UNIVERSITY OF ALBERTA

DISCRIMINANT VALIDITY OF THE STANFORD-BINET INTELLIGENCE SCALE: FOURTH EDITION

BY

CHRISTOPHER WILSON ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY.

IN

SCHOOL PSYCHOLOGY

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

SPRING 1995

ISBN 0-612-01778-8

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Christopher Wilson

TITLE OF THESIS: Discriminant Validity of the Stanford-Binet Intelligence Scale: Fourth Edition

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: 1995

Christopher Wilson

40 Shaneragh Road
Mandara
Harare
ZIMBABWE

Date: December 6, 1994

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 DISCRIMINANT VALIDITY OF THE STANFORD-BINET INTELLIGENCE SCALE: FOURTH EDITION submitted by CHRISTOPHER WILSON in partial fulfillment of the requirements for the degree DOCTOR OF PHILOSOPHY in SCHOOL PSYCHOLOGY.

H. L. Janzen (Supervisor)

J. G. Paterson

W. T. Rogers

L. L. Stewin

R. H. Short

K. L. Ward

G. W. Hynd

DATE: December 5, 1994

ABSTRACT

Performance on the Stanford-Binet Intelligence Scale: Fourth Edition (SB:FE) was examined in a Canadian clinic sample (N=1220), age 2 through 23 years, with a range of demographic characteristics and ability levels. Data were analyzed for the samples 2-6-11, 7-11-11, and 12-23-11. SB:FE subtest, Reasoning Area, and Composite Standard Age Scores (SAS's) decreased significantly ($p < .05$) with increasing age. Within each age group, the intercorrelations among subtests, the four Reasoning Area, and the Composite SAS's supported the four cognitive ability areas posited by Thorndike et al. (1986b). Performance of subjects on the SB:FE full battery and SB:FE General Purpose Abbreviated Battery (GPAB) were compared. Significant differences ($p < .05$), attributable to the large sample sizes, were found between means and variances in Reasoning Area and Composite SAS's. Uncorrected correlation coefficients among the two measures were significant ($p < .01$) and close to unity for the Verbal, Quantitative, Short Term Memory, and Composite SAS's. The correlations between Abstract/Visual SAS's, while significant, were somewhat lower. Also, similar and significant ($p < .05$) correlations were observed among the two versions of the SB:FE and the Wide Range Achievement Test-Revised (WRAT-R). Next, internally valid, reliable, and replicable groups displaying differences in profile elevation and/or shape were obtained through application of hierarchical agglomerative and iterative partitioning clustering procedures to SB:FE GPAB data. For the age sample 2-4-11, a two cluster solution, with high average and average groups was optimal. For the samples 5-6-11, 7-11-11, and 12-23-11, a three cluster solution comprising high, average, and low scoring groups was optimal. Mean WRAT-R subtest scores of the groups in all ages samples were significantly different ($p < .01$). However, when cluster solutions were compared with clinically derived _a priori_ learning disability models, clusters were more similar with respect to Composite SAS's or profile elevation, than educational diagnosis. In general, results suggest the SB:FE is most appropriately used as an index of global ability. Caution is needed interpreting Reasoning Area SAS's, although the GPAB may provide a reasonable representation of the full battery.

ACKNOWLEDGEMENTS

It is difficult to adequately express my gratitude to the many people who have contributed, not only this degree, but also to my growth and development over the years since I arrived at the University of Alberta. However, I will try.

First, I would like to express my sincere appreciation and gratitude to my committee: To Dr. Paterson, who, although becoming involved only in the later stages of the dissertation, so sagely contributed to the final examination; To Dr. L. Stewin, who has been involved in my academic progress from the day I arrived in Canada. Thank you for your support, encouragement, and belief in me over the years; To Dr. R. Short, who has also been involved in my academic program since I arrived in Alberta. Many thanks for your assistance over the years; To Dr. K. Ward, thank you for your involvement, insight, and constructive feedback; To Dr. G. W. Hynd many thanks for your invaluable assistance; To Dr. H. Janzen, my supervisor, thank you for your unwavering support and encouragement over the years; and finally sincere thanks to Dr. T. Rogers for your tremendous commitment and support in ensuring this document reached completion.

On a more personal note, thank you to the many people who have supported and encouraged me over the years while I have been in graduate school: In particular, thanks go to Dr. Rod Beattie, whose assistance and support, not only to myself, but to many graduate students over the years, has been immeasurable. To my mother Cicely, and brother David, in Zimbabwe, for their support and encouragement. Thank you. To my Canadian family, Hans, Christine, George, and Chris. Thank you for helping make Canada my home.

Last, and most important of all, I would like to thank Ellen, my wife, for her patience, understanding, support, and love throughout.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

In this chapter, the Stanford-Binet Fourth Edition (SB:FE) (Thorndike, Hagen, & Sattler, 1986a) is briefly introduced. The roles and purposes of educational assessment are discussed, and the need to establish the validity of the SB:FE in educational assessment is identified. In discussing educational assessment, particular emphasis was placed on investigations of learning disabilities and subtyping or classification research. Multivariate research with the SB:FE poses particular problems because of the way it was designed. Thus, as described in this chapter, a major purpose of the study was to explore the utility of the Stanford-Binet: Fourth Edition General Purpose Abbreviated Battery (SB:FE GPAB) as a means of helping to overcome these difficulties. Lastly, this chapter concludes with a presentation of the major research objectives and an outline of the organization of the dissertation.

The Stanford-Binet Intelligence Scale: Fourth Edition (SB:FE)

The SB:FE is "a major intelligence test" that provides a "continuous scale for appraising cognitive development from age 2 to adulthood" (Thorndike et al., 1986a, p. 8). A three level hierarchical model guided the construction of the SB:FE (Sattler, 1988, 1992). This model postulates a general intelligence factor, g, at the highest level of interpretation, and Crystallized, Fluid, and Short Term Memory factors at the second level. As shown in Figure 1, the three factors included at the third level are "nested" within the factors at the second level as follows: Verbal Reasoning and Quantitative Reasoning reflect Crystallized Abilities; Abstract/Visual Reasoning reflects Fluid Ability; and Short Term Memory which stands independently and does not subsume other factors. Each of Verbal Reasoning, Quantitative Reasoning, Abstract/Visual Reasoning, and Short Term Memory are then measured by specific subtests unique to each. Four subtests comprise the Verbal Reasoning area - - Vocabulary, Comprehension, Absurdities, and Verbal Relations; three subtests the Quantitative area - - Quantitative, Number Series, and Equation Building; four subtests the Abstract/Visual area - - Pattern Analysis, Copying, Matrices, and Paper Folding and Cutting; and four the Short Term Memory area - - Bead Memory, Memory for Sentences, Memory for Digits, and Memory for Objects. Up to 13 subtests of the SB:FE may be administered, depending on subject age, ability, and examiner choice (Keith, Cool, Novak, White, & Pottebaum, 1988a; Sattler, 1988). Despite differences in the number of subtests administered at the different age and ability levels, the same grouping of the subtests into ability domains is assumed (Molfese, Yaple, Helwig, Harris, & Connell, 1992).

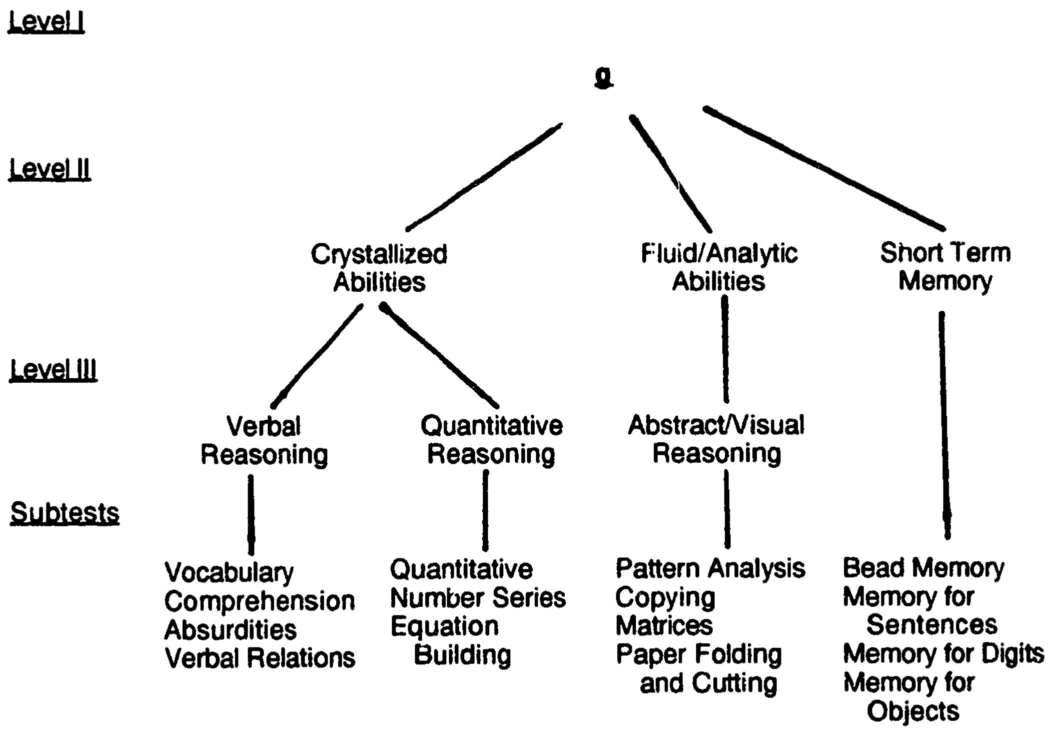

Figure 1. Theoretical model of intelligence underlying the SB:FE.

Thorndike et al. (1986a) noted that the hierarchical model of the SB:FE was adopted largely because of the way clinicians and educators have used the previous editions of the Binet -- "together with other information to make recommendations for educational intervention" (p. 9). These uses have included utilization of the Binet "to identify gifted students, to assess the cognitive abilities of mainstream students who were having difficulty learning, and to identify the mentally retarded" (Thorndike et al., 1986a, p. 9). These potential uses are also applicable to the revised SB:FE, given that the revision was intended "to assess the kinds of cognitive abilities years of research have shown are correlated with school progress" (Thorndike et al., 1986a, p. 9). The 15 subtests comprising the SB:FE, and the organization of these subtests into the four reasoning areas, make it possible to interpret profile elevations and depressions (Glutting, 1989). Thus, in light of the rationale underlying the construction of the SB:FE and the range of cognitive abilities tapped by the test, the instrument appears to hold promise as a means of providing diagnostic and remedial information in educational settings. Boyle (1989) suggested "the new instrument may well usher in an exciting era for cognitive measurement" with research and applied findings "pertaining to clinical, clinical neuropsychological, vocational and educational domains respectively" (p. 709).

There are abbreviated versions of the SB:FE version available that provide "a reasonably accurate estimate of overall cognitive level and pattern of cognitive abilities" (Thorndike et al., 1986a, p. 35). The four test Quick Screening Battery comprises four subtests administered at all age levels: Vocabulary, Bead Memory, Quantitative, and Pattern Analysis. The six test SB:FE General Purpose Abbreviated Battery (GPAB) includes all six subtests that are administered at all ages: Vocabulary, Bead Memory, Quantitative, Memory For Sentences, Pattern Analysis, and Comprehension. Both abbreviated versions require substantially less testing time than the complete battery and have acceptable internal consistency reliabilities (Thorndike et al., 1986b). For example, the SB:FE GPAB can be administered in about 60 minutes (Carvajal, McVey, Sellers, Weyand, & McKnab, 1987), whereas Sattler (1982) reported that the full battery is much too long to complete in most circumstances. The SB:FE GPAB can be used for placement decisions (Glutting, 1989; Thorndike et al., 1986a).

