Grade. In James V. Mitchell, Jr. (Ed.), The Ninth Mental Measurements Yearbook: Volume 1 (pp. 223 - 225). Lincoln, Nebraska: University of Nebraska Press.

Bolig, J. & Fletcher, G. (1973). The MRT vs. ratings of kindergarten teachers as predictors of success in first grade. Educational Leadership, 30, 637 - 640.

Book, R. M. (1974). Predicting reading failure: A screening battery for kindergarten children. Journal of Learning Disabilities, 7 (1), 52 - 56.

Book, R. M. (1980). Identification of educationally at-risk children during the kindergarten year: A four-year follow-up study of group test performance. Psychology in the Schools, 17, 153 - 158.

Bremer, N. (1959). Do readiness tests predict success in reading? The Elementary School Journal, 59, 222 - 224.

Brigance, A. (1982). The Brigance K and 1 Screen for Kindergarten and First Grade. North Billerica, MA: Curriculum Associates.

Burns, P. & Roe, B. (1980). Teaching Reading in Today's Elementary Schools (2nd ed.). Boston: Houghton Mifflin Company.

Chew, A. & Morris, J. (1984). Validation of the Lollipop Test: A diagnostic screening test of school readiness. Educational and Psychological Measurement, 44, 987 - 991.

Collis, B., Collis, L., & Yore, L. (1986). Predictive validity of the Canadian Readiness Test and selected measures of cognitive development for grade 1 reading and mathematics achievement. The Alberta Journal of Educational Research, 32, 2 - 11.

Cronbach, L. J. (1971). Essentials of psychological testing (Third Edition). New York: Harper and Row.

CTB/McGraw-Hill (1977). The California Achievement Tests. Monterey, CA: CTB/McGraw-Hill.

CTC/McGraw-Hill (1981). The Canadian Achievement Tests, Level 12, Form A, Examiner's Manual. Scarborough, Ontario: Canadian Test Center, McGraw-Hill Ryerson Ltd.

CTC/McGraw-Hill (1983). The Canadian Achievement Tests, Technical Bulletin.Scarborough, Ontario: Canadian Test Center, McGraw-Hill Ryerson Ltd.

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

Dietz, C. & Wilson, B. (1985). Beginning school age and academic achievement. Psychology in the Schools, 22, 93 - 94.

DiPasquale, G., Moule, A., & Flewelling, R. (1980). The birthdate effect. Journal of Learning Disabilities, 13, 4 - 8.

Doherty, J. & Conolly, M. (1985). How accurately can primary school teachers predict the scores of their pupils in standardized tests of attainment? A study of some non-cognitive factors that influence specific judgements. Educational Studies, 11, 41 - 60.

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

Dunn, L. & Markwardt, F. (1970). Peabody Individual Achievement Test. Circle Pines, Minn.: American Guidance Service.

Durost, W., Bixler, H., Wrightstone, G., Prescott, G., & Balow, J. (1976). Metropolitan Achievement Test. New York: Harcourt Brace Jovanovich, Inc.

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

Evans, R. & Ferguson, N. (1974). Screening school entrants. Association of Educational Psychologists (A.E.P.) Journal, 3 (6), 2 - 9.

Fedoruk, G. M. (1988). Unpublished doctoral dissertation proposal. University of Alberta, Edmonton.

Ferguson, G. (1981). Statistical Analysis in Psychology and Education. New York: McGraw-Hill Book Company.

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

Feshbach, S., Adelman, H. & Fuller, W. (1977). Prediction of reading and related academic problems. Journal of Educational Psychology, 69 (4), 299 - 308.

Feuerstein, R. (1979). Liberating Human Potential. The Dynamic Assessment of Retarded Performance (pp. 1 - 56). Baltimore: University Park Press.

Flynn, T., & Flynn, L. (1978). Evaluation of the predictive ability of five screening measures administered during kindergarten. Journal of Experimental Education, 46 (3), 65 - 70.

Frankenburg, W. (1985). The concept of screening revisited. In W. K. Frankenburg, R. N. Emde, & J. W. Sullivan (Eds.), Early identification of children at risk: An international perspective (pp. 3 - 17). New York: Plenum Press.

Frankenburg, W., Dodds, J., Fandal, A., Kazuk, E., & Cohrs, M. (1975). Denver Developmental Screening Test. Denver, CO; LADOCA Project and Publishing Foundation.

Freebody, P. & Rust, P. (1985). Predicting reading achievement in the first year of schooling: A comparison of readiness tests and instructional programs. Journal of School Psychology, 23, 145 - 155.

Friedman, R., Fuerth, J., & Forsythe, A. (1980). A brief screening battery for predicting school achievement at ages seven and nine years. Psychology in the Schools, 17, 340 - 346.

Garrett, H. (1954). Elementary Statistics. New York: Longmans, Green.

Glass, G. V., & Hopkins, K. D. (1984). Statistical Methods in Education and Psychology. Englewood Cliffs, NJ: Prentice Hall.

Glazzard, M. (1977). The effectiveness of three kindergarten predictors for first-grade achievement. Journal of Learning Disabilities, 10 (2), 36 - 40.

Gray, W. (1963). Gray Oral Reading Test. Indianapolis: Bobbs-Merrill.

Gredler, G. (1978). A look at some important factors in assessing readiness for school. Journal of Learning Disabilities, 11 (5), 25 - 31.

Gullo, D., Clements, D., & Robertson, L. (1984). Prediction of academic achievement with the McCarthy Screening Test and Metropolitan Readiness Test. Psychology in the Schools, 21, 264 - 269.

Hainsworth, P. & Siqueland, M. (1969). Early identificaiton of children with learning disabilities: The Meeting Street School Screening Test. Providence, RI.: Cripple Children and Adults of Rhode Island.

Hall, R. & Keogh, B. (1978). Qualitative characteristics of educationally high-risk children. Learning Disability Quarterly, 1, 62 - 68.

Hambleton, R. & Eignor, D. (1978). Guidelines for evaluating criterion-referenced tests and test manuals. Journal of Educational Measurement, 15 (4), 321 - 324.

Hase, H. (1977)). Predicting performance in the first grade with the First Grade Screening Test. Psychology in the Schools, 14 (4), 407 - 412.

Helfeldt, J. (1984). Test review: The Brigance K and 1 Screen for Kindergarten and First Grade. The Reading Teacher, May, 820 - 824.

Hildreth, G. H., Griffiths, N. L., & McGauvran, M. E. (1965). Metropolitan Readiness Tests. New York: Harcourt Brace Jovanovich.

Horn, W. & O'Donnell, J. (1984). Early identification of learning disabilities: A comparison of two methods. Journal of Educational Psychology, 76, 1106 - 1118.

Horn, W., & Packard, T. (1985). Early identification of learning problems: A meta-analysis. Journal of Educational Psychology, 77 (5), 597 - 607.

Ilg, F. & Ames, L. (1965). School Readiness. New York: Harper & Row Publishers, Inc.

Jansky, J. (1978). A critical review of "Some developmental and predictive precursors of reading disabilities". In A. L. Benton & D. Pearl (Eds.), Dyslexia: An appraisal of current knowledge (pp. 378 - 394). New York: Oxford University Press.

Jastak, J., Bijou, S., & Jasktak, S. (1965). Wide Range Achievement Test - Revised Edition. Delaware: Guidance Associates.

Kaplan, R., & Saccuzo, D. (1982). Psychological testing: Principles, applications, and issues. Monterey, CA: Brooks/Cole Publishing Co.

Kelley, T., Madden, R., Gardner, E., & Rudman, H. (1964). Stanford Achievement Test - Primary I Battery. New York: Harcourt, Brace & World.

Kelley, T., Madden, R., Gardner, E., & Rudman, H. (1963). Stanford Achievement Test - Primary II Battery. New York: Harcourt, Brace & World.

Keogh, B. & Becker, L. (1973). Early detection of learning problems: Questions, cautions, and guidelines. Exceptional Children, 40 (1), 5 - 11.

Keogh, B. K. & Daley, S. (1983). Early identification: One component of comprehensive services for at-risk children. Topics in Early Childhood Special Education, 3, 7 - 16.

Keogh, B. & Smith, C. (1970). Early identification of educationally high potential and high risk children. Journal of School Psychology, 8 (4), 285 - 290.

King, E., Hieronymus, A., Lindquist, E., & Hoover, H. (1982). Canadian Test of Basic Skills, Multilevel Edition, Levels 9 - 14, Forms 5 & 6, Teacher's Guide. Scarborough, Ontario: Nelson Canada Ltd.

Klein, A. (1977). The validity of the Screening Test of Academic Readiness in predicting achievement in first and second grades. Educational and Psychological Measurement, 37, 493 - 499.

Kornblau, B. & Keogh, B. (1980). Teachers' perceptions and educational decisions. In J. J. Gallagher (Ed.), New directions for exceptional children: No. 1 - Ecology of exceptional children (pp. 87 - 101). San Francisco: Jossey - Bass.

Leiter, J. & Brown, J. (1985). Determinants of elementary school grading. Sociology of Education, 58, 166 - 180.

Lesiak, J. (1984). The Bender Visual-Motor Gestalt Test: Implications for the diagnosis and prediction of reading achievement. Journal of School Psychology, 22, 391 - 405.

Lessler, K. & Bridges, J. (1973). The prediciton of learning problems in a rural setting: Can we improve on readiness tests? Journal of Learning Disabilities, 6(2), 36 - 40.

Lewis, A. (1980). The early identification of children with learning difficulties. Journal of Learning Disabilities, 13 (2), 51 - 57.

Lichenstein, R. (1981). Comparative validity of two preschool screening tests: Correlational and classificational approaches. Journal of Learning Disabilities, 14 (2), 68 - 72.

Lichenstein, R. (1982). New instruments, old problems for early identification. Exceptional Children, 49, 70 - 72.

Lichenstein, R., & Ireton, H. (1984). Preschool screening: Identifying young children with developmental and educational problems. Orlando, FL: Grune & Stratton, Inc.

Lindquist, G. T. (1982). Preschool screening as a means of predicting later reading achievement. Journal of Learning Disabilities, 15 (6), 331- 332.

Lindsay, G. A., & Wedell, K. (1982). The early identification of educationally 'at risk" children revisited. Journal of Learning Disabilities, 15 (4), 212 - 217.

MacGinitie, W. (1980). Gates-MacGinitie Reading Tests, Canadian Edition (Level A), Teacher's Manual. Scarborough, Ontario: Nelson Canada Ltd.

MacGinitie, W. (1978). Gates-MacGinitie Reading Tests. Boston: Houghton Mifflin.

Maitland, S., Nadeau, J., & Nadeau, G. (1974). Early school screening practices. Learning Disabilities, 7 (10), 55 - 59.

Mann, V. (1984). Longitudinal prediction and prevention of early reading difficulty. Annals of Dyslexia, 34, 117 - 136.

Mardell, C. D. & Goldenberg, D. S. (1975). Developmental Indicators for the Assessment of Learning. Edison, NJ: Childcraft.

May, D. & Welch, E. (1986). Screening for school readiness: The influence of birthdate and sex. Psychology in the Schools, 23, 100 - 105.

McInnis, C. (1985). Canadian Cognitive Abilities Test. In D. Keyser & R. Sweetland (Eds.), Test Critiques, Volume V. Kansas City, Missouri: Test Corporation of America.

Mercer, C., Algozzine, B., & Trifiletti, J. (1979a). Early identification: Issues and considerations. Exceptional Children, 46 (10), 52 - 54.

Mercer, C., Algozzine, B., & Trifiletti, J. (1979b). Early identification: An analysis of the research. Learning Disability Quarterly, 2, 12 - 24.

Myklebust, H. (1971). The Pupil Rating Scale. New York: Grune & Stratton.

Nadon, E. (1983). Early Indices of Grade One Performance. Unpublished master's thesis, University of Alberta, Edmonton, Alberta.

Nagle, R. (1979). The predictive validity of the MRT (1976 edition). Educational and Psychological Measurement, 39, 1043 - 1045.

Naslund, R., Thorpe, L., & Lefever, D. (1978). Science Research Associates Achievement Series. Chicago: Science Research Associates.

Nichta, L., Federici, L., & Schuereger, J. (1982). The Screening Test of Academic Readiness (STAR) as a predictor of third-grade achievement. Psychology in the Schools, 19, 190 - 193.

Nurss, J. R., & McGauvran, M. E. (1976). Metropolitan Readiness Tests, Level II, Teacher's Manual. New York: Harcourt Brace Jovanovich.

Otto, J. (1982). The new debate in reading. The Reading Teacher, 36, 14 - 18.

Paget, K., & Nagle, R. (1986). A conceptual model of preschool assessment. School Psychology Review, 15, 154 - 165.

Pate, J. & Webb, W. (1966). First Grade Screening Test. Circle Pines, Minn.: American Guidance Service.

Piersel, W., & Kinsey, J. (1984). Predictive validity of the First Grade Screening Test. Educational and Psychological Measurement, 44, 921 - 924.

Raccioppi, R. (1982). The promise of public law 94-142: Issues in the diagnostic process. Journal of Learning Disabilities, 15(4), 218 - 220.

Randel, M., Fry, M., & Ralls, E. (1977). Two readiness measures as predictors of first- and third-grade reading achievement. Psychology in the Schools, 14 (1), 37 - 40.

Randhawa, B., Hunt, D., & Rawlyk, S. (1974). Factorial structure, reliability, validity, effectiveness, and efficiency of the Canadian Cognitive Abilities Test. The Alberta Journal of Educational Research, 20 (2), 208 - 215.

Reinherz, H. (1977). Identifying children at risk. Health Education, 8 (4), 14 - 16.

Reynolds, C. (1979). A factor analytic study of the Metropolitan Readiness Test. Contemporary Educational Psychology, 4, 315 - 317.

Robinson, J., & Kovacevich, D. (1985). The Brigance Inventories. In D. Keyser and R. Sweetland (Eds.), Test Critiques, Volume III. Kansas City, Missouri: Test Corporation of America.

Rubin, R., Balow, B., Dorle, J. & Rosen, M. (1978). Preschool prediction of low achievement in basic school skills. Journal of Learning Disabilities, 11 (10), 62 - 65.

Salvia, J., & Ysseldyke, J. E. (1981). Assessment in Special and Remedial Education (2nd ed.). Boston: Houghton Mifflin.

Satz, P., & Fletcher, J. (1979). Early screening tests: Some uses and abuses. Journal of Learning Disabilities, 12 (1), 65 - 69.

Satz, P. & Friel, J. (1974). Some predictive antecedents of specific reading disability: A preliminary two-year follow-up. Journal of Learning Disabilities, 7 (7), 48 - 55.

Satz, P., Taylor, H., Friel, J., & Fletcher, J. (1978). Some developmental and predictive precursors of reading disabilities: A six-year follow-up. In A. L. Benton & D. Pearl (Eds.), Dyslexia: An appraisal of current knowledge (pp. 313 - 348). New York: Oxford University Press.

Schmidt, S. & Perino, J. (1985). Kindergarten screening results as predictors of academic achievement, potential, and placement in second grade. Psychology in the Schools, 22, 146 - 151.

Silver, A. (1978). Prevention. In A. L. Benton & D. Pearl (Eds.), Dyslexia: An appraisal of current knowledge (pp. 350 - 376). New York: Oxford University Press.

Simner, M. (1985). School readiness and the Draw-A-Man Test: An empirically derived alternative to Harris' scoring system. Journal of Learning Disabilities, 18 (2), 77 - 82.

Smith, M. & Glass, G. (1987). Research and Evaluation in Education and the Social Sciences. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Stahl, S. (1985). Gates-MacGinitie Reading Tests, Second Edition. In D. Keyser & R. Sweetland (Eds.), Test Critiques, Volume IV. Kansas City, Missouri: Test Corporation of America.

Steinbauer, E. & Heller, M. (1978). The Boehm Test of Basic Concepts as a predictor of academic achievement in grades 2 and 3. Psychology in the Schools, 15(3), 357 - 360.

Stevenson, H., Parker, T., Wilkinson, A., Hegion, A., & Fish, E. (1976). Predictive value of teacher's ratings of young children. Journal of Educational Psychology, 68 (5), 507 - 517.

Swanson, B., Payne, D., & Jackson, B. (1981). A predictive validity study of the Metropolitan Readiness Test and Meeting Street School Screening Test against first grade Metropolitan Achievement Test scores. Educational and Psychological Measurement, 41, 575 - 578.

Sweetland, J. & DeSimone, P. (1987). Age of entry, sex, and academic achievement in elementary school children. Psychology in the Schools, 24, 406 - 412.

Telegdy, G. A. (1974a). A factor analysis of four school readiness tests. Psychology in the Schools, 11 (1), 127 - 133.

Telegdy, G. A. (1974b). The relationship between socioeconomic status and school readiness. Psychology in the Schools, 11, 351 - 356.

Telegdy, G. A. (1975). The effectiveness of four readiness tests as predictors of first grade academic achievement. Psychology in the Schools, 12 (1), 4 - 11.

Thorndike, R., & Hagen, E. (1982). Canadian Cognitive Abilities Test, Primary Battery, Levels 1 & 2 (Form 3), Examiner's Manual. Scarborough, Ontario: Nelson Canada.

Tsushima, W., Onorato, V., Okumura, F., & Sue, D. (1983). The predictive validity of the STAR: A need for local validation. Educational and Psychological Measurement, 43, 663 - 665.

Vacc, N., Vacc, N., & Fogleman, M. (1987). Preschool screening using the DIAL as a predictor of first-grade performance. Journal of School Psychology, 25, 45 - 51.

Van Evra, J. (1983). Prevention. Psychological Disorders of Children and Adolescents (pp. 332 - 334). Boston: Little, Brown & Co.

Watkins, E. & Wiebe, M. (1984). Factor analysis of the Metropolitan Readiness Test on first grade children. Psychology in the Schools, 21, 429 - 432.

Weerdenburg, G. (1983). Predicting grade one success. Unpublished master's thesis, University of Alberta, Edmonton, Alberta.

Whyte, L. (1985). A review of Canadian Achievement Tests, Form A. In J. V. Mitchell, Jr. (Ed.), The Ninth Mental Measurements Yearbook: Volume 1 (pp. 257 - 258). Lincoln, Nebraska: University of Nebraska Press.

Wick, J. W. & Smith, J. K. (1981). Scott, Foresman Achievement Series. Glenview, IL.: Scott Foresman.

Wilson, B. J., & Reichmuth, M. (1985). Early screening programs: When is predictive accuracy sufficient? Learning Disability Quarterly, 8, 182 - 188.

Wolf, M. (1978). Social validity: The case for subjective measurement. Journal of Applied Behavior Analysis, 11, 203 - 214.

Wright, D. (1985). Review of Brigance K & 1 Screen for Kindergarten and First Grade. In James V. Mitchell, Jr. (Ed.), The Ninth Mental Measurements Yearbook: Volume 1. .(p. 225). Lincoln, Nebraska: University of Nebraska Presss.

Appendix A

Descriptive Summary of the BK1-1

Descriptive Summary of the BK1-1

Personal Data Response: child orally provides first name, middle name, last name, age address and birthdate

Point Value = 1 pt. each Total = 5

Color Recognition: child identifies and names ten colors presented in pictures of blocks and balls (red, blue, green, yellow, orange, purple, brown, black, pink, grey)

Point Value = .5 pt. each Total = 5

Picture Vocabulary: child recognizes and names ten common pictures (dog, cat, key, girl, boy, airplane, apple, leaf, cup, car)

Point Value = 5 pt. each Total = 5

Visual-Discrimination: child indicates the symbol that is different in a series of four lower case letters or short words using a pointing response.

Point Value = 1 pt. each Total = 10

Visual-Motor Skills: child copyies a circle, square, triangle, cross, and horizontal line from a model

Point Value = 1 pt. each Total = 5

Draw-A-Person: child draws a picture of a whole person with credit given for identifiable body parts (10 criteria: head, legs, eyes, nose, mouth, arms, trunk, hands, ears, neck)

Point Value =1 pt./body part Total = 10

Rote Counting: child counts by rote to 20

Point Value = 1 pt. given for each group of four numbers in sequence (eg. 1234, 5678) Total = 5

Recites Alphabet: child recites entire alphabet orally

Point Value = 1 pt. given for each five letters in sequence (eg. abcde, fghij) Total = 5

Numeral Comprehension: child matches quantities to printed numerals up to 10 using either fingers or objects, no verbal response required.