The Role of Assessment in Educational Diagnosis and Planning

There is an increasing emphasis in education on the use of individual educational plans (IEP's), especially for those students experiencing learning difficulties (Salvia & Ysseldyke, 1988; Sattler, 1988). The use of individualized intelligence tests is integral to the diagnosis of learning difficulties and the formulation of individual educational plans. Indeed, Mueller, Dennis, and Short (1986) suggest that the popularity of the Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974), for example, arises from its apparent attractiveness "for psychoeducational diagnostic purposes" (p. 22). Diagnosis provides one of the first steps in planning educational programs. Again, with reference to WISC scales, Kavale and Forness (1984) observed "the structure of the WISC leads to the assumption that there ought to be subtest patterns; in addition, established clinical practice operates as if such patterns were fact" (p. 150). Tests that merely yield single IQ scores are not useful for diagnosis and educational planning. Whilst the previous version of the Binet, the Stanford-Binet Form L-M, was one such test, in contrast, the revised structural design of the SB:FE, with its four cognitive ability or reasoning areas, represents a considerable departure from earlier editions (Keith et al., 1988a), and is "one that is better suited to educational planning" (Fritzke, 1988, p. 50). However, the validity of the SB:FE in educational and differential diagnosis needs to be established.

In the Standards for Educational and Psychological Testing (American Psychological Association, 1985) it is stated:

> Validity is the most important consideration in test evaluation. The concept refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. Test validation is the process of accumulating evidence to support such inferences. A variety of inferences may be made from scores produced by a given test, and therefore there are many ways of accumulating evidence to support any particular inference. Validity, however, is a unitary concept. Although evidence may be accumulated in many ways, validity always refers to the degree to which the evidence supports the inferences that are made from the scores. The inferences regarding specific uses of a test are validated, not the test itself. (APA, 1985, p. 9)

Standards 1.1. and 1.3 (APA, 1985) are particularly relevant for this dissertation. Standard 1.1 states: "Evidence of validity should be presented for the major types of inferences for which use of the test is recommended" (APA, 1985, p. 13). Standard 1.3 states: "Whenever interpretation of subscores, score differences, or profiles is suggested, the evidence justifying such interpretation should be made explicit" (APA, 1985, p. 13). The need to establish the validity of the SB:FE within these parameters in the context of educational diagnosis is paramount.

The Guidelines for Educational and Psychological Testing (Canadian Psychological Association) (CPA, 1987) were formulated to be generally consistent with the APA standards. However, in constructing the guidelines, allowances were made for differing legal and social facets. The guidelines are grounded within the Canadian context (CPA, 1987) and provide additional support for the need to establish evidence of the validity of the SB:FE in educational diagnosis and decision making, particularly within the context of Canadian samples. In the Principles for Fair Student Assessment Practices for Education in Canada (1993) further support for research endeavours of this nature can be found. The second part of this code applies to standardized assessment measures used in student admissions, placement, certification, and educational diagnosis. "Users should select methods that are appropriate for the intended purposes and suitable for the students to be assessed" (Principles for Fair Student Assessment Practices for Education in Canada, 1993, p. 15). One of the purposes of this dissertation is to explore the suitability and appropriateness of the SB:FE in educational assessment and diagnosis. These are the first steps in planning remediation interventions.

Perhaps one of the most important and contentious areas within education today is that of learning disabilities (Hooper & Willis, 1989). Individuals with learning disabilities experience extreme difficulties in making academic progress, despite at least average intelligence and conventional interventions (Childs & Finucci, 1983; Hooper & Willis, 1989; Hynd, 1988; Rourke, 1991; Wilson, 1985; Winzer, 1993). The consequences of learning disabilities are immeasurable, impacting negatively on the educational, emotional, and behavioural well-being of the student (Rourke, 1991). Moreover, the difficulties experienced by individuals with learning disabilities endure (Rourke & Fuerst, 1991; Spreen, 1988) and the deficits associated with learning disabilities typically persist into adulthood (Kaste, 1971; Mendelson, Johnson, & Stewart, 1971; Silver & Hagin, 1964; Spreen, 1982). Learning disabilities are the "largest single focus of special education in many school districts" (Winzer, 1993, p. 243). The study of learning disabilities and associated academic problems is burgeoning into an intensive area of investigation with growth that has been described "as little short of phenomenal" (Winzer, 1993, p. 243). Further, based on both prevalence and costs, the learning disabled population presents practical educational problems of major importance (Keogh, 1990).

Gaddes (1981) and Pirozzolo (1979) estimated that between 10 to 15% of children show seriously deficient academic attainment, although prevalence figures depend on the definition and procedures used to identify learning disabled subjects. In Alberta, it was estimated that, in

1991, there were just over 21,000 students identified as learning disabled, and of over 50,000 students identified as "exceptional", 40.6% were identified as learning disabled (Alberta Education, 1992). In Canada, students with learning disabilities make up the largest single group of children with disabilities, and generally, the best Canadian estimates place the number of students with learning disabilities at from 2 to 4% of the school-age population (Winzer, 1993). Duane (1979) estimated that with increased survival of high risk infants the population of children with learning disabilities will exceed the combined population of children with seizure disorders, cerebral palsies, and severe mental retardation. The consensus of scholarly opinion strongly suggests that cerebral dysfunctions underlie this disorder (Gaddes, 1985; Hooper & Willis, 1989; Hynd & Obrzut, 1981; Obrzut & Hynd, 1986; Pirozzolo, 1979; Rourke, 1991; Winzer, 1993) and the use of intelligence tests for classifying learning disabled children is "entrenched in every form of work with these children" (Francis, Espy, Rourke, & Fletcher, 1991, p. 15).

Winzer (1993, p. 243) maintained too, that the field of learning disabilities brought "changes and innovations to the entire field of special education" in the areas of instruction, assessment, and conceptualization of mild handicaps. Thus, it is important that efforts continue to be directed toward establishing reliable and valid means of identifying the presence of learning disabilities in order to facilitate remedial programming (Hooper & Willis, 1989). This is a complex task made all the more difficult by the fact that research suggests that there are numerous different subtypes of learning disability (Hynd, 1988), as well as controversy over conceptualization and operational definitions of learning disabilities (Keogh, 1990, Winzer, 1993). "The differential diagnosis of learning disability subtypes is a critical first step in developing theoretically sound programs of psychoeducational intervention" (Hynd, 1989, p. vii). Winzer (1993) stated "there is almost universal agreement on the need for efficient diagnosis of students with learning disabilities" (p. 253). Research techniques need to be directed toward the development of valid differential diagnostic procedures, based on theoretical clinically relevant classification schema (Adelman & Taylor, 1985, 1986). Evidence suggests such subtypes exist (Hynd, 1989) and currently much effort is being spent on determining subtypes according to patterns of disorder, particularly "different areas of underlying cognitive or psychologic dysfunction" (Forness, 1990, p. 195). Early attempts at subtyping were typically characterized by clinical inferential approaches (Forness, 1990). However, empirical multivariate cluster analytic procedures hold promise in developing classification schema in this area (Adelman & Taylor, 1985; Hooper & Willis, 1989; Kavale, 1990; Rourke, 1991) and research utilizing empirical approaches is growing significantly (Kavale, 1990). The role of the SB:FE in the diagnosis of learning disabilities, and its potential to contribute to learning disability subtyping research, in particular, needs to be investigated in order to improve current diagnosis and facilitate better remedial programming. Moreover, Lyon and Risucci (1988) emphasized that the scope of classification transcends mere categorization, but is also concerned with enhancing the theoretical understanding of learning disabilities. Similar lines of reasoning underscore the need to accurately identify and further understand mild handicaps such as Mental Retardation (Shepard, 1989) and the use of standardized intelligence tests is integral to work with such populations (Winzer, 1993). The potential of the SB:FE to contribute to these areas of understanding is presently unclear.

However, the SB:FE poses particular problems for multivariate research. The SB:FE scales are characterized by their adaptive testing and age scale formats (Thorndike et al., 1986a). It is possible for each of the four Reasoning Area scores, and therefore Composite Standard Age Scores (SAS's) to be composed of various numbers and different subtests. This tendency becomes more marked at different age levels. These difficulties are further compounded by the possibility that several tasks are thought to involve different abilities which depend on developmental levels (Keith, 1987; Keith et al., 1988a). It is worth noting that these methodological concerns, in general, may be applicable to virtually all measures of intelligence that span various age ranges. These concerns are exacerbated by the SB:FE's adaptive format and wide age coverage.

In light of the difficulties inherent in multivariate research with the SB:FE, a major purpose of this proposed research was to investigate the relationship between the SB:FE GPAB and the SB:FE, and to investigate the utility of the abbreviated battery in educational diagnosis and classification. Ideally, use of the abbreviated version may provide a means of ameliorating the difficulties that arise from the idiosyncratic nature of the SB:FE adaptive testing format. As noted, the complete battery for the SB:FE consists from eight to thirteen tests, depending on the age and ability level of the subject. The GPAB comprises six subtests that should be administered to any subject assessed with the SB:FE, regardless of age or ability level (Thorndike, Hagen, & Sattler, 1986b). Thorndike et al. (1986b) indicated that for all age ranges internal consistency (KR-20) (Kuder & Richardson, 1937) reliabilities of the abbreviated version are satisfactory (around .95), and Composite SAS's derived from the abbreviated version correlate very highly with Composite SAS's derived from the complete battery (r=.94-.98). These claims need to be investigated, particularly in terms of the relationship of the abbreviated battery to academic achievement within a Canadian population.

Research Objectives

The research reported here was designed to meet four main objectives:

1. To provide comprehensive descriptive data about the SB:FE used with a Canadian clinic sample;
2. To explore the relationship between the SB:FE full battery and the SB:FE GPAB within this sample;
3. Given close agreement between the two, to then investigate the applicability of multivariate cluster analytic procedures to SB:FE GPAB data in order to derive reliable and replicable (internally valid) groups of individuals with distinct cognitive profiles; and
4. To explore the external validity of groups derived through application of multivariate clustering procedures to SB:FE GPAB data through investigation of subgroup differences on the basis of external achievement criteria and to explore the agreement between empirically derived subtypes and clinical inferential models.

Throughout this study, short hand notation has been used to designate the age groupings. For example, 2-23-11 means 2 years through 23 years, 11 months; 2-6-11 means 2 years through

6 years, 11 months; 7-11-11 means 7 years through 11 years, 11 months; and 12-23-11 means 12 years through 23 years, 11 months.

Delimitations

The primary restrictions of this research centered around the possible idiosyncratic nature of the sample. Issues relating to the sample are more fully addressed in Chapter Three and the final chapter. The study was restricted to a post hoc or retrospective analysis of data. With respect to exploration of learning disabilities, only academic deficits which occurred in the three major areas of difficulty (reading, arithmetic, and spelling) (Winzer, 1993) were considered. The intent of the study was to provide evidence of the SB:FE's validity with Canadian subjects. Messick (1989) pointed out that "validity is an inductive summary of both the existing evidence for and the potential consequences of score interpretation and use" and then cautioned that "validity is a matter of degree, not all or none" maintaining that it is "an evolving property" and "a continuing process" (p. 13). This research was designed to be a part of this ongoing process in generating empirical evidence about the validity of the SB:FE in educational settings, within the parameters identified.

Outline of the Dissertation

This remaining portion of this dissertation is organized in seven chapters. In Chapter Two, a review of the development of the Binet intelligence scales culminating in the 1986 SB:FE is presented. An overview of learning disabilities is also provided, and empirical and clinical clustering or subgrouping research procedures are reviewed. The methodology used to address the objectives of the study (see p. 8) is presented in Chapter Three. The results are presented in four chapters corresponding to the research objectives. In Chapter Four the descriptive data about the SB:FE used with a Canadian clinic sample is presented. In Chapter Five the relationship between the SB:FE full battery and abbreviated version is presented. The results of the multivariate clustering procedures applied to SB:FE GPAB data are presented in Chapter Six. Chapter Seven contains the results of external validation procedures applied to the clusters derived through application of empirical clustering procedures to SB:FE data. A summary of the study, together with a discussion of the conclusions and the implications for school psychology practice and future research appear in the final chapter.

CHAPTER TWO: REVIEW OF RELEVANT LITERATURE

Overview

The present chapter begins with a description of the development of the SB:FE and the SB:FE GPAB. Next, their psychometric properties are examined. This is followed by a review of trends in educational assessment and planning. Emphasis is placed on a discussion of learning disabilities and definitional and historical issues related to learning disabilities. The conceptualization of learning disabilities as a heterogeneous multi-factor construct is examined. Finally, the notion of subtypes in learning disabilities is introduced and the validity of subtyping or subgrouping in research is reviewed.