Point Value = 1 pt. each — Total = 10

Letter Recognition: child recognizes and names all 26 lowercase letters

Point Value = .5 pt. each — Total = 13

Auditory Discrimination: child indicates whether two words sound alike or different (eg. met-met; sum-sun) Note: student must get both responses correct to receive credit for each item

Point Value = 1 pt. given for each correct set — Total = 10

Prints Personal Data: child prints first and last name

Point Value = 5 pt. for each — Total = 10

Numerals in Sequence: child prints the numbers 1 through 7 in sequence

Point Value = 1 pt. each — Total = 7

Total Score = 100

Appendix B

Data Form

Data Form

I.D. #________ Gender: M F Birthdate ______________

Occupation of Head of Household ___________

1. BK1-1 Scores (C.A. at time of administration: ____________________)
(Date of Administration: ____________________)

	Subtest	Total Possible	Obtained Score
a)	Personal Data Response	5	_______
b)	Color Recognition	5	_______
c)	Picture Vocabulary	5	_______
d)	Visual Discrimination	10	_______
e)	Visual-Motor Skills	5	_______
f)	Draw-A-Person	10	_______
g)	Rote Counting	5	_______
h)	Recites Alphabet	5	_______
i)	Numeral Comprehension	10	_______
j)	Recognition of Lowercase Letters	13	_______
k)	Auditory Discrimination	10	_______
l)	Prints Personal Data	10	_______
m)	Numerals in Sequence	7	_______
	Total Test Score	100	_______

2. CCAT Scores (Date of administration: _______________)

SAS ______

3. GMRTCE Scores

		Grade 1	Grade 2	Grade 3
		Date:_____	Date:_____	Date: _____
Vocabulary	Raw Score	_____	_____	_____
Comprehension	Raw Score	_____	_____	_____
Total Test	Raw Score	_____	_____	_____

4. CAT Scores (Grade (s)___ Date(s) of Administration: ________________)

Total Spelling	**Raw Score**	_____	_____
Phonic Analysis	Raw Score	_____	_____
Structural Analysis	Raw Score	_____	_____
Reading Vocabulary	Raw Score	_____	_____
Reading Comp.	Raw Score	_____	_____
Total Spelling	**Raw Score**	_____	_____
Total Language	**Raw Score**	_____	_____
Language Mechanics	Raw Score	_____	_____
Language Expression	Raw Score	_____	_____
Total Mathematics	**Raw Score**	_____	_____
Computation	Raw Score	_____	_____
Concepts & Applic.	Raw Score	_____	_____

Total Battery **Raw Score** _____ _____

5. Report Card Marks

	Grade 1	Grade 2	Grade 3
Teacher:	_____	_____	_____
Reading	_____	_____	_____
Written Exp.	_____	_____	_____
Mathematics	_____	_____	_____
Printing	_____	_____	_____

6. Grade Retention (Circle one) Y N

   If "Yes", state grade(s): ______________________________

7. Special Education Services

   a) Resource Room (Circle one) Y N

      If "yes", state when: ______________________________

   b) Special Education Class Placement (Circle one) Y N

      If "yes", state when and type of placement: ______________________

   c) Enrichment Program participation (Circle one) Y N

      If "yes", state when: ______________________________

Appendix C

Correspondence

January 2, 1989

(name of superintendent), Superintendent
(name of participating school district)
(name of town)
(postal code)

Dear Mr. (name of superintendent),

As you are aware, I am completing my Master's degree in Educational Psychology with a specialization in School Psychology at the University of Alberta. At the present time, I am working on my thesis under the supervision of Dr. Lorraine Wilgosh (432-3738). I have previously discussed with you the possibility of conducting my research at (name of participating school). Please accept this letter as my formal request for approval to conduct research in (name of participating school) in order to collect the data necessary to complete my Master's thesis.

In my previous position as a resource room teacher at (name of participating school), I was responsible for administering the Brigance K and 1 screening test to children at the end of kindergarten. I individually assessed the kindergarten classes of 1985, 1986, and 1987. In addition, last year when I was on a leave of absence, I assisted Mr. (name of principal) in assessing the 1988 kindergarten class. Because of my contact with the Brigance K and 1 screening test, I have developed an interest in researching its development and technical properties. Therefore, the topic I have chosen to research is: The Predictive Validity of the Brigance K and 1 Screen.

I have attached a detailed outline of my research project for your consideration. The outline presents the background and rationale of my research topic and the research questions to be investigated. In addition, the data collection procedures to be implemented, including a consideration of how the anonymity of subjects and confidentiality of information will be protected and how the the outcomes of the research will benefit (name of participating school) are described.

Thank you for your consideration of my request.

Sincerely,

Anne T. Shillington

c.c. (name of principal), Principal
(name of participating school)

February 8, 1989

Mrs. Anne Shillington
5708 - 50 Street
STONY PLAIN, AB.
TOE 2GO

Dear Anne:

The Board of Trustees on January 9, 1989, approved your request to conduct research at __________ Catholic School to collect data for your thesis. This approval was subject to the guidelines and procedures to be worked out in consultation with the Superintendent.

As you are aware, the current School Act Section (18) addresses student records and there have been draft regulations with respect to the area of access to student records. Under the proposed regulation, access is available only to parents or students if 16 or older, to an employee of the board if need be, or to anyone with written permission from the parent or parent and student.

To facilitate your request it will be necessary for us to obtain permission from parents concerned before you can have access to the records. I am prepared to send a letter to the parents' outlining your request and soliciting parental permission. When I have received the parent authorization, I will give you access to the student's file. The files may not be removed from the school.

I trust that this arrangement will be suitable to accomodate your needs.

Sincerely,

J. Collins
J. COLLINS
SUPERINTENDENT

JC/blf

February 14, 1989
(April 3, 1989)

Dear ____________________________,

My name is Mrs. Anne Shillington and I was employed as the Resource Room (T.M.) teacher at (name of participating school)from 1984 - 1987. When your child, ____________, finished kindergarten and/or entered grade one at (name of participating school) in _______________, I administered the Brigance Grade One Screening Test to him/her. This test assesses skills which are commonly thought to be necessary for success in grade one (e.g., naming letters, naming colours, counting, copying shapes, etc.). The scores from this test are used to give grade one teachers a preliminary idea of childrens' areas of strength as well as areas in which they may need some extra assistance during grade one.

I am presently completing my Master's degree in Educational Psychology at the University of Alberta under the supervision of Dr. Lorraine Wilgosh. The final requirement for my degree is the completion of a research project. I have been granted approval by the (name of participating school district) School Board to conduct this project at (name of participating school). My research study is intended to determine how well the Brigance Grade One Screening Test predicts academic achievement at the end of grades one, two, and three.

In order to obtain the necessary data for my study, I require the following information:

1) your child's age and gender;
2) your child's scores from the Brigance Grade One Screening Test;
3) your child's scores from academic ability and achievement tests which have been routinely administered to all children at (name of participating school) in grades one through three;
4) your child's year-end report card marks in reading and math in grades one through three;
5) an indication of whether your child was ever included in either the T.M. or enrichment programs at (name of participating school).

All of the information mentioned above is currently included in your child's file (cumulative record) at (name of participating school). However, before I can obtain this information from your child's file, I require your written permission. You can indicate your permission by signing the attached form and returning it to the school by March 1, 1989 (April 10, 1989). Additionally, when conducting a research project, it is necessary to have a description of the average occupational level of the families of children included in the study. Therefore, please indicate the occupation of the primary wage earner in your household on the attached form.

All information will be treated as confidential. Your child's name will not be recorded on the form used to obtain the information needed for the study and the children will not be identified individually in the study. I anticipate that my study will be completed by the end of June, 1989. A copy of my study will be made available to the school should you be interested in the results. If you have any questions or concerns, please do not hesitate to contact me at my home: 963 - 5864.

Thank you in advance for your cooperation.

Sincerely,

Mrs. Anne Shillington

- -

Mrs. Anne Shillington has my permission to obtain the information necessary for her research study in my child's, ______________________________, cumulative record at
Child's Name

(name of participating school).

________________________ ____________________________
Date Parent's Signature

Occupation of the primary wage earner in our household:

____

*Please have your child return this form to the school by March 1, 1989 (April 10, 1989).

February 15, 1989

Dear Parents:

The __________ Catholic School Board has given Mrs. Anne Shillington permission to gather information for her thesis at the University of Alberta from the documentation which is available from the Brigance Screening Tests used at ______ ______ Catholic School. It is necessary that you, the parent, give signed authorization for someone other than parent or employee to have access to the student record.

Mrs. Shillington taught at __________ Catholic School for three years as teacher in the Resource Room. She was also involved in administering the tests to the students during those years.

Attached is a copy of the letter from Mrs. Shillington explaining her project in detail. With your approval, she would have access to the information at the school. The records will not leave the school.

I would appreciate your cooperation in this regard and support the work that Mrs. Shillington is undertaking.

Please return the form attached to Mrs. Shillington's letter to the __________ School Board Office at __________ Catholic School.

If you have any concerns or questions, please contact me at 963-3685.

Yours truly,

J. Collins
J. COLLINS
SUPERINTENDENT

JC/blf

encs.

Appendix D

Supplementary Tables

KEY TO VARIABLES - PREDICTOR VARIABLES

1. PDR - BK1-1 Personal Data Response Subtest Raw Score
2. CR - BK1-1 Colour Recognition Subtest Raw Score
3. PV - BK1-1 Picture Vocabulary Subtest Raw Score
4. VD - BK1-1 Visual Discrimination Subtest Raw Score
5. VM - BK1-1 Visual-Motor Skills Subtest Raw Score
6. DAP - BK1-1 Draw-A-Person Subtest Raw Score
7. RC - BK1-1 Rote Counting Subtest Raw Score
8. RA - BK1-1 Recites Alphabet Subtest Raw Score
9. NC - BK1-1 Numeral Comprehension Subtest Raw Score
10. RLL - BK1-1 Recognition of Lower Case Letters Subtest Raw Score
11. AD - BK1-1 Auditory Discrimination Subtest Raw Score
12. PPD - BK1-1 Prints Personal Data Subtest Raw Score
13. NS - BK1-1 Numbers in Sequence Subtest Raw Score
14. BKIT - BK1-1 Total Test Raw Score

KEY TO VARIABLES - CRITERION VARIABLES

---

1. CA - Chronological Age in months
2. CCAT - Canadian Cognitive Abilities Test Standard Age Score
3. GRVI - Gates-MacGinitie Reading Test (GMRT) Vocabulary Subtest Raw Score Grade 1)
4. GRV2 - GMRT Vocabulary Subtest Raw Score (Grade 2)
5. GRV3 - GMRT Vocabulary Subtest Raw Score (Grade 3)
6. GRCI - GMRT Comprehension Subtest Raw Score (Grade 1)
7. GRC2 - GMRT Comprehension Subtest Raw Score (Grade 2)
8. GRC3 - GMRT Comprehension Subtest Raw Score (Grade 3)
9. GRTI - GMRT Total Test Raw Score (Grade 1)
10. GRT2 - GMRT Total Test Raw Score (Grade 2)
11. GRT3 - GMRT Total Test Raw Score (Grade 3)
12. GTV2 - GMRT Vocabulary Subtest T-score (Grade 2)
13. GTV3 - GMRT Vocabulary Subtest T-score (Grade 3)
14 GTC2 - GMRT Comprehension Subtest T-score (Grade 2)
15. GTC3 - GMRT Comprehension Subtest T-score (Grade 3)
16. GTT2 - GMRT Total Test T-score (Grade 2)
17. GTT3 - GMRT Total Test T-score (Grade 3)
18. CATRP2 - Canadian Achievement Test (CAT) Phonetic Analysis Subtest Raw Score (Grade 2)
19. CATRP3 - CAT Phonetic Analysis Subtest Raw Score (Grade 3)
20. CATRS2 - CAT Structural Analysis Subtest Raw Score (Grade 2)
21. CATRS3 - CAT Structural Analysis Subtest Raw Score (Grade 3)
22. CATRV2 - CAT Reading Vocabulary Subtest Raw Score (Grade 2)
23. CATRV3 - CAT Reading Vocabulary Subtest Raw Score (Grade 3)
24. CATRC2 - CAT Reading Comprehension Subtest Raw Score (Grade 2)
25. CATRC3 - CAT Reading Comprehension Subtest Raw Score (Grade 3)
26. CATRR2 - CAT Total Reading Area Raw Score (Grade 2)
27. CATRR3 - CAT Total Reading Area Raw Score (Grade 3)
28. CATRSP2 - CAT Total Spelling Area Raw Score (Grade 2)
29. CATRSP3 - CAT Total Spelling Area Raw Score (Grade 3)
30. CATRLM2 - CAT Language Mechanics Subtest Raw Score (Grade 2)

31. CATRLM3 - CAT Language Mechanics Subtest Raw Score (Grade 3)
32. CATRLE2 - CAT Language Expression Subtest Raw Score (Grade 2)
33. CATRLE3 - CAT Language Expression Subtest Raw Score (Grade 3)
34. CATRL2 - CAT Total Language Area Raw Score (Grade 2)
35. CATRL3 - CAT Total Language Area Raw Score (Grade 3)
36. CATRCO2 - CAT Computation Subtest Raw Score (Grade 2)
37. CATRCO3 - CAT Computation Subtest Raw Score (Grade 3)
38. CATRA2 - CAT Concepts and Applications Subtest Raw Score (Grade 2)
39. CATRA3 - CAT Concepts and Applications Subtest Raw Score (Grade 3)
40. CATRM2 - CAT Total Math Area Raw Score (Grade 2)
41. CATRM3 - CAT Total Math Area Raw Score (Grade 3)
42. CATRB2 - CAT Total Test Battery Raw Score (Grade 2)
43. CATRB3 - CAT Total Test Battery Raw Score (Grade 3)
44. RD1 - Reading Report Card Mark (Grade 1)
45. WE1 - Written Expression Report Card Mark (Grade 1)
46. PR1 - Printing Report Card Mark (Grade 1)
47. MA1 - Math Report Card Mark (Grade 2)
48. RD2 - Reading Report Card Mark (Grade 2)
49. WE2 - Written Expression Report Card Mark (Grade 2)
50. PR2 - Printing Report Card Mark (Grade 2)
51. MA2 - Math Report Card Mark (Grade 2)
52. RD3 - Reading Report Card Mark (Grade 3)
53. WE3 - Written Expression Report Card Mark (Grade 3)
54. PR3 - Printing Report Card Mark (Grade 3)
55. MA3 - Math Report Card Mark (Grade 3)

Table D.1

Summary of Descriptive Statistics

	Total Group				Boys				Girls			
Variable	N	M	SD	range	N	M	SD	range	N	M	SD	range
CA	149	71.92	4.5	63 - 85	80	72.85	4.34	65 - 85	69	70.84	3.65	63 - 77
CCAT	149	109.50	15.09	50 - 143	80	108.24	16.20	50 - 141	69	111.03	13.67	69 - 143
PDR	149	3.80	.90	2 - 5	80	3.66	.95	2 - 5	69	4.03	.80	2 - 5
CR	149	4.91	.41	2 - 5	80	4.89	.40	2 - 5	69	4.92	.43	2 - 5
PV	149	4.99	.12	4 - 5	80	4.98	.16	4 - 5	69	5.0	.00	5 - 5
VD	149	8.31	1.72	0 - 10	80	8.13	1.71	0 - 10	69	8.52	1.72	0 - 10
VM	149	4.72	.57	3 - 5	80	4.66	.59	3 - 5	69	4.78	.54	3 - 5
DAP	149	7.64	1.23	3 - 10	80	7.56	1.24	5 - 10	69	7.74	1.21	3 - 10
RC	149	4.86	.51	2 - 5	80	4.83	.55	3 - 5	69	4.90	.46	2 - 5
RA	149	3.99	1.59	0 - 5	80	3.64	1.80	0 - 5	69	4.41	1.18	0 - 5
NC	149	9.06	1.46	0 - 10	80	8.96	1.59	0 - 10	69	9.17	1.36	3 - 10
RLL	149	9.89	2.84	0 - 13	80	9.18	2.98	1 - 13	69	10.71	2.44	0 - 13
AD	149	9.01	1.97	0 - 10	80	8.84	2.20	0 - 10	69	9.20	1.66	2 - 10
PPD	149	6.71	2.52	0 - 10	80	5.94	2.12	0 - 10	69	7.61	2.66	0 - 10
NS	149	6.62	1.31	0 - 7	80	6.45	1.57	0 - 7	69	6.81	.90	1 - 7
BKIT	149	84.53	10.50	27 - 99	80	81.72	10.32	27 - 97	69	87.80	9.80	28 - 99
GRV1	122	31.85	7.73	11 - 45	66	30.82	7.78	11 - 45	56	33.07	7.56	18 - 45

Summary of Descriptive Statistics (continued)

Variable	Total Group N	Total Group M	Total Group SD	Total Group range	Boys N	Boys M	Boys SD	Boys range	Girls N	Girls M	Girls SD	Girls range
GRV2	33	34.00	6.62	19 - 43	16	34.44	5.72	26 - 42	17	33.59	7.53	19 - 43
GRV3	57	33.72	5.84	13 - 42	32	34.56	5.21	21 - 42	25	32.64	6.51	13 - 40
GRC1	122	26.83	7.66	6 - 40	66	24.14	7.11	6 - 40	56	30.00	7.11	13 - 39
GRC2	33	29.15	6.04	14 - 41	16	28.13	5.51	14 - 37	17	30.12	6.52	15 - 41
GRC3	57	34.53	5.76	19 - 42	32	34.84	5.27	20 - 42	25	34.12	6.42	19 - 42
GRT1	122	58.68	14.26	20 - 83	66	54.96	13.80	20 - 83	56	63.07	13.64	32 - 83
GRT2	53	64.70	11.94	35 - 84	23	61.13	10.56	40 - 79	30	67.43	12.38	35 - 84
GRT3	57	68.33	10.44	32 - 82	32	69.41	9.02	48 - 82	25	66.96	12.06	32 - 82
GTV2	58	53.29	7.31	39 - 73	31	54.03	7.43	45 - 73	27	52.44	7.21	39 - 67
GTV3	86	52.28	7.06	33 - 72	44	51.64	5.97	38 - 63	42	52.95	8.06	33 - 72
GTC2	58	50.31	8.03	34 - 76	31	49.94	7.62	34 - 66	27	50.74	8.60	36 - 76
GTC3	86	55.11	7.56	38 - 69	44	54.02	6.95	39 - 67	42	56.24	8.08	38 - 69
GTT2	78	52.92	8.33	37 - 74	38	51.61	7.65	40 - 72	40	54.18	8.84	37 - 74
GTT3	86	53.64	6.75	35 - 69	44	52.82	5.80	42 - 63	42	54.50	7.60	35 - 69
CATRP2	57	19.74	4.24	10 - 25	30	19.23	4.4	10 - 25	27	20.30	4.07	10 - 25
CATRP3	33	16.10	3.56	7 - 20	18	16.56	2.98	18 - 62	15	15.53	3.80	7 - 20
CATRS2	57	8.47	2.08	4 - 11	30	8.13	2.3	4 - 11	27	8.85	1.79	4 - 11
CATRS3	33	8.73	1.44	5 - 11	18	8.72	1.67	5 - 11	15	8.73	1.16	7 - 10
CATRV2	57	11.65	2.30	5 - 15	30	11.43	2.18	6 - 15	27	11.89	2.44	5 - 15

Summary of Descriptive Statistics (continued)

Variable	Total Group N	Total Group M	Total Group SD	Total Group range	Boys N	Boys M	Boys SD	Boys range	Girls N	Girls M	Girls SD	Girls range
CATRV3	33	12.67	2.13	6 - 15	18	12.83	2.09	8 - 15	15	12.47	2.32	6 - 15
CATRC2	57	16.75	3.37	5 - 20	30	16.60	3.42	5 - 20	27	16.93	3.36	5 - 20
CATRC3	33	21.67	2.92	14 - 26	18	21.06	2.88	14 - 25	15	22.40	2.90	15 - 26
CATRR2	57	56.61	9.05	25 - 69	30	55.40	8.64	35 - 69	27	57.96	9.46	25 - 68
CATRR3	33	59.15	7.02	42 - 67	18	59.17	6.70	43 - 67	15	59.13	7.62	42 - 66
CATRSP2	57	15.04	2.74	10 - 20	30	15.00	3.20	10 - 20	27	15.07	2.18	11 - 20
CATRSP3	33	14.94	2.19	10 - 20	18	14.50	2.28	10 - 20	15	15.47	2.03	12 - 19
CATRLM2	57	14.39	3.70	5 - 20	30	13.33	3.54	5 - 19	27	15.56	3.58	5 - 20
CATRLM3	33	15.97	2.20	12 - 20	18	15.78	2.21	12 - 20	15	16.20	2.42	12 - 19
CATRLE2	57	17.18	2.70	10 - 20	30	17.20	2.54	11 - 20	27	17.15	2.92	10 - 20
CATRLE3	33	20.42	2.25	13 - 24	18	20.22	1.40	18 - 22	15	20.67	3.02	13 - 24
CATRL2	57	31.56	5.13	16 - 40	30	30.53	4.82	20 - 38	27	32.70	5.32	16 - 40
CATRL3	33	36.39	3.10	30 - 43	18	36.00	2.33	32 - 41	15	36.87	3.87	30 - 43
CATRCO2	57	21.72	4.03	8 - 26	30	21.57	4.22	8 - 26	27	21.89	3.89	10 - 26
CATRCO3	33	27.82	26.65	15 - 39	18	29.33	7.28	18 - 39	15	26.00	5.49	15 - 34
CATRA2	57	30.60	4.90	17 - 38	30	30.50	5.35	17 - 38	27	30.70	4.45	17 - 37
CATRA3	33	32.97	5.91	18 - 43	18	33.61	6.26	23 - 43	15	32.20	5.57	18 - 41
CATRM2	57	52.26	8.31	25 - 63	30	52.07	8.95	25 - 63	27	52.48	7.70	30 - 61
CATRM3	33	60.82	10.83	42 - 79	18	62.94	12.00	44 - 79	15	58.27	8.96	42 - 75