Development of the Stanford-Binet Intelligence Scales

Binet and Simon (1905, 1908) initially developed the Binet-Simon scale to provide a screening instrument that would enable the French Minister of Public Instruction to identify mentally retarded children (Fancher, 1985). The scale comprised a series of 30 tasks of increasing levels of difficulty, standardized on groups of about 50 normal children of varying ages and 45 subnormals of varying degrees (Fancher, 1985). The 1905 test was atheoretical and empirically derived. In 1908, Binet and Simon published an extensive revision consisting of 58 items, again arranged in order of increasing difficulty, located at specific age levels between three and 13 years. The concept of mental "level" was also introduced (Freeman, 1955), although Binet cautiously did not use the term "mental age" (Fancher, 1985). In 1911, the third revision of the Binet-Simon scale was published by Binet alone, who extended the scale to include 15 year olds and a limited adult category (Fancher, 1985). Several items were relocated to higher age levels and several omitted, so that there were five items for each age level (Anastasi, 1982).

In 1909, Goddard translated the Binet-Simon Scale from French to English and introduced the scale to the United States with a number of revisions (Sattler, 1988, 1992). In 1911, Goddard tested 2,000 children for standardization purposes (Thorndike & Lohman, 1990) and became one of the world's leading proponents of Binet's testing methods (Fancher, 1985). Other researchers (Kuhlmann, 1912) published English translations of the Binet scales in the United

States (Thorndike & Lohman, 1990). However, in 1916, Terman of Stanford University, completed and published the most successful revision of the 1908 scale. This test was called The Stanford Revision of the Binet-Simon Scale (Terman, 1916). The revised test covered the age range from three to 16 years and was standardized on 1000 Californian children. In addition, groups of items were included for average and superior adult levels. A total of 90 items were included in the 1916 scale, of which 54 had been adopted from the 1911 Binet scale. Although the test remained basically an age-scale yielding a mental age, Stern's Intelligence Quotient (IQ) ratio (Stern, 1914) was adopted in order to report responses on the age scale in a condensed form. This ratio was calculated by the formula:

$$IQ = \frac{\text{Mental Age (MA)}}{\text{Chronological Age (CA)}} \times 100$$

Standardized administration procedures were also instituted, although these remained somewhat subjective and problematic (Freeman, 1955). The test was primarily designed as a measure of global intelligence and no attempt was made to measure separate mental faculties, although the distributions of IQ's was basically normal at each age group (Freeman, 1955). The test subsequently became known as the Stanford-Binet Intelligence Scale, and despite its limitations, became the standard against which all subsequent American intelligence tests would be measured (Fancher, 1985).

Terman and Merrill (1937) revised the scale again, extending the age range of the instrument from age 2 to 18 years and attempting to improve the standardization (based on 3,184 white American born subjects). The emphasis on measurement of general intellectual ability, rather than specific abilities was maintained. Age scale formats and the ratio IQ were retained. Two forms, L and M, were developed, each comprising 129 items. Sattler (1988) commented that the 1937 revision was "recognized as a milestone in the progress of the individual testing of intelligence" and noted that it had "excellent reliability and validity" (p. 246). Factor analytic studies indicated that most tests loaded heavily on a common factor and the tests "served as important tools in clinical and educational settings" (Sattler, 1988, p. 248).

The instrument was revised for a third time in 1960 (Terman & Merrill, 1960). No new content was introduced in this revision, but the best items from the two forms were selected and merged into the Form L-M. Deviation IQ's or standard scores derived from Yerkes (1917) and Wechsler (1939), although with a mean of 100 and a standard deviation of 16, were adopted for the first time in order to present test results for the sample age 2 through 18 years. The concept of mental age, however, was not abandoned. Norms were based on the 1937 sample and a sample of 4,498 subjects who had taken the scale between 1950 and 1954 was used to explore changes in item difficulty and to determine placement of items on the new form. The test remained a measure of general ability (Sattler, 1982).

R. L. Thorndike restandardized the 1960 revision, providing new interpretive norms in 1972, based on a more representative sample of 2,100 nonwhite and white children (Sattler, 1988). The revised norms were published in 1973 (Terman & Merrill, 1973). However, the tests in the scale, and directions for scoring and administration remained the same, and the test yielded a single score or measure of general intelligence. Salvia and Ysseldyke (1985) noted weaknesses in the norming, reliability, validity, and standardization of the 1972 Binet scale. Sattler (1982) provided a comprehensive review of the test.

It is worth noting that the successive revisions of the Stanford-Binet were intended as scales with a unitary focus. All purported to measure general intelligence. However, factor analytic studies (Burt & John, 1942a, 1942b; Hallahan, Ball, & Payne, 1973; Jones, 1949; McNemar, 1942; Ramsey & Vane, 1970; Thompson, 1984; Wright, 1939) suggested the presence of group factors (e.g., memory, verbal, visual-spatial, or numerical) in all editions of the test. In addition, earlier versions of the Stanford-Binet were consistently criticized for over-emphasizing verbal abilities (Krohn & Lamp, 1989). Moreover, the need for a revision of the Binet was paramount as during the 1970's and 1980's use of the Binet scales declined drastically (Lubin, Larsen, & Matarazzo, 1984; Lubin, Larsen, Matarazzo, & Seever, 1985). Thus, in light of current educational trends emphasizing differential abilities and specific areas of educational need (Sattler 1988), previous factor analytic findings, and criticisms of earlier versions of the Binet scales, continuity and logical evolutionary processes in test development can be found between the SB:FE and prior versions of the Binet. Substantial differences are also apparent.

The Stanford-Binet Intelligence Scale: Fourth Edition

Evolutionary factors and developmental continuity notwithstanding, Thorndike et al. (1986a), in creating the SB:FE, produced the "most radical revision of the Binet-Simon scale since its inception" (Thorndike & Lohman, 1990, p. 93). The test abandoned its age scale format, made use of advances in psychometric theory, and was revised according to an explicitly stated theoretical model of intelligence (Thorndike & Lohman, 1990). This model reflected a strong commitment to general intelligence, as well as a ready acceptance of group factors and an acknowledgment of the importance of both in educational applications. Adherence to a theoretical model in test development is consistent with contemporary psychometric theory, typically espoused in the Standards for Educational and Psychological Measurement (APA, 1985) and Guidelines for Educational and Psychological Testing (CPA, 1987). Thorndike et al. (1986b) attempted to maintain the strengths of previous versions and retain some degree of continuity. The revised instrument was designed to yield clinically useful profiles, as well as an overall general ability score and effort was made to de-emphasize verbal skills (Keith et al., 1988a).

As noted in Chapter One, a three level hierarchical model of the structure of cognitive abilities was adopted to guide construction of the SB:FE (Thorndike et al., 1986a). General reasoning ability, or g, traditionally associated with Binet scales, comprises the first level (Glutting, 1989). The second level comprises three group factors identified as Crystallized Abilities, Fluid-Analytic Abilities, and Short-term Memory. The first two dimensions at this level originate from the Cattell-Horn theory of intelligence (Horn, 1976, 1982; Horn & Cattell, 1966). The additional component, Short-term Memory, not contained in the Cattell-Horn theory, derives from information processing theory (Baddeley, 1976; Vallar & Baddeley, 1982). Factors at the third level are identified in terms of three facets of reasoning: Verbal; Quantitative; and Abstract/Visual Reasoning. The Verbal and Quantitative Reasoning Areas reflect Crystallized Abilities and Abstract/Visual Reasoning Area reflect Fluid-Analytic Abilities and together with the Short-term

Memory factor at the third level comprise the four Reasoning Area scores derived by the SB:FE. The Short Term Memory Area in the SB:FE does not subsume specific factors. Fifteen subtests form the SB:FE's foundation (Glutting, 1989). In addition, the SB:FE is a power test; only one subtest, Pattern Analysis, requires mandatory time limits (Glutting, 1989).

Sattler (1988) noted that, within each area, specific subtests are designated. These designations were not based on factor analytic results, but rather were rationally derived. The test is designed for use with subjects age 2 to 23 years, 11 months. Fifteen subtests were included in the SB:FE in order to measure the various factors; nine were retained from the older version and six were newly developed. Raw scores are converted into three types of score: Standard Age Scores for the subtests (M=50, SD=8); four Reasoning Area scores (M=100, SD=16); and a Composite Standard Age Score (M=100, SD=16). The Composite Standard Age Score (SAS) is a global estimate of intellectual functioning and is similar to the Deviation IQ employed in the Wechsler Scales (Glutting, 1989; Sattler, 1988). The term Standard Age Score was adopted to avoid some of the connotations associated with the term IQ (Murphy & Davidshofer, 1991; Thorndike & Lohman, 1990).

Thorndike et al.'s (1986a, 1986b) conceptualization of g is similar to that proposed by Spearman (1904). Thus, g is viewed as "consisting of the cognitive assembly and control processes that an individual uses to organize adaptive strategies for solving novel problems" (Thorndike et al., 1986a, p. 3). The Crystallized Ability factor, which is similar to Horn and Cattell's (1966) postulate, subsumes Verbal and Quantitative Reasoning. Within the framework postulated by Horn and Cattell (1966), crystallized intelligence reflects mental abilities that depend on experience with the world, e.g., formal schooling and informal learning experiences (Schaie & Willis, 1991). The Fluid-Analytic Abilities factor subsumes the Abstract/Visual reasoning area, and shares commonalties with Horn and Cattell's (1966) notion of fluid intelligence or Wechsler's "native mental ability" (Schaie & Willis, 1991, p. 402). This is identified as the innate-biological element of intelligence (Horn, 1982; Horn & Cattell, 1966). As noted, the Short-term Memory factor in the SB:FE does not subsume specific factors, but in 1976, in a review of human abilities, Horn acknowledged that a memory factor exists, although not including it in earlier work. The conceptualization of short term memory as a "passive storage buffer" is being replaced with an emphasis on short term memory as an active component with separate visual, auditory, and kinesthetic systems (Baddeley,1976; Vallar & Baddeley, 1982). In addition, "the information-processing emphasis in modern cognitive research relates short term memory to more complex aspects of cognitive performance" (Thorndike et al., 1986a, p. 4). This is reflected in the SB:FE.

Thorndike et al. (1986a, 1986b) also freely incorporated many of Vernon's ideas into the construction of the SB:FE, particularly with respect to verbal-educational and practical mechanical factors. Vernon (1961) proposed a hierarchical model of intelligence with g at one level, and a few highly specific group factors. Vernon (1961) maintained that once the effects of g are removed two specific group factors remain, verbal educational factors and spatial practical mechanical factors. These two specific factors can be broken down into factors of a more minor nature. However, Vernon (1961) was critical of multiple factor theorists who carried this process of reduction too far, noting that while group factors may be infinitely subdivisible, the range of performance they account for becomes meaningless. Kail and Pelligrino (1985) identified many similarities between Horn and Cattell's (1966) and Vernon's (1961) theories.

The SB:FE has been accepted as a replacement for the third edition (Molfese et al., 1992; Sattler, 1988, 1992; Thorndike & Lohman, 1990). Generally, the greatest area of contention, and concomitantly research, has focused on issues relating to the correspondence between the theory underlying the SB:FE and its factor structure. Two issues are paramount. The first relates to whether the SB:FE Composite scores adequately reflect general intelligence. The second relates to whether the SB:FE measures separate ability domains (Molfese et al., 1992). Data support the presence of g in the Composite Score (Sattler, 1992; Thorndike et al., 1986b). However, a number of researchers contend that the hierarchical theory behind the SB:FE is not confirmed (Sattler, 1992), although the presence of the four cognitive areas has generally received modest support from the results of factor analytic studies. This line of research will be reviewed in a later section.

Psychometric Properties of the SB:FE

Standardization Data

The SB:FE standardization sample consisted of 5,013 individuals in 17 age groups (2 years 0 months through 23 years 11 months). The sample was designed to be nationally representative of the United States population and was selected on the basis of five primary standardization variables: age, sex, ethnicity, geographic location, and community size (Thorndike et al., 1986b). Data from the 1980 U.S. Census were used to allocate sample proportions for these variables. In addition to these specific stratification variables, the socioeconomic status (parental occupation and educational attainment) of examinees was monitored.

A significantly larger number of children who came from upper socioeconomic (SES) categories were sampled (Thorndike et al., 1986a, 1986b) (43.1% vs. 19.0% of the U.S. population) and lower SES examinees were underrepresented (10.6% vs. 29.2% respectively). These disproportions have been attributed to the necessity of obtaining examinees consent for testing and to the fact that "parents of higher SES tended to return permission forms more frequently than those of lower SES" (Thorndike et al., 1986c, p. 24). Weighting procedures were adopted to overcome this discrepancy. Children from advantaged backgrounds were counted as only a fraction of one case, while children from disadvantaged backgrounds were counted as more than one case. The specific weights that were applied appear in The Stanford-Binet Intelligence Scale: Fourth Edition, Technical Manual (SB:FE Technical Manual) (Thorndike et al., 1986b, p. 25). Sandoval and Irwin (1986), Slate (1986), and Vernon (1987) were critical of this weighting procedure. However, Keith et al. (1988a) agreed with the test publishers that the weighting procedure used was probably the best method to deal with the overrepresentation/underrepresentation in the sample. Glutting (1989) essentially concurred, noting that weighting is an accepted procedure for standardizing group tests (p. 74) and commented that the advantage of weighting was that it accounted for all scores in the samples. However, Glutting (1989) indicated that weighting probably produced estimates of higher reliability than would elimination.

Reliability

Thorndike et al. (1986a, 1986b) provided reliability estimates for the SB:FE derived from the norming study data. Kuder-Richardson Formula 20 (KR-20) (Kuder & Richardson, 1937) internal consistency reliability coefficients for the SB:FE Composite, Reasoning Area, and subtest scores were calculated. In order to use the KR-20 formula, all items below the basal level were assumed to be passed and all items above the ceiling level assumed to be failed. Thorndike et al. (1986b) conceded that this procedure tends to yield inflated values and cautioned that the internal consistency coefficient should be considered "upper bound estimates" (p. 38).

The most reliable (internally consistent) score is the Composite SAS. Internal consistency estimates for the Composite and four Reasoning Area SAS's are presented in Table 1.

Table 1

Internal Consistency Estimates for SB:FE Reasoning Area and Composite SAS's

Age	Ver. SAS	A/V SAS	Qnt. SAS	STM SAS	Com. SAS
2	.90-.93	.85	.81	.90	.95
3	.92-.94	.87	.84	.92	.96
4	.91-.93	.91	.87	.90	.97
5	.91-.93	.93	.88	.92	.97
6	.87-.91	.92	.80	.90	.96
7	.86-.90	.92-.94	.89	.88-.93	.97
8	.89-.92	.93-.95	.92	.88-.93	.97
9	.88-.92	.94-.95	.91	.86-.91	.97
10	.93-.95	.94-.96	.93	.87-.92	.98
11	.92-.94	.94-.95	.93	.88-.93	.98
12	.94-.97	.94-.96	.94-.96	.88-.92	.98
13	.94-.97	.93-.96	.95-.96	.90-.94	.99
14	.94-.97	.95-.96	.94-.96	.88-.93	.98
15	.94-.96	.95-.96	.95-.96	.89-.93	.98
16	.94-.96	.95-.96	.95-.96	.88-.93	.98
17	.95-.97	.95-.96	.95-.96	.91-.94	.99
18-23	.95-.97	.96-.97	.96-.97	.92-.95	.99

Note: Ranges for internal consistency estimates are presented as estimates and depend on the number of subtests used to form the reasoning areas and may differ at each age level. Adapted from Thorndike et al. (1986b).

As can be seen from Table 1, the internal consistency estimates typically increase with age, particularly with respect to Composite SAS scores, with a median internal consistency estimate of .97. Based on these internal consistency estimates the Standard Errors of Measurement (SEM's) within the Composite SAS range from 1.6 at age 18-23, to 3.6 at age 2. Within the four Reasoning Areas, internal consistency estimates also typically increase with age, but are also influenced by the number of subtests that are used in the calculation of the reliability estimates. The internal consistency estimates for the four Reasoning Areas range from a low of .80 in the Quantitative Reasoning Area at age 2 to a high of .97 at age 18-23 in the Verbal, Abstract/Visual, and Quantitative Reasoning Areas.

Reliability estimates for the subtests are lower than the estimates obtained for the total test and four Reasoning Area SAS's. As shown in Table 2, median internal consistency reliabilities across all ages for the subtests range from .73 for Memory for Objects to .94 for Paper Folding and Cutting. Subtest internal consistencies differ according to age group, ranging from a low of .66 for Memory for Objects at age 10, to a high of .96 for Pattern Analysis at ages 18-23 (Sattler, 1988).

Table 2

Median Internal Consistency Estimates and Ranges for SB:FE Subtests Across Ages 2-23-11

SB:FE Subtest	Median Reliability	Range
Verbal Reasoning Area		
Vocabulary	.87	.78-.94
Comprehension	.89	.79-.96
Absurdities	.87	.74-.94
Verbal Relations	.91	.90-.91
Abstract/Visual Reasoning Area		
Pattern Analysis	.92	.81-.96
Copying	.87	.74-.92
Matrices	.90	.88-.92
Paper Folding and Cutting	.94	.93-.94
Quantitative Reasoning Area		
Quantitative	.88	.80-.95
Number Series	.90	.86-.93
Equation Building	.91	.90-.92
Short Term Memory Area		
Bead Memory	.87	.82-.95
Memory For Sentences	.89	.85-.94
Memory For Digits	.83	.80-.88
Memory for Objects	.73	.66-.78

Adapted from Thorndike et al. (1986b)

Test-retest reliability estimates were obtained from retesting 112 children. The sample consisted of two groups; the first tested at approximately five years of age (n=57), and the second tested at approximately eight years of age (n=55). The length of time between testing varied from 2 to 8 months, with the average interval between sessions 16 weeks. As can be seen from Table 3, the test-retest estimates are generally lower than the respective internal consistency estimates (cf., Table 2) derived using the KR-20 formula (Kuder & Richardson, 1937). Thorndike et al. (1986b) suggested this may reflect restrictions in range of scores for both age groups as well as some degree of practice effect. However, caution may be needed in interpreting test scores relying on Standard Errors of Measurement (SEM's) derived from K-R 20 values. In addition, Sattler (1988) cautioned against relying on individual subtests to provide stable measures of ability.

Table 3

Stability Estimates for SB:FE Subtest Scores, Reasoning Area, and Composite SAS's

	Stability Coefficient	
SB:FE Test	Sample One	Sample Two
Verbal Reasoning Area	.88	.87
Vocabulary	.75	.75
Comprehension	.69	.86
Absurdities	.70	.74
Verbal Relations	N/A	N/A
Abstract/Visual Reasoning Area	.81	.67
Pattern Analysis	.77	.78
Copying	.71	.46
Matrices	N/A	.63
Paper Folding and Cutting	N/A	N/A
Quantitative Reasoning Area	.71	.28
Quantitative	.71	.61
Number Series	N/A	N/A
Equation Building	N/A	N/A
Short Term Memory Area	.78	.81
Bead Memory	.56	.62
Memory For Sentences	.78	.72
Memory For Digits	N/A	.66
Memory for Objects	N/A	.61
Composite SAS	.91	.91

Adapted from Thorndike et al. (1986b)

The reliability estimates (internal consistency and stability) presented by Thorndike et al. (1986b) were reviewed by Sandoval and Irwin (1986) and Vernon (1987). In general, support was found for the internal consistency of the instrument, although Sandoval and Irwin (1986) suggested that standard errors of measurement should be based upon test-retest reliability estimates rather than on internal consistency reliability estimates. They also suggested that (SEM's) should be provided for each ability level.

Validity

In the SB:FE Technical Manual (Thorndike et al., 1986b) evidence of the construct validity of the SB:FE is reviewed. Thorndike et al. (1986b) viewed construct validity by focusing on three types of evidence. First, they examined the correlations among the subtests, which are then clarified by factor analysis. Second, they examined the correlations of the SB:FE with scores on other tests deemed to measure the same or similar constructs. Third, they examined the performance of groups on the SB:FE identified by indices or criteria other than the SB:FE. The evidence relating to the validity of the SB:FE that is presented in the SB:FE Technical Manual (Thorndike et al., 1986b) is discussed in this order.

Construct Related Evidence

Intercorrelations and Factor Analysis

Thorndike et al. (1986b) maintained that factor analysis supports a number of the theoretical dimensions of the scale including adequate to high g loadings of the subtests. Median correlations based on those age groups who had taken particular pairs of tests were calculated. Median correlations were used, according to Thorndike et al. (1986b), in order to reduce the weight given to extreme correlation values. An unspecified variant of confirmatory factor analysis was carried out on the median correlations (Thorndike et al., 1986b). A general factor was extracted first, then a group factor that accounted for the residual correlation among tests assigned to a given content area. This was performed in the following order: Verbal; Memory; Quantitative; and Abstract/Visual. Thorndike et al. (1986b) maintained that all tests received substantial loadings on g and that confirmation was received for each of the four area scores. All four verbal tests loaded on the Verbal factor with loadings ranging from .26 to .47. Three memory tests loaded significantly on the Memory factor, although Bead Memory showed loadings of only .13 on the Memory factor and .13 on the Abstract/Visual factor. Thorndike et al. (1986b) also noted that the three quantitative tests showed modest loadings (.21 to .49) on the Quantitative factor (p. 54). The final factor, the Abstract/Visual area was, according to Thorndike et al. (1986b), less well defined, with substantial loadings on Pattern Analysis (.65), modest loadings on Paper Folding and Cutting (.23), and loadings of only .04 for Matrices and .15 for Copying.

The SB:FE's authors proceeded to discuss additional results of factor analyses for three age ranges, 2 through 6 years (2-6-11), 7 through 11 years (7-11-11), and 12 through 18-23 years (12-23-11). These age groups were selected because the age cohorts typically took the same subtests from among the 15 subtests included in the standardization. Correlations were calculated separately for each age group and median correlations used for analyses. For the 2-6-11 age group, a Verbal factor appeared, with a primarily Abstract/Visual second factor. No memory factor emerged, with the verbal memory test (Memory for Sentences) loading on the Verbal factor, and the abstract visual memory task (Bead Memory) loading on the Abstract/Visual factor. Only one quantitative test was administered at this level, thus no common quantitative factor was expected. Results of the factor analysis for age 7-11-11, indicated a strong Verbal factor, with loadings on Vocabulary, Comprehension, and marginal loadings on Absurdities and Memory for Sentences. Three tests, Memory for Sentences, Memory for Digits, and Memory For Objects defined a Memory factor. A small Abstract/Visual factor defined primarily by Pattern Analysis was also obtained. The Bead Memory subtest loaded only on the general factor as did the Number Series and Quantitative tests. In the group 12-23-11, a well defined Verbal factor, comprising Vocabulary, Comprehension, and Verbal Relations emerged. An Abstract/Visual factor comprising Pattern Analysis and Paper Folding and Cutting was identified. All four memory subtests, Memory for Sentences, Memory for Digits, Memory for Objects, and Bead Memory loaded on a well-defined Memory factor. In addition, a Quantitative factor comprising the Quantitative, Number Series, and Equation Building subtests was also identified.

Thorndike et al. (1986b) concluded:

> Generally, the factor structure provides positive support for the rationale underlying the battery. Clearly all the tests share primarily their function as measures of general cognitive ability. However, the area scores do have a defined structure, apart from their function as measures of *g* (p. 55).

However, critical examination of the conclusions reached by Thorndike et al. (1986b) about the underlying structure of the SB:FE suggests this interpretation cannot be fully endorsed, particularly across all age groupings, and in light of the lack of clarity concerning the specific analyses used. Fritzke (1988) reported that, in a personal written communication with R. L. Thorndike dated June 2, 1987, it was conceded by Thorndike that the "procedure of (factor) analysis was slightly idiosyncratic" (p. 120). The assertions made by Thorndike et al. (1986b) need to be substantiated and will be addressed in a review of subsequent research which follows this section.

Correlations of the SB:FE with Other Tests of Cognitive Ability

Thorndike et al. (1986b, p. 58) presented further evidence pertaining to the construct validity of the SB:FE as a measure of intelligence by determining correlations with scores on the SB:FE and other individual intelligence tests. This evidence, based on both exceptional and non-exceptional samples, suggested there was comparability among the different cognitive ability tests used and that the underlying constructs measured share something in common. Non-exceptional sample results are presented first.