Summary of Descriptive Statistics (continued)

Variable	Total Group N	M	SD	range	Boys N	M	SD	range	Girls N	M	SD	range
CATRB2	57	155.18	20.77	82 - 185	30	152.43	20.12	106 - 185	27	158.22	21.42	82 - 182
CATRB3	33	171.30	17.36	144 - 206	18	172.61	18.31	144 - 206	15	169.73	16.64	148 - 198
RD1	24	4.15	.68	3 - 5	13	4.05	.74	3 - 5	11	4.27	.61	3 - 5
WE1	24	3.74	.72	3 - 5	13	3.67	.79	3 - 5	11	3.82	.64	3 - 5
PR1	24	3.97	.76	3 - 5	13	3.92	.82	3 - 5	11	4.03	.72	3 - 5
MA1	24	4.31	.58	3 - 5	13	4.18	.66	3 - 5	11	4.45	.45	4 - 5
RD2	33	4.00	.77	3 - 5	18	3.94	.83	3 - 5	15	4.07	.72	3 - 5
WE2	33	3.98	.65	3 - 5	18	3.89	.74	3 - 5	15	4.09	.51	3 - 5
PR2	33	3.90	.63	3 - 5	18	3.78	.59	3 - 5	15	4.04	.67	3 - 5
MA2	33	4.34	.67	3 - 5	18	4.33	.69	3 - 5	15	4.35	.67	3 - 5
RD3	30	4.09	.66	3 - 5	12	3.86	.64	3 - 5	18	4.24	.65	3 - 5
WE3	30	3.82	.76	2 - 5	12	3.56	.78	2 - 5	18	4.00	.71	3 - 5
PR3	30	3.86	.77	2 - 5	12	3.61	.81	2 - 5	18	4.02	.72	3 - 5
MA3	30	4.18	.81	2 - 5	12	4.17	1.03	2 - 5	18	4.18	.65	3 - 5

Table D.2

Significant Differences Between Variable Means by Gender

Variable	Boys n	Boys M	Girls n	Girls M	T Value	p level
CA	80	72.85	69	70.84	3.03	.003*
CCAT	80	108.24	69	111.03	-1.13	.262
PDR	80	3.66	69	4.03	-2.51	.013*
CR	80	4.89	69	4.92	-0.039	.697
PV	80	4.98	69	5.00	-1.32	.188
VD	80	8.13	69	8.52	-1.41	.161
VM	80	4.66	69	4.78	-1.28	.201
DAP	80	7.56	69	7.74	-0.88	.382
RC	80	4.83	69	4.90	-0.88	.379
RA	80	3.64	69	4.41	-3.03	.003*
NC	80	8.96	69	9.17	-0.87	.388
RLL	80	9.18	69	10.71	-3.39	.001*
AD	80	8.84	69	9.20	-1.13	.261
PPD	80	5.94	69	7.61	-4.27	<.001*
NS	80	6.45	69	6.81	-1.69	.093
BKIT	80	81.72	69	87.80	-3.67	<.001*
GRV1	66	30.82	56	33.07	-1.62	.109
GRV2	16	34.44	17	33.59	0.36	.719
GRV3	32	34.56	25	32.64	1.24	.220
GRC1	66	24.14	56	30.00	-4.54	<.001*
GRC2	16	28.13	17	30.12	-0.95	.352
GRC3	32	34.84	25	34.12	0.47	.642
GRT1	66	54.95	56	63.07	-3.26	.001*
GRT2	23	61.13	30	67.43	-1.96	.056

Significant Differences Between Variable Means by Gender (continued)

Variable	Boys n	Boys M	Girls n	Girls M	T Value	p level
GRT3	32	69.41	25	66.96	0.88	.385
GTV2	31	54.03	27	52.44	0.82	.414
GTV3	44	51.64	42	52.95	-0.86	.391
GTC2	31	49.94	27	50.74	-0.38	.707
GTC3	44	54.02	42	56.24	-1.37	.176
GTT2	38	51.61	40	54.18	-1.37	.175
GTT3	44	52.82	42	54.50	-1.16	.251
CATRP2	30	19.23	27	20.30	-0.94	.349
CATRP3	18	16.56	15	15.53	0.87	.392
CATRS2	30	8.13	27	8.85	-1.31	.197
CATRS3	18	8.72	15	8.73	-0.02	.983
CATRV2	30	11.43	27	11.89	-0.75	.459
CATRV3	18	12.83	15	12.47	0.49	.630
CATRC2	30	16.60	27	16.93	-0.36	.719
CATRC3	18	21.06	15	22.40	-1.33	.193
CATRR2	30	55.40	27	57.96	-1.07	.290
CATRR3	18	59.17	15	59.13	0.01	.989
CATRSP2	30	15.00	27	15.07	-0.10	.920
CATRSP3	18	14.50	15	15.47	-1.27	.212
CATRLM2	30	13.33	27	15.56	-2.36	.022*
CATRLM3	18	15.78	15	16.20	-0.54	.591
CATRLE2	30	17.20	27	17.15	0.07	.943
CATRLE3	18	20.22	15	20.67	-0.56	.580
CATRL2	30	30.53	27	32.70	-1.62	.112
CATRL3	18	36.00	15	36.87	-0.79	.433
CATRCO2	30	21.57	27	21.89	-0.30	.766

Significant Differences Between Variable Means by Gender (continued)

Variable	Boys n	Boys M	Girls n	Girls M	T Value	p level
CATRCO3	18	29.33	15	26.00	1.46	.155
CATRA2	30	30.50	27	30.70	-0.16	.877
CATRA3	18	33.61	15	32.20	0.68	.503
CATRM2	30	52.07	27	52.48	-0.19	.853
CATRM3	18	62.94	15	58.27	1.25	.222
CATRB2	30	152.43	27	158.22	-1.05	.297
CATRB3	18	172.61	15	169.73	0.47	.643
RD1	13	4.05	11	4.27	-0.79	.439
WE1	13	3.67	11	3.82	-0.51	.618
PR1	13	3.92	11	4.03	-0.34	.739
MA1	13	4.18	11	4.45	-1.16	.257
RD2	18	3.94	15	4.07	-0.45	.657
WE2	18	3.89	15	4.09	-0.88	.387
PR2	18	3.78	15	4.04	-1.22	.232
MA2	18	4.33	15	4.35	-0.09	.932
RD3	12	3.86	18	4.24	-1.58	.125
WE3	12	3.56	18	4.00	-1.61	.119
PR3	12	3.61	18	4.02	-1.44	.160
MA3	12	4.17	18	4.18	-0.06	.954

*Indicates a significant difference at $p \leq .05$.

Table D.3

Pearson Correlation Coefficients: BK1-1 Subtest and Total Test Scores with Canadian Cognitive Abilities Test (CCAT) for Total Group, Boys, and Girls

	CCAT Total Group (N = 149)	CCAT Boys (n = 80)	CCAT Girls (n = 69)
PDR	.092	.087	.055
CR	.311***	.333***	.287**
PV	.233*	.286**	-
VD	.328***	.341***	.297**
VM	.240**	.326**	.093
DAP	.166*	.141	.190
RC	.348***	.403***	.247*
RA	.332***	.334***	.298**
NC	.470***	.475***	.452***
RLL	.467***	.546***	.318**
AD	.359***	.368***	.329**
PPD	.209**	.093	.304**
NS	.293***	.347***	.148
BKIT	.523***	.578***	.439***

Note: " - " printed if a coefficient could not be calculated.

* $p \leq .05$
** $p \leq .01$
*** $p \leq .001$

Table D.4

Pearson Correlation Coefficients: BK1-1 Subtests and Total Test Score with Gates-MacGinitie Reading Test Scores for Total Group

	N	PDR	CR	PV	VD	VM	DAP	RC	RA	NC	FILL	AD	PPD	NS	BKIT
GRV1	122	.163*	.302***	-.014	.136	.260**	.156*	.100	.268***	.403***	.512***	.091	.235**	.362***	.516***
GRC1	122	.114	.269***	.093	.179*	.235**	.169*	.212**	.378***	.360***	.604***	.191*	.326***	.250**	.610***
GRT1	122	.150*	.310***	.043	.170*	.267***	.166*	.166*	.349***	.422***	.602***	.162*	.302***	.331***	.606***
GRV2	33	.214	.233	-	.350*	.213	.132	.405**	.275	.178	.373*	.164	.156	-.060	.400**
GRC2	33	.153	.254	-	.451**	.351*	.100	.393**	.150	.347*	.350*	.403*	.306*	.171	.508***
GRT2	53	.265*	.267*	-	.401***	.260*	.141	.379**	.245*	.330**	.473***	.276*	.308**	.077	.565***
GRV3	57	.066	-.099	-	.080	.130	-.005	.200	-.100	.288*	.160	.216	-.144	.060	.148
GRC3	57	-.024	.018	-	-.018	-.145	.150	.233*	-.202	.246*	.303**	.181	.064	.161	.220*
GRT3	57	.044	-.044	-	.039	-.004	.083	.242*	-.161	.265*	.276*	.200	-.053	.118	.200
GTV2	58	-.015	.188	-	.343**	.163	.197	.217*	.275*	.164	.284**	.003	-.016	-.075	.276*
GTC2	58	.093	.182	-	.348*	.238*	.164	.239*	.205	.242*	.301**	.060	.120	.063	.364**
GTT2	76	.131	.210*	-	.369***	.192*	.210*	.246*	.259**	.310**	.403***	.075	.163	.012	.459***
GTV3	86	.129	-.066	-	.129	.135	.132	.197*	.063	.304**	.335***	.214*	-.010	.064	.322***
GTC3	86	.044	.078	-	.083	.041	.235*	.240*	-.066	.346***	.403***	.163	.069	.165	.347**
GTT3	86	.102	.0001	-	.123	.113	.201*	.241*	.033	.353***	.403***	.210*	.039	.124	.373***

Note: " - " printed if a coefficient could not be calculated.