In a sample of 139 subjects with a mean age of 6 years 11 months, SB:FE Composite SAS scores correlated .81 with Stanford-Binet: Form L-M (S-B:LM) (Terman & Merrill, 1973) total scores. The SB:FE mean Composite SAS was 105.8, with a standard deviation of 13.8. The mean total score on the S-B:L-M was 108.1, with a standard deviation of 16.7. The SB:FE Verbal, Abstract/Visual, Quantitative, and Short Term Memory Reasoning Area scores correlated .76, .56, .70, and .67 respectively with the total scores on the S-B:L-M.

In a sample of 205 subjects with a mean age of 9 years and 5 months, a correlation of .83 was obtained between Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974) Full Scale IQ scores and SB:FE Composite SAS scores. Significant correlations, ranging from .63 to .73, were obtained between all four SB:FE Reasoning Area scores and the Verbal, Performance, and Full Scale IQ scores of the WISC-R. In a sample of 75 subjects with a mean age of 5 years 6 months, a correlation of .80 was obtained between the SB:FE Composite SAS and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) (Wechsler, 1967). Significant correlations, ranging from .46 to .80, were obtained between all four SB:FE Reasoning Area scores and the Verbal, Performance, and Full Scale IQ scores of the WPPSI. The relationship between the Wechsler Adult Intelligence Scale-Revised (WAIS-R) (Wechsler 1981) and the SB:FE was investigated in a sample of 47 adults with a mean age of 19.5. The correlation between the respective total score indices was .91, with correlations ranging from .65 to .86, between the four SB:FE Reasoning Area scores and the Verbal, Performance, and Full Scale IQ scores on the WAIS-R.

Finally, in a sample of 175 subjects with a mean age of 7 years, a significant correlation of .89 was obtained between the Kaufman Assessment Battery for Children (K-ABC) (Kaufman & Kaufman, 1983) Mental Processing Composite and the SB:FE Composite SAS. Correlations, ranging from .68 to .83, were also obtained between the SB:FE Reasoning Area scores and K-ABC Sequential and Simultaneous Processing and Achievement scores.

Results with samples of exceptional children are presented next in the SB: FE Technical Manual (Thorndike et al., 1986b). Two studies of gifted children were conducted. The SB:FE correlated with the S-B:L-M (r=.27) in a sample of 82 subjects with an average age of 7 years 4 months. In a sample of 19 gifted children with an average age of 12 years 11 months, the SB:FE correlated positively with the WISC-R (r=.69). SB:FE Composite SAS scores and WISC-R Full Scale IQ scores were very similar, 116.3 and 117.7 respectively. Three studies were conducted with students designated as learning disabled by their schools. The SB:FE correlated with the S-B:L-M (r=.79) in a sample of 14 subjects with a mean age of 8 years 4 months. The SB:FE correlated significantly (r=.87) with the WISC-R in a sample of 90 subjects with an average age of 11 years. In a sample of 30 subjects with an average age of 8 years 11 months, the SB:FE correlated with the K-ABC (r=.74). Three studies were conducted with sample of subjects designated by their schools as mentally retarded. The SB:FE correlated with the S-B:L-M (n=22, r=.91), with the WISC-R (n=61, r=.66), and with the WAIS-R (n=21, r=.79).

Studies of Differences in Performance of Exceptional Children

Finally, evide:r.e pertaining to the construct validity of the SB:FE, cited in the SB:FE Technical Manual (Thorndike et al., 1986b) related to a series of studies investigating differences in performance of special groups of examinees: gifted; learning disabled; and mentally retarded. Differences in mean Composite SAS's were noted between the three groups in the direction anticipated, i.e., gifted subjects scored significantly above the standardization mean (SB:FE Composite SAS M=123.2), with learning disabled and mentally retarded subjects scored significantly below the standardization mean (SB:FE Composite SAS M=85.1 and M=54.9 respectively). The group of mentally retarded subjects scored significantly below the learning disabled group.

The evidence presented by Thorndike et al. (1986a, 1986b) relating to the validity of the SB:FE has not been accepted unequivocally. Vernon (1987) was particularly critical of the SB:FE, maintaining that it seemed to be more similar in structure to the Wechsler Scales than the original Binet scales. Vernon (1987) also noted that the SB:FE may measure different things at different age levels, suggesting a lack of comparability in the content of area scores at different ages. Spelliscy (1991) noted that the validity of the factorial results in particular has been questioned. Spelliscy's (1991) reservations are consistent with concerns expressed by Glutting (1989), Keith et al. (1988a), Keith, Cool, Novak, White, and Pottebaum (1988b), Molfese et al. (1992), Osberg (1986), Sandoval and Irwin (1986), Sattler (1988), Sim (1987), Slate (1986), and Vernon (1987). However, Laurent, Swerdlik, and Ryburn (1992) in a comprehensive review of the validity research on the SB:FE concluded that the results cited by (Thorndike et al., 1986b) suggested good to very good criterion-related validity (most in the range .80 to .90).

Results of Factor Analytic Studies Conducted With the SB:FE Since Thorndike et al. (1986b)

Sattler (1988) reported that the factor analytic findings cited in the SB:FE Technical Manual, supported the existence of adequate to high g loadings of the subtests, with specific factors at various age levels of the scale. In addition, all the subtests correlate moderately to highly positively with the Composite SAS's (Sattler, 1988). However, Sattler (1988) reported on a principal components analysis with varimax rotation conducted on the SB:FE standardization data that differs from the results of the analysis presented by Thorndike et al. (1986b). Sattler (1988) noted that, given the lack of continuity of subtests and the fact that different subtests are administered at different ages, the factor structure of the scale differs at different ages. Sattler (1988) maintained that at ages 2 through 6, a two factor solution characterized the scale. These two factors are labelled Verbal Comprehension and Nonverbal reasoning. At ages 7 through 23, Sattler (1988) suggested a three factor solution is the most appropriate. These three factors are labelled Verbal Comprehension, Nonverbal Reasoning, and Memory. Consequently, Sattler (1988, 1992), as one of the SB:FE's authors, cautioned against using content area scores for interpretive purposes when using the SB:FE and advocated an interpretive model different from that proposed by Thorndike et al. (1986b), based on factor scores. In addition, the presentation of divergent models (Sattler, 1988) clearly set the stage for ongoing debate about the theoretical constructs underlying the SB:FE.

Reynolds, Kamphaus, and Rosenthal (1988) used exploratory factor analysis procedures to investigate the factor structure of the SB:FE using the correlation matrix from the standardization sample (Thorndike et al., 1986b). They reported findings that did not support Thorndike et al.'s (1986b) claims, and recommended using only SB:FE composite scores as they found a large first factor, with little support for the four Reasoning Areas.

Fritzke (1988) examined the concurrent and construct validity of the SB:FE by exploring the component structures of the WISC-R, the SB:FE, and the joint battery of the WISC-R and SB:FE utilizing a Canadian clinical sample of 168 children aged 6-0 to 16-11. Fritzke (1988) divided the sample into three age groups, 6-0 to 8-11, 9-0 to 12-11, and 13- to 16-11, and found significant correlations between the WISC-R Full Scale scores and the SB:FE Composite Scores of .87, .91, and .97 respectively for each age group sampled. The Verbal Scale of the WISC-R was also found to correlate significantly with the Verbal Reasoning Area of the SB:FE .81, .89, and .96 respectively for each age group. Correlations between the Performance Scale of the WISC-R and the Abstract/Visual Reasoning Area were .61, .89, and .92 respectively for each age group.

Fritzke (1988) then performed principal component analysis followed by quartimax and varimax rotations. Support for the organization of the SB:FE into the four cognitive areas was not found. However, based on Fritzke's (1988) results the SB:FE appeared to be a good measure of general intelligence. Some care needs to be exercised in interpreting Fritzke's (1988) factor analytic findings. Sample sizes are relatively small, and Fritzke used different age groupings from those used by Thorndike et al. (1986b). As Spelliscy (1991) noted, confirmatory factor analysis would have been more appropriate in testing the specific model outlined by Thorndike et al. (1986b). It is important to note that Sattler's (1988), Reynolds et al.'s (1988) and Fritzke's (1988) results are based on exploratory factor analysis.

However, Kline (1989), used confirmatory factor analyses to determine the goodness of fit of Thorndike et al.'s (1986b) model and Sattler's (1988) model to subtest intercorrelations within each of the 17 age groups of the standardization sample. Results tended to support Sattler's two factor Verbal and Nonverbal model for the subjects 2 through 6 years and the three factor, Verbal, Nonverbal and Memory for older subjects (Kline, 1989). Kline (1989) found little support for the model proposed by Thorndike et al. (1986b). Laurent et al. (1992) were critical of Kline (1989) for failing to specify whether a strict or relaxed model was used in the analysis and noted that lack of procedural detail in the study precluded full comment on the results.

R. M. Thorndike (1990) reviewed and critiqued some of the factor analytic studies conducted on the SB:FE. Thorndike (1990) noted in particular that Reynolds et al.'s (1988) factor analysis provided nothing more than "a computer center default solution" (p. 418). Thorndike (1990) reported on a reanalysis of the standardization data, noting the results provide strong evidence of g as a second-order factor of general intelligence in the test. Results of analyses using a principal axis factoring of correlation matrices conducted on each age group were not entirely supportive of the four factor model proposed by Thorndike et al. (1986b). At age 2-7 years, a two factor model (verbal and nonverbal ability) seemed most appropriate, whereas for subjects above age 7, a three factor model seemed most appropriate, with verbal, abstract/spatial, and memory factors emerging. Results thus seemed more supportive of Sattler's (1988) findings. Thorndike (1990), however, also concluded that analysis of the median correlations across all ages supported the existence of the four factors hypothesized by the test constructors (Thorndike et al., 1986b).

Gridley and McIntosh (1991) used confirmatory factor analysis on two samples, one age 2-6 years (n=50), and one aged 7-11 years (n=137) that were not part of the standardization sample. Support for the strict four factor model proposed by Thorndike et al. (1986b) was not found for either sample. Support for the interpretive structure suggested by Sattler (1988) was found for the subjects age 2-6, but not for the older subjects. However, an alternative model with verbal, quantitative, abstract/visual, and memory factors that allowed some subtests to load on more than one factor was proposed. This conceptualization offered partial support for the theoretical structure hypothesized by Thorndike et al. (1986b).

Ownby and Carmin (1988) used confirmatory factor analytic procedures, using the standardization data, to test four models of hypothetical models underlying the SB:FE. The first model was based on the general ability model presented by Thorndike et al. (1986b) as the overarching construct behind the SB:FE. The second model was based on Wechsler's Verbal-Nonverbal model. The third model proposed included a third factor, memory, in addition to verbal and non-verbal factors. The fourth model was based on the four cognitive area factors proposed by Thorndike et al. (1986b). The analysis was conducted on the entire sample and separately on 2, 4, 6, 8, and 10 year old samples. A four factor model fit the data best for the analysis conducted on the entire sample. For the 2-year-olds a two factor model was best, but for the 4-, 6- and 10-year olds, four factor models were satisfactory. For the 8-year old group, the four factor model was best. The g only model did not provide the best fit for the data. Laurent et al. (1992) noted that it was unfortunate that Ownby and Carmin (1988) tested only strict models that would not allow subtests to load on more than one factor, although, in general, the results provided some support for the theoretical structure porposed by Thorndike et al. (1986b).

Keith et al. (1988a) investigated the construct validity of the SB:FE using confirmatory factor analysis, with the specific intent of testing the "explicit underlying theory" of the SB:FE (p. 258) to determine the extent to which the existence of the four cognitive areas of ability was supported.

Keith et al. (1988a) noted that the factor analytic results presented by Thorndike et al. (1986b) provide inadequate information about the test's factor structure, because the factor loadings reported were generally small. Confirmatory factor analysis involves specifying the number of factors, determines whether such factors are correlated or uncorrelated, and identifies the specific tests that load on each factor. Results of confirmatory factor analytic procedures include statistics (goodness of fit) that indicate how well the specified model fits that data and, in addition, provide information on how to modify the model in order to better fit the data (Keith et al., 1988a). Keith et al. (1988a) noted that confirmatory factor analysis provides a less subjective and "much stronger test of the underlying structure of a test instrument than does exploratory factor analysis" (p. 258).

First-order confirmatory factor analyses were performed on the entire SB:FE standardization data utilizing median intercorrelation data obtained from the SB:FE Technical Manual (Thorndike et al., 1986b). Four analyses were conducted by Keith et al. (1988a). Consistent with Thorndike et al. (1986a, 1986b), these analyses were performed on samples from ages 2-6, 7-11, 12-23, and a combination of ages 2-23. Models were specified that matched the hypothesized four areas of cognitive ability, i.e., the first level proposed by Thorndike et al. (1986b). Thus, Vocabulary, Comprehension, Absurdities, and Verbal Relations subtests were specified as loading only on a verbal reasoning factor. Pattern Analysis, Copying, Matrices, and Paper Folding and Cutting subtests were specified to load on an abstract/visual factor. Quantitative, Number Series, and Equation Building subtests were allowed to load on a quantitative factor. Finally, Bead Memory, Memory for Sentences, Memory for Digits, and Memory for Objects subtests were specified to load on a short term memory factor. These four areas or factors were allowed to correlate with each other, consistent with the underlying theory of the SB:FE (Thorndike et al., 1986b). Hierarchical analysis specifying two levels of factors beyond the subtest level was not performed as the first order analysis was considered by Keith et al. (1988a) to be a reasonable initial confirmatory step. In addition, Keith et al. (1988a) noted that given the theoretical model underlying the SB:FE it was expected that verbal reasoning and quantitative reasoning factors would correlate more highly with each other as measures of crystallized intelligence than with the other two factors.

The results, based on the entire sample, tended to be supportive of the first two levels (subtest and cognitive areas) of the theoretical model underlying the SB:FE proposed by Thorndike et al. (1986a), as well as confirming the existence of g. Factor loadings and intercorrelations were significant. The researchers, however, note some inconsistencies with the SB:FE's theoretical model in that the correlation between the verbal reasoning factor and the quantitative factor was the lowest intercorrelation. Indeed, the quantitative factor correlated most highly with the abstract/visual factor which reflects Fluid-Analytic abilities, according to the proposed theoretical model.

A more relaxed model was tested next, allowing various subtests (Absurdities, Memory for Sentences, and Bead Memory) to load outside of their hypothesized areas. A better fit to the standardization data was obtained under these constraints. In the revised model, Bead Memory loaded on the abstract/visual reasoning factor, while Memory for Sentences loaded on the verbal reasoning factor.

Confirmatory factor analysis was then conducted for each of the three age groups using the same constraints as applied in the analysis involving the whole sample. Results were "quite similar to those for the entire standardization sample" (Keith et al., 1988a, p. 264) for the 12- 23 age group. Results for the 7-11 age group were also quite similar. Keith et al. (1988a) noted that the analysis on the sample aged 2-6 proved to be the most difficult, leading them to conclude the theory underlying the SB:FE does not adequately explain the structure of the instrument with that particular age group. They suggested a three factor model underlies the SB:FE at that age group. The three factors are verbal, quantitative, and abstract visual reasoning.

Keith et al. (1988a) concluded "the results of their analyses offer reasonably good support for the construct validity of the new Binet" (p. 271). They maintained that the first level of the SB:FE theory provided a generally good fit to the standardization data for the total standardization sample and for two out of three of the age groups. As well, results of analyses generally supported the existence of the four cognitive ability areas for most age levels and the entire standardization sample, with the caveat that, in the age group 2-6, a memory factor was difficult to isolate. The presence of g in the SB:FE was also strongly supported, given the strong correlations among the first order factors. In addition, findings by Keith et al. (1988a) rebut the views expressed by Sandoval and Irwin (1986) and Slate (1986) that the SB:FE measures little other than g. They tested the strict SB:FE model against a model that assumes that all of the SB:FE subtests measure only g to see which provided a better fit to the standardization data. "The 'g-only model' provided a significantly worse fit to the data than did the strict models" (Keith et al., 1988a, p. 270). They concluded that, whilst the SB:FE was heavily laden with g, results suggested that the SB:FE measured other constructs as well. The level of the SB:FE theoretical model which posits division of the scale into Crystallized abilities (Verbal and Quantitative Reasoning Areas) and Fluid-Analytic abilities (Abstract/Visual Reasoning Area) received little support, suggesting caution is needed in interpreting this level. Keith et al. (1988a) also suggested that confirmatory hierarchical factor analysis be conducted to further examine the second level of the SB:FE theory. Keith et al. (1988b) extended their (Keith et al., 1988a) study to include higher order confirmatory factor analysis of SB:FE standardization data. As expected, evidence supported the existence of the four area scores, but little support was found for the Crystallized and Fluid-Analytic dimensions.

Hoffman, Carleton, Bishop-Marbury, and Goodwin (1988) applied similar confirmatory factor analytic techniques to the standardization data. Results provided support for the four factor structure of the instrument posited by Thorndike et al. (1986b).

Boyle (1989) also criticized the confirmatory factor analytic evidence presented by Thorndike et al. (1986b) and identified several inconsistent factor loadings in the original work, without reasonable explanation. Boyle (1988) refactored the data in the SB:FE Technical Manual (Thorndike et al., 1986b) using an exploratory principal components factor analytic procedure applied to the total standardization sample. Boyle (1989) noted that the inconsistencies identified in the original work were likely an artifact of the procedures used by Thorndike and colleagues

(1986b) and concluded that the refactored results provided unequivocal support for the structural dimensionality of the revised test (p. 714).

Boyle (1990) conducted a LISREL (Joreskog & Sorbom, 1986) reanalysis of the subtest intercorrelations provided by Thorndike et al. (1986b). Confirmatory congeneric factor analysis strongly supported the four area dimensions posited in the theoretical model (Thorndike et al., 1986b) and confirmed the findings of Keith et al. (1988a, 1988b). Some subtests loaded differently from those reported by Keith et al. (1988a), although each area dimension was supported strongly. Significant intercorrelations were reported among the four reasoning area scores confirming the influence of g in the SB:FE. Confirmatory factor analysis was also applied to the three factor model of the SB:FE proposed by Sattler (1988), although goodness of fit indices suggested a less adequate fit than for the four factor model proposed by Thorndike et al. (1986b). However, Boyle (1990) suggested that the Absurdities, Copying, and Memory for Objects subtest be excluded in order to improve accuracy and shorten the test as they appeared to be contributing little to their defined Reasoning Areas.

McCallum, Karnes, and Crowell (1988) investigated the structure of the SB:FE with a sample of 60 gifted elementary school children between the ages of 9 and 12 years. They concluded that the factor structure of the test provided partial support for the rationally derived model proposed by Thorndike et al (1986b), but noted that caution may be needed in interpreting results in light of the relatively smaller sample and restricted variance due to the nature of the sample.

Spelliscy (1991) analyzed the protocols of 371 subjects tested at the University of Alberta, Education Clinic using exploratory and confirmatory factor analytic techniques. Age categorizations adopted by Spelliscy (1991) were similar to those used by Thorndike et al. (1986b). Results were compared to the factorial studies conducted by Thorndike et al. (1986b). Principal components analysis followed by varimax rotations, yielded results that tended to be most closely comparable to Sattler's (1988) findings, yet did not support the underlying SB:FE model proposed by Thorndike et al. (1986b). However, the results of Spelliscy's (1991) confirmatory factor analyses stand in contrast. Orthogonal procrustees analyses were generally supportive of the four cognitive areas of ability, particularly for the middle and older samples. The existence of a general factor was also confirmed. In addition, results of multiple group factor analyses followed by oblique transformations provided further support for the four cognitive areas across the three age groups.

Oblique hierarchical analysis was conducted by Spelliscy (1991) in order to investigate the relationship between higher order factors. The results of the hierarchical process were "complex and generally equivocal" (Spelliscy, 1991, p. 203) and it was concluded that the three level hierarchical model proposed by Thorndike et al. (1986b) was not supported by clinic data. However, Spelliscy (1991) maintained that a two tier model with a general factor at the highest level and Verbal, Abstract/Visual, Quantitative, and Short Term Memory factors at a lower level best described the SB:FE.

Congruence is noted between Spelliscy's (1991) factor analytic findings, based on a Canadian clinical sample, and Boyle (1989, 1990) and Keith et al.'s (1988a, 1988b) factor analytic findings, based on nationally representative U.S. norms. A number of advantages arise from adoption of Spelliscy's (1991) model and other research that supports the presence of the four cognitive ability areas. The model adheres to two levels of the SB:FE theoretical framework, avoids the complexities of multi-level interpretations that lack empirical support and, most significantly, is generally supported across all three age groups, and particularly strongly supported within the older two age groups.

The limited fit of the SB:FE to the theoretical model at the age level 2-6 reported by some researchers may reflect the fact that the SB:FE does not assess memory effectively at this age (Keith et al., 1988a), or it may reflect the fact that cognitive abilities are less differentiated for younger than older children (Garrett, 1965; Keith et al., 1988a). The latter may be the most acceptable interpretation, but caution may be needed interpreting these factors at this age. The WISC-R, the WPPSI, and the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R) (Wechsler, 1989) use similar subtests. Factor analyses of the WPPSI generally produces Verbal and Performance factors (Carlson & Reynolds, 1981). Factor analyses of the WISC-R typically produces three factors (Kaufman, 1975), with the third factor being interpreted as a memory factor (Freedom from Distractibility).

The Relationship Between the SB:FE and Other Measures of Cognitive Ability

Empirical investigations of the SB:FE have generally supported the relationships documented by Thorndike et al. (1986b) between the test and other measures of intelligence with a variety of populations.

A positive relationship was found between the SB:FE and Wechsler Preschool and Primary Scale of Intelligence (WPPSI) by Carvajal, Hardy, Smith, and Weaver (1988). As well, the relationship between the SB:FE and Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R) was investigated by Carvajal, Parks, Bays, Logan, Lujano, Page, and Weaver (1991) in a sample of 51 preschool, kindergarten, and 1st and 2nd grade children. A correlation of .61 was obtained between the total scores yielded by the two tests.

Rothlisberg (1987) examined the concurrent validity of the SB:FE and the WISC-R in a sample of 32 non-exceptional elementary school children. Significant correlations were obtained between the SB:FE Verbal Reasoning Area and the WISC-R Verbal Scale IQ, between the SB:FE Quantitative Reasoning Area and the WISC-R Performance Scale IQ, and between the SB:FE Composite Score and the WISC-R Full Scale IQ score. The correlations obtained were r=.70, r=.40, and r=.77, respectively. Rothlisberg (1987) concluded that the results generally supported assertions made by Thorndike et al. (1986b).

In a sample of 166 subjects with a mean age of 10.3, a correlation of .55 was found between SB:FE and WISC-R scores (Livesay, 1986). Similar scores were obtained (SB:FE M=121.47, SD=9.72; WISC-R M=123.33, SD=9.58). Lukens (1990) in a sample of 80 mentally handicapped subjects age 8-15 years obtained a correlation of .83 between total tests scores on the SB.FE and WISC-R. The WISC-R Full Scale IQ scores were significantly lower than the SB:FE Composite SAS scores and there was a median difference of four points between the two tests.

Greene, Sapp, and Chissom (1990) administered the SB:FE and WISC-R to 51 black urban children in grade 1-8 who were educably mentally retarded, learning disabled, or behaviourally disturbed. Significant correlations were obtained between the two measures. In addition, tabular comparison of precision of classification between the two tests yielded 78% agreement. However, it was noted that the SB:FE appeared most effect in assessing subjects in the educable mentally retarded range, but use of the SB:FE was not clearly supported with learning disabled subjects. Additional studies have clearly documented the relationship between the SB:FE and the WISC-R (Brown & Morgan, 1991; McCallum & Karnes, 1987; Phelps, 1989; Phelps, Bell, & Scott, 1988; Prewett & Giannuli, 1991; Robinson & Nagle, 1992). Carvajal, Hayes, Lackey, Rathke, Wiebe, & Weaver (1993) found a relationship between the SB:FE and the WISC-III (Wechsler, 1991).