* p ≤ .05
** p ≤ .01
*** p ≤ .001

Table D.5

Pearson Correlation Coefficients: BK1-1 Subtests and Total Test Score with Gates-MacGinitie Reading Test Scores for Boys

	n	PDR	CH	PV	VD	VM	DAP	RC	RA	NC	RLL	AD	PPD	NS	BKIT
GRV1	66	-.098	.291**	-.035	.003	.223*	.172	.051	.200	.404***	.574***	.010	.270*	.397***	.473***
GRC1	66	-.098	.271*	.090	.064	.271*	.246*	.146	.381***	.436***	.639***	.070	.280**	.195	.552***
GRT1	66	-.106	.303**	.027	.038	.265*	.225*	.108	.309**	.452***	.653***	.042	.300**	.324**	.551***
GRV2	16	-.176	.465*	-	.073	.208	.231	.254	.258	-.172	.271	.435*	.120	-.023	.317
GRC2	16	-.321	.332	-	.133	.119	.008	-.042	.024	.151	.337	.696***	.168	.169	.341
GRT2	23	-.142	.325	-	.068	.100	-.053	.166	.075	.159	.367*	.425*	.163	.016	.304
GRV	32	.041	-.225	-	-.063	.300*	-.174	.305*	-.121	.196	.072	.051	-.265	.092	-.056
GRC3	32	-.163	-.040	-	-.120	.043	.252	.017	-.235	.185	.325*	-.165	-.162	.203	-.0004
GRT3	32	-.072	-.154	-	-.106	.198	.050	.186	-.207	.221	.231	-.061	-.271	.172	-.033
GTV2	31	-.240	.266	-	.128	.175	.236	.104	.310*	.004	.254	-.064	-.158	-.052	.163
GTC2	31	-.134	.208	-	.177	.171	.160	.106	.221	.141	.343*	-.092	-.113	.062	.216
GTT2	36	-.111	.243	-	.174	.167	.126	.119	.223	.173	.349*	-.121	-.093	-.038	.224
GTV3	44	-.069	-.228	-	-.062	.216	-.113	.230	-.041	.206	.236	.070	-.196	.095	.045
GTC3	44	-.325*	.012	-	-.149	.162	.221	.107	-.195	.286*	.360**	-.111	-.113	.165	.067
GTT3	44	-.207	-.124	-	-.122	.235	.057	.191	-.100	.260*	.334*	-.020	-.172	.138	.072

Note: " - " printed if a coefficient could not be calculated.

* p ≤ .05
** p ≤ .01
*** p ≤ .001

Table D.6

Pearson Correlation Coefficients: BK1-1 Subtests and Total Test Scores with Gates-MacGinitie Reading Test Scores for Girls

	n	PDR	CR	PV	VD	VM	DAP	RC	RA	MC	SLL	AD	PPD	KS	SKDT
GRV1	56	.461***	.456***	-	.266*	.265*	.117	.146	.341**	.364**	.361**	.166	.133	.366**	.569***
GRC1	56	.222*	.354**	-	.197	.108	.094	.262*	.213	.264*	.406**	.326**	.169	.331**	.648***
GRT1	56	.371**	.436***	-	.250*	.215	.114	.217	.301**	.366**	.424**	.276*	.177	.344**	.661***
GRV2	17	.564**	-	-	.526*	.253	.142	.465*	.467*	.421*	.506*	-.016	.260	-.117	.671**
GRC2	17	.602**	-	-	.662***	.511*	.060	.589**	.263	.489*	.363	.106	.300	.064	.594**
GRT2	30	.514**	-	-	.683***	[illegible]	.163	.496**	.406*	.436**	.644***	.126	.231	.037	.643***
GRV3	25	.219	.269	-	.247	-.010	.275	.109	-.027	.363*	.491*	.439*	.110	.066	.619**
GRC3	25	.187	.242	-	.068	-.201	.019	.410*	-.157	.292	.334*	.576***	.381*	.266	.649**
GRT3	25	.224	.260	-	.192	-.163	.162	.201	-.067	.331	.406*	.466**	.237	.207	.667**
GTV2	27	.422*	.123	-	.546**	.164	.226	.323*	.496*	.366*	.426*	.160	.212	-.060	.626**
GTC2	27	.386*	.180	-	.503**	.302	.164	.346*	.209	.331*	.260	.221	.279	.064	.612**
GTT2	40	.321*	.150	-	.515***	.169	.256	.334*	.316*	.407**	.440**	.249	.229	.049	.696***
GTV3	42	.292*	.277*	-	.312*	.079	.371**	.174	.211	.374**	.437**	.343*	.046	.136	.665***
GTC3	42	.353**	.296*	-	.294*	-.077	.223	.349**	.036	.403**	.421**	.413**	.146	.260*	.577***
[illegible]	42	.337*	.302*	-	.335*	.023	.339*	.278*	.150	.424**	.466**	.416**	.096	.216	.622***

Note: " - " printed if a coefficient could not be calculated.

* p ≤ .05

** p ≤ .01

*** p ≤ .001

Table D.7

Pearson Correlation Coefficients: BK1-1 Subtest and Total Test Scores with Grade 2 Canadian Achievement Test (CAT) for Total Group (N = 57)

	PDR	CR	PV	VD	VM	DAP	RC	RA	NC	RU	AD	PPD	NS	BKIT
CATRP2	.263*	.112	-	.160	.353**	.160	.197	.153	.474***	.442***	.196	.298**	.103	.507***
CATRS2	.211	.011	-	.123	.193	.153	.131	-.026	.178	.265*	.122	.054	.030	.241*
CATRV2	.164	.171	-	.211	.012	-.075	.112	.012	.222*	.316**	.446***	.095	.087	.290*
CATRC2	.014	-.007	-	.167	.100	-.067	.133	-.165	.126	.047	.375**	.071	.107	.111
CATRR2	.267	.096	-	.226*	.250*	.087	.200	.0000	.366**	.371**	.373**	.203	.117	.408***
CATRSP2	.216	.231*	-	-.125	.160	.016	.105	.113	.219*	.303**	.457***	.187	-.117	.291*
CATRLM2	.112	.027	-	-.100	.028	.073	.054	.141	.066	.150	.222*	.196	.304**	.234*
CATRLE2	.241*	-.005	-	.065	.181	-.088	.097	.198	.142	.117	.144	-.041	-.090	.134
CATRL2	.207	.017	-	-.038	.115	.006	.090	.205	.137	.170	.238*	.120	.172	.239*
CATRCO2	.275*	.092	-	-.053	.397***	.174	.463***	.093	.308**	.235*	.175	.102	.040	.323**
CATRA2	.272*	.073	-	-.091	.318**	.041	.335**	.170	.262*	.326**	.125	-.079	.005	.236**
CATRM2	.267*	.066	-	-.024	.363**	.103	.431***	.145	.305*	.300**	.165	-.0002	.020	.291*
CATTB2	.293*	.072	-	.080	.300**	.075	.292*	.140	.329**	.342**	.336**	.155	.061	.367***

Note: " - " printed if a coefficient could not be calculated.

* p ≤ .05
** p ≤ .01
*** p ≤ .001

Table D.8

Pearson Correlation Coefficients: BK1-1 Subtest and Total Test Scores with Grade 2 Canadian Achievement Test (CAT) for Boys (n = 30)

	PDR	CR	PV	VD	VM	DAP	RC	RA	NC	RLL	AD	PPD	NS	BKIT
CATRP2	.168	.096	-	-.050	.263	.146	.182	.084	.446**	.461**	.137	.099	.044	.360*
CATRS2	.071	-.032	-	-.059	.280	.128	.175	-.116	.090	.160	.065	.053	-.006	.112
CATRV2	-.046	.157	-	-.070	.061	-.096	-.049	-.074	.077	.236	.570***	-.007	.127	.170
CATRC2	-.133	-.003	-	-.047	.116	-.061	-.133	-.300	.165	-.047	.422**	.027	.163	.007
CATRR2	.040	.079	-	-.077	.270	.060	.074	-.125	.336*	.318*	.396*	.074	.117	.269
CATRSP2	.022	.227	-	-.250	.285	-.030	.059	.101	.352*	.173	.528***	.256	-.163	.243
CATRLM2	-.030	-.060	-	-.129	.279	-.021	.071	.016	.242	-.066	.181	-.069	.363*	.085
CATRLE2	.076	-.066	-	-.236	.471**	-.193	.313*	.210	-.099	-.116	.205	-.106	-.105	-.050
CATRL2	.018	-.079	-	-.220	.452**	-.117	.217	.122	.126	-.110	.241	-.107	.211	.036
CATRCO2	.226	.091	-	-.141	.642***	.267	.338*	.178	.355*	.118	.147	-.083	.056	.271
CATRA2	.173	.082	-	-.251	.553***	.154	.300	.172	.330*	.266	.075	-.264	.031	.200
CATRM2	.210	.080	-	-.216	.633***	.216	.339*	.187	.365*	.215	.114	-.197	.045	.247
CATRB2	.137	.026	-	-.191	.527***	.087	.239	.096	.340*	.189	.344*	-.028	.080	.253

Note: " - " printed if a coefficient could not be calculated.