Similar findings are reported with respect to the relationship between the SB:FE and the WAIS-R (Wechsler, 1982). Spruill (1991) found no significant differences between mean WAIS IQs and SB:FE SAS scores in a sample of 32 mentally retarded adults below the age of 24 years. Significant correlations were also noted between the two scales. Carvajal, Gerber, Hewes, and Weaver (1987) in a sample of 32 normal college students obtained a correlation of .91 between the WAIS-R and the SB:FE. Similar scores were obtained (SB:FE $\underline{M}$=103.5, $\underline{SD}$=10.9); WAIS-R $\underline{M}$=100.9, $\underline{SD}$=9.1).

Studies by Carvajal, Gerber, and Smith (1987), Johnson, Howie, Owen, and Baldwin (1993), Molfese, Helwig, and Holcomb (1993), and Tarnowski and Kelly (1987) have also documented the relationship between SB:FE and the Peabody Picture Vocabulary Test-Revised (PPVT-R) (Dunn & Dunn, 1981). Hodapp (1993), in a study of a sample of 42 children, ranging in age from three to six years, referred for learning difficulties, obtained a correlation of .54 between the SB:FE and PPVT-R.

Hartwig, Sapp, and Clayton (1987) in a sample of 30 normal children found that the SB:FE correlated with Stanford-Binet Form L-M ($\underline{r}$=.72) and produced similar overall scores (SB:FE $\underline{M}$=114.4, $\underline{SD}$=12.3; S-B:L-M $\underline{M}$=113.1, $\underline{SD}$=15.2). Livesay (1986) administered the SB:FE and

S-B:L-M to a sample of 120 gifted children with a mean age of 6-11. The correlation was significant (r=.64). The SB:FE yielded lower overall scores (M=122.5, SD=9.17) than the S-B:L-M (M=130.5, SD=11.2). Lukens (1987) administered the SB:FE to 32 mentally handicapped adolescents who had previously completed the S-B:L-M. A significant correlation of .86 between the two sets of test scores was obtained. Kluever and Green (1990) compared the performance of 51 gifted children on the S-B:L-M and the SB:FE and found significant correlations between the two instruments, although it was noted that SB:FE Composite SAS cutoff values to identify gifted subjects may need to be explored and modified.

The SB:FE has also been validated against the Kaufman Assessment Battery for Children (K-ABC) (Kaufman & Kaufman, 1983) Significant correlations have been obtained (Hayden, Furlong, Linnemeyer, 1988; Hendershott, Searight, Hatfield, & Rogers, 1990; Knight, Baker & Minder, 1990; Krohn & Lamp, 1989; Rothlisberg & McIntosh, 1991; Smith, St. Martin, & Lyon, 1989).

Atkinson, Bevc, Dic[illegible] and Blackwell (1992) assessed the concurrent validity of the SB:FE against the Leiter Inter[illegible]onal [illegible]erformance Scale (LIPS) (Leiter, 1948) in a sample of 24 developmentally delayed children ranging in age from 4 to 11 years. A correlation of .78 was obtained between the total scores on the two scales, although in 29% of cases significant intra-individual differences were noted between the two scores.

Meloff (1987) analyzed results of the SB:FE, administered to a sample of 153 individuals, from a Canadian clinic population. The sample was divided into four groups - - learning disabled, mentally retarded, average, and gifted. Results of the study suggested the SB:FE was able to distinguish learning disabled from mentally handicapped populations and gifted from normal populations. Meloff (1987), however, did not examine differences at given age levels.

Clarke (1988) analyzed the results from administration of the SB:FE, WISC-R (Wechsler, 1974), and the Wide Range Achievement Test-Revised (WRAT-R) (Jastak & Wilkinson, 1984) to 326 Canadian subjects ranging in age from 2 to 27 years. Clarke (1988) divided the sample into five groups which included those with learning difficulties, mentally handicapped, gifted, normally achieving, and preschool subjects below age 5-00. Clarke (1988) suggested that there were patterns of high and low scores distinct to each group. The learning difficulty group showed strengths in perceptual tasks, less developed abilities in verbal skills, and weaknesses in attentional tasks. The normal and mentally handicapped groups showed strengths in perceptual abilities. The gifted group showed strengths in verbal and conceptual abilities. Clarke (1988) also suggested accurate prediction of membership in a specific educational group was possible using SB:FE subtest and area scores as criterion. Clarke (1988) concluded that the SB:FE possessed reasonable concurrent, convergent, and discriminant ability. Caution may be needed in interpretation of Clarke's (1988) results as individuals were allocated to each educational group largely on the basis of referral data, and not on objective criteria. In addition, the learning difficulty group was treated as a homogeneous entity.

The SB:FE has been useful in research with groups presenting with learning difficulties (Hollinger & Baldwin, 1990; Knight, Baker, & Minder, 1990; Krohn & Lamp, 1989; Phelps, Bell, & Scott, 1988; Smith, St. Martin, & Lyon, 1989). However, much of this research has involved limited samples.

A number of studies have investigated the utility of the SB:FE with gifted populations (Carvajal & McNab, 1990; Hayden, Furlong, & Linnemeyer, 1988; Karnes & D'Ilio, 1987; Kitano & de Leon, 1988; Kluever & Green, 1990; McCall, Yates, Hendricks, Turner, & McNabb, 1989; Phelps, 1989; Robinson, Dale, & Landesman, 1990). In general, whilst results of these studies support the validity of the SB:FE in identifying gifted children, the results also suggest that less children are likely to be identified as gifted with the SB:FE (Kluever & Green, 1990).

Blakeslee (1987) found significant correlations between the SB:FE Composite SAS scores and the WRAT-R (Jastak & Wilkinson, 1984) in a sample of 47 emotionally disturbed children. The correlations with WRAT-R Reading, Spelling, and Arithmetic scores were .48, .37, and .70 respectively. Lewis-O'Donnell (1986) also found significant correlations between the SB:FE Composite SAS scores and the WRAT-R in a sample of 49 emotionally disturbed children. The correlations with WRAT-R Reading, Spelling, and Arithmetic scores were .51, .55, and .58 respectively. Powers, Church, and Treolar (1989), in a study of 60 regular education third grade students, found the SB:FE Composite score was the single best predictor of all the achievement subtests on the Woodcock-Johnson Tests of Achievement (Woodcock, 1977). Interestingly, Powers et al. (1989) found that the SB:FE Area and Composite scores were better predictors of academic achievement than were Sattler's (1988) factor scores. Krohn and Lamp (1990) reported that the SB:FE Composite score correlated with the Metropolitan Readiness Test (MRT) (Nurss & McGauvern, 1986). Prewett and McCaffery (1993) found significant correlations between the SB:FE and the Kaufman Test of Educational Achievement (K-TEA) (Kaufman & Kaufman, 1985) Math, Reading, and Spelling scores in a sample of 75 students, age 6-15 years. Subjects were referred for academic difficulties.

Summary of Research Evidence Related to the Validity of the Stanford-Binet: Fourth Edition

Sattler (1988) concluded the SB:FE is "potentially a powerful tool for assessment of cognitive ability of young children, adolescents, and young adults" (p. 290). It possesses excellent internal consistency reliability and adequate concurrent validity (Sattler, 1988, 1992), and has been accepted as a replacement to the previous edition. However, questions remain whether the scores derived from the scale can be used to reflect separate ability domains (Molfese et al., 1992).

Incontrovertible evidence is available suggesting that the SB:FE is a valid measure of g, or overall general ability (Glutting 1989), and that it is strongly related to major intelligence measures in current use, such as the WPPSI (Wechsler, 1967), WPPSI-R (Wechsler 1989), WISC-R (Wechsler, 1974), WISC-III (Wechsler, 1991), K-ABC (Kaufman & Kaufman, 1983), and WAIS-R (Wechsler, 1982). Laurent et al. (1992) suggest the SB:FE "appears to be at least as good a measure of g as other existing measures of intelligence" (p. 108). In addition, one of the strengths of the SB:FE is that there is a well defined and articulated theory underlying the scale (Keith et al., 1988a). A "construct of interest should be embedded in a conceptual framework, no matter how imperfect the framework may be" (CPA, 1987, p. 9). However, the evidence to suggest the SB:FE actually matches its intended theory is controversial and likely to remain so for the foreseeable future.

Factor analyses have yielded equivocal evidence about the construct validity of the instrument, although the results appear to depend on the type of factor analysis conducted. Exploratory factor analyses have been less supportive of the four factor model proposed by Thorndike et al. (1986b). Evidence contradicting the model proposed by Thorndike et al. (1986b) is available (Fritzke, 1988; Reynolds, Kamphaus, & Rosenthal, 1988; Sattler, 1988; Thorndike, 1990). However, at present, it appears more compelling evidence, particularly from confirmatory factor analysis, supporting the presence of the four cognitive ability areas, can be found (Boyle, 1989, 1990; Gridley & McIntosh, 1991; Keith et al., 1988a, 1988b; McCallum, Karnes, & Crowell, 1988; Ownby & Carmin, 1988; Spelliscy, 1991; Thorndike et al., 1986b). Little or no support can be garnered for the hierarchical structure and the Crystallized and Fluid-Analytic intelligence levels proposed by Thorndike et al. (1986b). For example, Keith et al. (1988a, 1988b) suggested that the Quantitative Reasoning Area, which Thorndike et al (1986b) proposed, along with the Verbal Reasoning Area comprised Crystallized intelligence, was more highly correlated with Abstract/Visual Reasoning factor, which was associated with Fluid-Analytic intelligence.

Sattler (1988) suggested that the SB:FE may be especially helpful in clinical and psychoeducational work. Some research has addressed this already (e.g., Hollinger & Baldwin, 1990; Knight, Baker, & Minder, 1990; Krohn & Lamp, 1989), but more is needed. Thorndike et al. (1986b) asserted that the SB:FE measured cognitive abilities that are associated with school progress. Applications of the SB:FE frequently address questions that concern performance of children in certain ability domains (Molfese et al., 1992) and this is an area that needs to be addressed in demonstrating the validity of the SB:FE as a psychometric instrument for use with special populations. Vernon (1987) noted that more research is needed with the SB:FE by Canadian psychologists. Spelliscy (1991) asserted empirical literature on the SB:FE is limited, noting that evidence for interpretation of SB:FE subscores, score differences and profiles is "at present lacking" (p. 229). However, interpretive tables for examining test scatter on the SB:FE have been published (Naglieri, 1988; Rosenthal & Kamphaus, 1988; Sattler, 1988; Thorndike et al., 1986b) suggesting interpretive practice along these dimensions is likely. Hopkins (1988) maintained, for example, that the four Reasoning Area scores should be interpreted on the basis of clinical necessity and the need to generate hypotheses about examinee performance. If one accepts that the four cognitive areas underlying the model do possess some validity, as research may suggest, then a logical step in adding to knowledge about the validity of the SB:FE is to explore the validity of these cognitive area scores, in terms of profile analysis and educational diagnosis or classification.

The SB:FE General Purpose Abbreviated Battery (GPAB)

Thorndike et al. (1986a, 1986b) made a number of recommendations about the use of abbreviated batteries (the 6-test General Purpose Abbreviated Battery, the 4-test Quick Screening Battery, and the 2-test Battery) when the research or clinical situation warrants (Delaney & Hopkins, 1987). The SB:FE GPAB is based on six core tests administered at all levels: Vocabulary; Bead Memory; Quantitative; Memory for Sentences; Pattern Analysis; and Comprehension. The Quick Screening Battery is based on four subtests that measure all four areas: Vocabulary; Bead Memory; Quantitative; and Pattern Analysis. The 2-test Battery is comprised of the Vocabulary and Pattern Analysis subtests. Delaney and Hopkins (1987) maintained that the shortened batteries possess reasonable reliability estimates and correlate highly with full batteries at the composite score level. Thorndike, Hagen, and Sattler (1986c) indicated that the GPAB can be used for placement purposes (p. 50). It is this assertion that makes it important to fully investigate the utility of the GPAB and its relationship with the SB:FE full battery.

Psychometric Properties of the SB:FE GPAB

Reliability

Thorndike et al. (1986b) indicated that for all age ranges internal consistency (KR-20) (Kuder & Richardson, 1937) reliabilities of the GPAB are satisfactory (around .95), and Composite SAS's derived from this abbreviated version correlate very highly with Composite SAS's derived from the complete battery (*r*=.94-.98). Thorndike et al. (1[illegible]) cautioned against using the 2- and 4-test abbreviated batteries for placement decisions. The internal consistency estimates for the GPAB appear in Table 4.