* $p \leq .05$
** $p \leq .01$
*** $p \leq .001$

Table D.9

Pearson Correlation Coefficients: BK1-1 Subtest and Total Test Scores with Grade 2 Canadian Achievement Test (CAT) for Girls (n = 27)

	PDR	CR	PV	VD	VM	DAP	RC	RA	NC	RLL	AD	PPD	NS	BKIT
CATRP2	.365	.163	-	.423*	.467**	.118	.266	.225	.525**	.394*	.319	.446**	.266	.649***
CATRS2	.390*	.057	-	.346*	.110	.091	.180	.057	.310	.438**	.143	-.149	-.023	.314
CATRV2	.368*	.341*	-	.461**	-.026	-.119	.211	.092	.351*	.368*	.290	.099	-.067	.362*
CATRC2	.197	-.092	-	.388*	.088	-.119	.291	-.036	.092	.150	.371*	.079	-.087	.208
CATRR2	.395*	.136	-	.504**	.246	-.005	.306	.118	.407*	.406*	.371*	.217	.057	.511**
CATRSP2	.423*	.406*	-	.055	-.007	.112	.176	.140	.234	.549**	.297	.171	.176	.429*
CATRLM2	.197	.125	-	-.127	-.195	.030	.118	.216	-.035	.334*	.118	.157	.017	.197
CATRLE2	.452**	.262	-	.341*	-.067	.049	.014	.210	.340*	.373*	.090	.010	-.111	.347
CATRL2	.360*	.228	-	.102	-.168	.047	.087	.260	.163	.429*	.128	.110	.049	.323*
CATRCO2	.342*	.140	-	.170	.142	.004	.646***	-.070	.267	.365*	.293	.268	-.055	.409*
CATRA2	.433*	.175	-	.103	.051	-.165	.430*	.165	.195	.415*	.327*	.060	-.140	.309
CATRM2	.408*	.168	-	.205	.108	-.118	.571***	.061	.248	.422*	.329*	.170	-.113	.360*
CATRB2	.459**	.218	-	.327*	.105	-.021	.380*	.153	.333*	.493**	.344*	.202	-.010	.486**

Note: " - " printed if a coefficient could not be calculated.

* p ≤ .05
** p ≤ .01
*** p ≤ .001

Table D.10

Pearson Correlation Coefficients: BK1-1 Subtest and Total Test Scores with Grade 3 Canadian Achievement Test (CAT) for Total Group (N = 33)

	PDR	CR	PV	VD	VM	DAP	RC	RA	NC	RLL	AD	PPD	NS	BKIT
CATRP3	.323*	.085	-	.221	.459**	.167	.005	.036	.737***	.281	.032	.230	.227	.466**
CATRS3	.066	.241	-	-.235	.251	.115	.091	-.056	.355*	.219	.355*	.229	.129	.278
CATRV3	.213	.008	-	.031	.390**	.076	.140	.037	.558***	.202	.091	-.019	.200	.273
CATRC3	.286*	.104	-	-.112	.114	-.225	.410**	-.121	.460**	.324*	.365**	.323*	.059	.320*
CATRR3	.552*	.160	-	.020	.437**	.034	.234	-.034	.787***	.376*	.276	.266	.221	.496**
CATRSP3	.275	.007	-	.307*	.071	.114	-.087	.217	.380*	.263	.339*	.336*	-.096	.437**
CATRLM3	.165	-.057	-	-.242	.361*	.017	-.166	.344*	.036	.046	.198	-.014	-.055	.107
CATRLE3	.250	.123	-	.311*	.288*	.208	-.126	.209	.571***	.185	-.181	.279	.225	.433**
CATRL3	.312*	.049	-	.054	.465**	.163	-.209	.396**	.440**	.167	.009	.193	.124	.390**
CATRCO3	.304*	.175	-	-.282	.043	-.067	.130	.208	.096	.206	.243	-.242	-.123	.061
CATRA3	.134	.308*	-	-.120	.272	.070	.455**	.164	.499**	.574***	.128	-.058	.245	.397**
CATRM3	.257	.273	-	-.237	.176	.0002	.312*	.220	.326*	.432**	.218	-.183	.056	.250
CAT[illegible]3	.394**	.245	-	-.091	.376*	.057	.241	.222	.638***	.465**	.292*	.079	.134	.462**

Note: " - " printed if a coefficient could not be calculated.

* $p \leq .05$
** $p \leq .01$
*** $p \leq .001$

Table D.11

Pearson Correlation Coefficients: BK1-1 Subtest and Total Test Scores with Report Card Marks by Grade for Total Group

	N	PDR	CR	PV	VD	VM	DAP	RC	RA	NC	RLL	AD	PPD	N3	BKIT
RD1	24	.103	.119	-	.343*	.230	-.088	.113	.191	.165	.532**	.361*	.167	-.046	.400*
WE1	24	-.054	.001	-	.122	.259	-.146	.163	.032	.351*	.477**	.260	.172	-.044	.307
PR1	24	-.175	.019	-	-.014	.181	-.173	.207	.231	.191	.264	.256	.064	-.102	.161
MA1	24	.242	.322	-	.186	.267	.152	.097	.125	.365*	.636**	.336	.310	.012	.537**
O2	33	.409**	-.036	-	.213	-.107	.104	-.0001	.0003	.304*	.237	.153	.080	.076	.284
WE2	33	.257	-.023	-	.132	-.242	.284	-.066	.063	.064	.268	.301*	.168	.241	.327*
PR2	33	.118	-.131	-	.161	-.277	.252	-.313*	.107	-.247	-.070	-.111	.160	.065	.060
MA2	33	.250	.091	-	.254	-.241	.255	.003	.110	.154	.252	.067	.019	.266	.260
RD3	30	.160	.242	-	.056	.064	.202	.311	.114	.347*	.401**	.064	.376*	-	.621**
WE3	30	.329*	.115	-	.160	-.006	-.007	.204	.001	.220	.561***	.028	.420**	-	.630***
PR3	30	.154	.302	-	.264	.055	-.009	.464**	.219	.324*	.321*	.165	.415*	-	.519**
MA3	30	.160	.060	-	.107	-.036	-.070	.276	.077	.407*	.527***	.083	.190	-	.427**

Note: " - " printed if a coefficient could not be calculated.

* p ≤ .05
** p ≤ .01
*** p ≤ .001

Table D.12

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade Two Gates-MacGinitie Reading Test (GMRT) Total Test T-Scores for Boys (n = 38)

a)

BKIT1		GMRT: Difficulty (≤ 45)	GMRT: No Difficulty (≥ 46)	Row Total
	At risk (≤ 71.5)	1	2	3
	Not at risk (≥ 72)	7	28	35
	Column Total	8	30	38

Overall Hit Rate: 29/38 = 76% False Positive Rate: 2/38 = 5% False Negative Rate: 7/38 = 18%
Specificity Rate: 28/30 = 93% Sensitivity Rate: 1/8 = 13%

b)

BKIT2		GMRT: Difficulty (≤ 45)	GMRT: No Difficulty (≥ 46)	Row Total
	At risk (≤ 76.5)	2	3	5
	Not at risk (≥ 77)	6	27	33
	Column Total	8	30	38

Overall Hit Rate: 29/38 = 76% False Positive Rate: 3/38 = 8% False Negative Rate: 6/38 = 16%
Specificity Rate: 27/30 = 90% Sensitivity Rate: 2/8 = 25%

c)

BKIT3		GMRT: Difficulty (≤ 45)	GMRT: No Difficulty (≥46)	Row Total
	At risk (≤ 82.5)	4	9	13
	Not at risk (≥ 83)	4	21	25
	Column Total	8	30	38

Overall Hit Rate: 25/38 = 66% False Positive Rate: 9/38 = 24% False Negative Rate: 4/38 = 11%
Specificity Rate: 21/30 = 70% Sensitivity Rate: 4/8 = 50%

Note: All percentages rounded to the nearest whole number.

Table D.13

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade Two Gates-MacGinitie Reading Test (GMRT) Total Test T-Scores for Girls (n = 40)

a)

BKIT1	GMRT: Difficulty (≤ 45)	GMRT: No Difficulty (≥ 46)	Row Total
At risk (≤ 78.5)	2	1	3
Not at risk (≥ 79)	4	33	37
Column Total	6	34	40

Overall Hit Rate: 35/40 = 88% False Positive Rate: 1/40 = 3% False Negative Rate: 4/40 = 10%
Specificity Rate: 33/34 = 97% Sensitivity Rate: 2/6 = 33%

b)

BKIT2	GMRT: Difficulty (≤ 45)	GMRT: No Difficulty (≥ 46)	Row Total
At risk (≤ 82.5)	4	2	6
Not at risk (≥ 83)	2	32	34
Column Total	6	34	40

Overall Hit Rate: 36/40 = 90% False Positive Rate: 2/40 = 5% False Negative Rate: 2/40 = 5%
Specificity Rate: 32/34 = 94% Sensitivity Rate: 4/6 = 67%

c)

BKIT3	GMRT: Difficulty (≤ 45)	GMRT: No Difficulty (≥46)	Row Total
At risk (≤ 87.5)	5	7	12
Not at risk (≥ 88)	1	27	28
Column Total	6	34	40

Overall Hit Rate: 32/40 = 80% False Positive Rate: 7/40 = 18% False Negative Rate: 1/6 = 3%
Specificity Rate: 27/34 = 79% Sensitivity Rate: 5/6 = 83%

Note: All percentages rounded to the nearest whole number.

Table D.14

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade Two Gates-MacGinitie Reading Test (GMRT) Total Test Raw Scores for the Total Group (N = 53)

a)

		GMRT		
		Difficulty (≤ 53)	No Difficulty (≥ 54)	Row Total
BKIT1	At risk (≤ 74.5)	2	3	5
	Not at risk (≥ 75)	8	40	48
	Column Total	10	43	53

Overall Hit Rate: 42/53 = 79% False Positive Rate: 3/53 = 6% False Negative Rate: 8/53 = 15%
Specificity Rate: 40/43 = 93% Sensitivity Rate: 2/10 = 20%

b)

		GMRT		
		Difficulty (≤ 53)	No Difficulty (≥ 54)	Row Total
BKIT2	At risk (≤ 79.5)	5	3	8
	Not at risk (≥ 80)	5	40	45
	Column Total	10	43	53

Overall Hit Rate: 45/53 = 85% False Positive Rate: 3/53 = 6% False Negative Rate: 5/53 = 9%
Specificity Rate: 40/43 = 93% Sensitivity Rate: 5/10 = 50%

c)

		GMRT		
		Difficulty (≤ 53)	No Difficulty (≥54)	Row Total
BKIT3	At risk (≤ 84.5)	7	9	16
	Not at risk (≥ 85)	3	34	37
	Column Total	10	43	53

Overall Hit Rate: 41/53 = 77% False Positive Rate: 9/53 = 17% False Negative Rate: 3/53 = 6%
Specificity Rate: 34/43 = 79% Sensitivity Rate: 7/10 = 70%

Note: All percentages rounded to the nearest whole number.

Table D.15

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade Three Gates-MacGinitie Reading Test (GMRT) Total Test T-Scores for Boys (n = 44)

a)

BKIT1		GMRT Difficulty (≤ 47)	GMRT No Difficulty (≥ 48)	Row Total
	At risk (≤ 71.5)	0	3	3
	Not at risk (≥ 72)	7	34	41
	Column Total	7	37	44

Overall Hit Rate: 34/44 = 77% False Positive Rate: 3/44 = 7% False Negative Rate: 7/44 = 16%
Specificity Rate: 34/37 = 92% Sensitivity Rate: 0/7 = 0%

b)

BKIT2		GMRT Difficulty (≤ 47)	GMRT No Difficulty (≥ 48)	Row Total
	At risk (≤ 76.5)	1	3	4
	Not at risk (≥ 77)	6	34	40
	Column Total	7	37	44

Overall Hit Rate: 35/44 = 80% False Positive Rate: 3/44 = 7% False Negative Rate: 6/44 = 14%
Specificity Rate: 34/37 = 92% Sensitivity Rate: 1/7 = 14%

c)

BKIT3		GMRT Difficulty (≤ 47)	GMRT No Difficulty (≥48)	Row Total
	At risk (≤ 82.5)	1	14	15
	Not at risk (≥ 83)	6	23	29
	Column Total	7	37	44

Overall Hit Rate: 24/44 = 55% False Positive Rate: 14/44 = 32% False Negative Rate: 6/44 = 14%
Specificity Rate: 23/37 = 62% Sensitivity Rate: 1/7 = 14%

Note: All percentages rounded to the nearest whole number.

Table D.16

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade Three Gates-MacGinitie Reading Test (GMRT) Total Test T-Scores for Girls (n = 42)

a)

		GMRT		
		Difficulty (≤ 47)	No Difficulty (≥ 48)	Row Total
BKIT1	At risk (≤ 78.5)	1	0	1
	Not at risk (≥ 79)	5	36	41
	Column Total	6	36	42

Overall Hit Rate: 37/42 = 88% False Positive Rate: 0/42 = 0% False Negative Rate: 5/42 = 12%
Specificity Rate: 36/36 = 100% Sensitivity Rate: 1/6 = 17%

b)

		GMRT		
		Difficulty (≤ 47)	No Difficulty (≥ 48)	Row Total
BKIT2	At risk (≤ 82.5)	3	4	7
	Not at risk (≥ 83)	3	32	35
	Column Total	6	36	42

Overall Hit Rate: 35/42 = 83% False Positive Rate: 4/42 = 10% False Negative Rate: 3/42 = 7%
Specificity Rate: 32/36 = 89% Sensitivity Rate: 3/6 = 50%

c)

		GMRT		
		Difficulty (≤ 47)	No Difficulty (≥48)	Row Total
BKIT3	At risk (≤ 87.5)	5	7	12
	Not at risk (≥ 88)	1	29	30
	Column Total	6	36	42

Overall Hit Rate: 34/42 = 81% False Positive Rate: 7/42 = 17% False Negative Rate: 1/42 = 2%
Specificity Rate: 29/36 = 81% Sensitivity Rate: 5/6 = 83%

Note: All percentages rounded to the nearest whole number.

Table D.17

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade Three Gates-MacGinitie Reading Test (GMRT) Total Test Raw Scores for the Total Group (N = 57)

a)

		GMRT		
		Difficulty (≤ 58)	No Difficulty (≥ 59)	Row Total
BKIT1	At risk (< 74.5)	1	3	4
	Not at risk (> 75)	9	44	53
	Column Total	10	47	57

Overall Hit Rate: 45/57 = 79% False Positive Rate: 3/57 = 5% False Negative Rate: 9/57 = 16%
Specificity Rate: 44/47 = 94% Sensitivity Rate: 1/10 = 10%

b)

		GMRT		
		Difficulty (≤ 58)	No Difficulty (≥ 59)	Row Total
BKIT2	At risk (< 79.5)	2	10	12
	Not at risk (> 80)	8	37	45
	Column Total	10	47	57

Overall Hit Rate: 39/57 = 68% False Positive Rate: 10/57 = 18% False Negative Rate: 8/57 = 14%
Specificity Rate: 37/47 = 79% Sensitivity Rate: 2/10 = 20%

c)

		GMRT		
		Difficulty (≤ 58)	No Difficulty (≥ 59)	Row Total
BKIT3	At risk (< 84.5)	8	20	28
	Not at risk (> 85)	2	27	29
	Column Total	10	47	57

Overall Hit Rate: 35/57 = 61% False Positive Rate: 20/57 = 35% False Negative Rate: 2/57 = 4%
Specificity Rate: 27/47 = 57% Sensitivity Rate: 8/10 = 80%

Note: All percentages rounded to the nearest whole number.

Table D.18

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade Three Canadian Achievement Test (CAT) Total Reading Area Raw Scores for the Total Group (N = 33)

a)

		CAT		
		Difficulty (≤ 52)	No Difficulty (≥ 53)	Row Total
BKIT1	At risk (< 74.5)	1	2	3
	Not at risk (> 75)	4	26	30
	Column Total	5	28	33

Overall Hit Rate: 27/33 = 82% False Positive Rate: 2/33 = 6% False Negative Rate: 4/33 = 12%
Specificity Rate: 26/28 = 93% Sensitivity Rate: 1/5 = 20%

b)

		CAT		
		Difficulty (≤ 52)	No Difficulty (≥ 53)	Row Total
BKIT2	At risk (< 79.5)	1	7	8
	Not at risk (> 80)	4	21	25
	Column Total	5	28	33

Overall Hit Rate: 22/33 = 67% False Positive Rate: 7/33 = 21% False Negative Rate: 4/33 = 12%
Specificity Rate: 21/28 = 75% Sensitivity Rate: 1/5 = 20%

c)

		CAT		
		Difficulty (≤ 52)	No Difficulty (≥ 53)	Row Total
BKIT3	At risk (< 84.5)	4	12	16
	Not at risk (> 85)	1	16	17
	Column Total	5	28	33

Overall Hit Rate: 20/33 = 61% False Positive Rate: 12/33 = 36% False Negative Rate: 1/33 = 3%
Specificity Rate: 16/28 = 57% Sensitivity Rate: 4/5 = 80%

Note: All percentages rounded to the nearest whole number.

Table D.19

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade Three Canadian Achievement Test (CAT) Total Math Area Raw Scores for the Total Group (N = 33)

a)

		CAT		
		Difficulty (≤ 50)	No Difficulty (≥ 51)	Row Total
BKIT1	At risk (< 74.5)	2	1	3
	Not at risk (> 75)	6	24	30
	Column Total	8	25	33

Overall Hit Rate: 26/33 = 87% False Positive Rate: 1/33 = 3% False Negative Rate: 6/33 = 18%
Specificity Rate: 24/25 = 96% Sensitivity Rate: 2/8 = 25%

b)

		CAT		
		Difficulty (≤ 50)	No Difficulty (≥ 51)	Row Total
BKIT2	At risk (< 79.5)	2	6	8
	Not at risk (> 80)	6	19	25
	Column Total	8	25	33

Overall Hit Rate: 21/33 = 64% False Positive Rate: 6/33 = 18% False Negative Rate: 6/33 = 18%
Specificity Rate: 19/25 = 76% Sensitivity Rate: 2/8 = 25%

c)

		CAT		
		Difficulty (≤ 50)	No Difficulty (≥ 51)	Row Total
BKIT3	At risk (< 84.5)	3	13	16
	Not at risk (> 85)	5	12	17
	Column Total	8	25	33

Overall Hit Rate: 15/33 = 45% False Positive Rate: 13/33 = 39% False Negative Rate: 5/33 = 15%
Specificity Rate: 12/25 = 48% Sensitivity Rate: 3/8 = 38%

Note: All percentages rounded to the nearest whole number.

Table D.20

Classificational Analysis: BK1-1 Total Test Score (BKIT) with Grade Three Canadian Achievement Test (CAT) Total Battery Raw Scores for the Total Group (N = 33)

a)

		CAT		
		Difficulty (≤ 154)	No Difficulty (≥ 155)	Row Total
BKIT1	At risk (< 74.5)	1	2	3
	Not at risk (> 75)	4	26	30
	Column Total	5	28	33

Overall Hit Rate: 27/33 = 82% False Positive Rate: 2/33 = 6% False Negative Rate: 4/33 = 12%
Specificity Rate: 26/28 = 93% Sensitivity Rate: 1/5 = 20%

b)

		CAT		
		Difficulty (≤ 154)	No Difficulty (≥ 155)	Row Total
BKIT2	At risk (< 79.5)	1	7	8
	Not at risk (> 80)	4	21	25
	Column Total	5	28	33

Overall Hit Rate: 22/33 = 67% False Positive Rate: 7/33 = 21% False Negative Rate: 4/33 = 12%
Specificity Rate: 21/28 = 75% Sensitivity Rate: 1/5 = 20%

c)

		CAT		
		Difficulty (≤ 154)	No Difficulty (≥ 155)	Row Total
BKIT3	At risk (< 84.5)	3	13	16
	Not at risk (> 85)	2	15	17
	Column Total	5	28	33

Overall Hit Rate: 18/33 = 55% False Positive Rate: 13/33 = 39% False Negative Rate: 2/33 = 6%
Specificity Rate: 15/28 = 54% Sensitivity Rate: 3/5 = 60%

Note: All percentages rounded to the nearest whole number.

Table D.21

Intercorrelations of BK1-1 Subtest and Total Test Scores (N = 149)

	PDR	CR	PV	VD	VM	DAP	RC	RA	NC	RLL	AD	PPD	NS	BKIT
PDR	-	.202**	.173*	.281***	.288***	.129	.214**	.296***	.280***	.287***	.126	.246***	.191**	.452***
CR		-	.044	.351***	.375***	.348***	.227**	.360***	.423***	.463***	.213**	.253***	.396***	.576***
PV			-	.225**	.147*	-.082	.198**	.147*	.320***	.191**	.268***	.196**	.279***	.326***
VD				-	.310***	.280***	.337***	.273***	.503***	.385***	.261***	.220**	.419***	.640***
VM					-	.213**	.306***	.386***	.427***	.296***	.146*	.174*	.298***	.492***
DAP						-	-.005	.155*	.164*	.287***	.079	.166*	.273***	.411***
RC							-	.351***	.406***	.351***	.373***	.137*	.254***	.489***
RA								-	.338***	.548***	.156*	.248***	.194**	.[illegible]02***
NC									-	.584***	.381***	.306***	.454***	.742***
RLL										-	.270***	.367***	.337***	.791***
AD											-	.270***	.255***	.534***
PPD												-	.313***	.614***
NS													-	.598***
BKIT														-

* p ≤ .05
** p ≤ .01
*** p ≤ .001