Table 4

Internal Consistency Reliability Estimates for the SB:FE GPAB Composite SAS's

Age	Composite SAS
2	.95
3	.96
4	.96
5	.97
6	.95
7	.95
8	.96
9	.95
10	.97
11	.97
12	.97
13	.98
14	.97
15	.98
16	.98
17	.98
18-23	.99

Adapted from Thorndike et al. (1986b)

Validity

Hoffman (1988) examined the relationship between the full battery of the SB:FE and abbreviated versions in a sample of 108 children. All correlations were significant and Hoffman (1988) suggested that the abbreviated batteries provide good estimates of full battery composite scores. DeLamatre and Hollinger (1990) extracted GPAB scores from full battery administrations given to 19 developmentally handicapped children ranging in age from 7 to 17 years. The Pearson correlation between the two test composites was .94. Short form scores were higher in 47% of cases, lower in 21% of cases, and the same in 32% of cases. The researchers concluded that the abbreviated battery could provide accurate estimates of full battery composite scores, but cautioned that the results could misrepresent scores. McCallum and Karnes (1990) compared scores from a short form of the SB:FE with scores obtained from the long version in a sample of 33 gifted children between the ages of 9 and 11 years. Means from the Abstract/Visual Reasoning, Short Term Memory, and Composite SAS's derived from the abbreviated version were significantly lower than the corresponding means of the full battery. Correlations were significant and ranged from .54 to .93. Atkinson (1991) assessed the validity of the SB:FE GPAB in a sample of 53 children between 6 and 11 years with low intelligence (SB:FE full battery Composite SAS's under 79). A correlation of .98 was obtained between the full and abbreviated composite scores. Atkinson (1991) suggested that GPAB abbreviated composite scores accurately approximated the full battery composite and thus may be of clinical utility with children with low levels of cognitive ability. Prewett (1992) evaluated three short forms of the SB:FE [the 2-, 4-, and 6-test versions proposed by Thorndike et al. (1986b)] among 150 low achieving students aged 5 through 15 years. The sample included males and females and white and black subjects. The 4- and 6-test short forms obtained the highest and most similar validity coefficients.

Carvajal and Weyand (1986) administered the SB:FE GPAB and the WISC-R to 24 Grade 3 children. A significant correlation of r=.78 was obtained between the SB:FE Composite Score and the WISC-R Full Scale Score. Carvajal and Gerber (1987) administered the Quick Screening Battery and the GPAB from the SB:FE to 32 undergraduates. The Quick Screening Battery yielded a Composite SAS M of 101.0, and the GPAB a Composite SAS M of 100.6 The researchers concluded that the GPAB was preferable as a substitute to the full version than the Quick Screening Battery. Carvajal, McVey, Sellers, Weyand, and McKnab (1987) found that scores on the SB:FE GPAB correlated significantly with scores on the Columbia Mental Maturity

Scale (CMMS) (Burgemeister, Blum, & Lorge, 1972) and the Peabody Picture Vocabulary Test-Revised (PPVT-R) (Dunn & Dunn, 1981) in a sample of 23 third grade children. In a second study, Carvajal, Hardy, Harmon, Sellers, and Holmes (1987) administered the GPAB, the PPVT-R and the CMMS to 21 kindergarten children. The correlation of the SB:FE GPAB with the PPVT-R and the CMMS was statistically significant. Carvajal et al. (1993) found a significant correlation between SB:FE GPAB composite scores and the Wechsler Intelligence Scale for Children-Third Edition (WISC-III) (Wechsler, 1991) based on a sample of 32 grade 3, 4, and 5 subjects.

Not all evidence pertaining to the GPAB has been supportive of its validity. Nagle and Bell (1993) investigated the relationship of the SB:FE GPAB and the full battery Composite SAS in a sample of 38 college students and suggested that there was a lack of comparability between the short-form and complete battery SAS's.

Results in general suggest, however, that the 6-test abbreviated version of the SB:FE has potential as a brief, valid, and reliable measure of cognitive ability. Glutting (1989) suggested that the GPAB appeared to be a reliable and valid measure, but stressed that more information is needed about the difference between estimated Composite and Reasoning Areas scores from abbreviated and full batteries. Glutting (1989) also stressed information is needed concerning the relationship of full and abbreviated batteries and external criterion. Laurent et al. (1992), in their extensive review of the validity research on the SB:FE, concluded that research suggested the abbreviated battery "provides as valid a measure of intellectual ability as the complete SB:FE" (p. 107). Its use may overcome difficulties associated with multivariate research with the full battery, as well as alleviating some the concerns identified with respect to the difference in test compositions at different ages. This needs to be explored empirically with a large sample of diverse subjects.

Educational Planning and Mild Handicaps

Intelligence tests, like the SB:FE, form an integral part of educational assessment, diagnosis, and classification (Sattler, 1988; Winzer, 1993). Shepard (1989) emphasized the controversial nature of identification and diagnosis of mild handicaps (mental retardation, learning disabilities, and emotional/behavioural disorders). The controversy relates largely to strongly held beliefs about the positive or negative consequences of identification (Shepard, 1989). Sattler (1988) maintained that "diagnosis and classification play crucial roles in the assessment process" (p. 551), lending organization to heterogeneous and complex issues associated with exceptionality. Diagnosis and classification are essential to developing and administering treatment programs and research (Sattler, 1988; Winzer, 1993).

Within the field of educational psychology, efforts need to be directed toward determining reliable and valid means of identifying academic or learning problems in order to facilitate remedial programming (Hooper & Willis, 1989; Rourke, 1991; Shepard, 1989). This is particularly important in the area of learning disabilities in order to develop sound psychoeducational interventions (Hooper & Willis, 1989; Hynd, 1989; Winzer, 1993). Research techniques must be directed toward the development of valid differential diagnostic procedures based on theoretical clinically relevant classification schema (Adelman & Taylor, 1986). On a more general level, Shepard (1989) suggested:

> the search for subtypes is essential, whether the goal is to find new biological causes for a subgroup of EMR, LD, ED, or to discover aptitude treatment interactions whereby children of a certain type benefit more from a particular strategy of instruction (pp. 569-570).

Shepard (1989) stressed this is essentially a measurement task. Thorndike et al. (1986b) presented a similar view.

Profile Analysis

Although research in the area of intelligence has focused on inter-individual differences, Binet was one of the first psychometricians to recognize that various psychological processes may arrange themselves in a different configuration within each individual (Frank, 1983). Kramer, Henning-Stout, Ullman, and Schellenberg (1987) defined this scatter as the "extent of variability of subtest scores" (p. 37). Intelligence tests have thus traditionally been subjected to profile analysis or scatter analysis. Scatter patterns are used to generate hypotheses about an individual's potential or to develop educational or clinical interventions (Sattler, 1988). Miller, Stoneburger, and Brecht (1978) maintained that differences in academic functioning should be reflected in differential WISC-R performance patterns.

Mueller et al. (1986) stated that the analysis of WISC-R subtest patterning has become increasingly important in the differential diagnosis of learning problems in children. This trend, Mueller et al. (1986) indicated, is a consequence of the fact that a number of the WISC-R features render it particularly attractive for psychoeducational diagnostic purposes. Bush and Waugh (1976) and Sacuzzo and Lewandowski (1976) pursued similar lines of reasoning. Kaufman (1980) maintained that the WISC-R possesses a number of subtests that purportedly measure a wide variety of skills and traits. Moreover, each subtest generally displays adequate and reliable specific variance (Kaufman, 1979, 1980).

Sattler (1982) cited studies that appear to give credence to the practice of profile analysis, although cautioning that WISC-R patterns cannot, without additional information, identify learning disabilities. However, a number of studies have suggested a rank ordered list of Wechsler subtest scores, and the four most difficult tests for the reading disabled individual, Arithmetic, Coding, Information, and Digit Span, are given the acronym ACID (Sattler, 1988). Rourke, Bakker, Fisk, and Strang (1983) found the ACID pattern to be a reliable discriminator of reading disabled groups. Such findings need to be viewed with caution. Mueller et al. (1986) observed that these tests appear to be routinely difficult even for normal children. Spelliscy (1991) reported that there is an abundance of literature suggesting profile analysis may be of limited value.

Terman and Merrill (1973) warned against attempts to construct psychologically meaningful profiles from instruments yielding single measures of general ability. Wechsler (1974) stated the WISC-R was designed to measure "global intelligence" (p. 1). Earlier versions of the Binet tests have most strongly measured general ability. The "utility of the Stanford-Binet L-M in educational settings has been hampered by the lack of subtest/subscale scores on which profiles can be based" (Fritzke, 1988, p. 4[illegible]). However, the SB:FE lends itself to profile analysis. The revised structure provides measures of four areas of cognitive ability, specifically designed to be of utility in educational planning (Thorndike et al., 1986a). Naglieri (1988), Rosenthal and Kamphaus

(1988), Sattler (1988), and Thorndike et al. (1986b) provided information necessary to identify statistically significant differences between subtest SAS's and between Reasoning Area SAS's.

Reynolds and Clarke (1985) supported the use of profile analysis, maintaining the procedure has traditionally been wrongly implemented and criticized. Much of the literature pertaining to profile analysis is based on the Wechsler Intelligence Scale for Children (WISC) (Wechsler, 1949) or Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974), and on the older versions of the Stanford-Binet. Research into profile analysis using earlier versions of the Binet may no longer be valid, and evidence concerning the efficacy of profile analysis with the SB:FE is needed.

The SB:FE's intended amenability to profile analysis is consistent with trends in education of moving away from instruments that tap single general intelligence quotients toward evaluation materials that assess specific areas of educational need (Molfese et al., 1992; Sattler, 1982). However, whilst there exists a vast body of research examining the utility of the WISC-R, and, to a lesser extent the WAIS-R, and subtest performance profiles, little such research has been conducted with the SB:FE. Moreover, "in spite of the massive research effort that has gone into differentiating exceptional groups on the basis of WISC-R subtest profiles, the results of the studies have been less than fully consistent" (Mueller et al., 1986, p. 22). One of the criticisms aimed at such research is a failure to include samples of normal and exceptional children. Multivariate cluster analysis of large undifferentiated samples may be the most effective means of addressing issues relating to profile analysis. In addition, the broad range of cognitive abilities tapped by the SB:FE suggests that research with the SB:FE may be more productive than that conducted with the Wechsler scales as well as previous versions of the Binet.

Learning Disabilities

Definitional Issues in Learning Disabilities

Learning disorders in children, adolescents, and adults have been recognized for several decades and represent a major educational/clinical enterprise and substantive topic of inquiry (Keogh, 1990, p. 15; Winzer, 1993). Diagnostic and definitional issues have plagued contemporary learning disability classification research (Hooper & Willis, 1989; Winzer, 1993). Kirk (1963), however, is credited with establishing the term "learning disabilities" and in defining the term carefully, exerted a significant impact on the field (Winzer, 1993). Kirk (1963, p. 3) stated:

> Recently, I have used the term "learning disabilities" to describe a group of children who have disorders in the development in language, speech, reading, and associated communication skills needed for social interaction. In this group, I do not include children who have sensory handicaps, such as blindness and deafness, because we have methods of managing and training the deaf and the blind. I also exclude from this group children who have generalized mental retardation.

Winzer (1993) noted that Kirk's definition isolated the general characteristics of the population to be subsumed under the label of learning disabilities and precipitated a vote to reorganize diverse groups within North America into the Association for Children with Learning Disabilities, which has been renamed in Canada as the Learning Disabilities Association of Canada (LDAC) and which continues to be a very powerful lobby group. Most importantly, Kirk's work served as a catalyst to stimulate interest in the field and heralded an era of unprecedented and "astonishing growth" in learning disability research, identification, and conceptualization, as well as development of a number of professional organizations within the field (Winzer, 1993, p. 243). The tremendous growth in the field of learning disabilities was accompanied by a great deal of debate and controversy over conceptualization, definition, and diagnosis which prevails to this day.

There are, however, some commonly held assumptions about learning disabilities that tend to bring consensus to the various conceptualizations or definitions (Keogh, 1990). These are briefly reviewed.

First, it is assumed that the learning disability is locused, or situated, within the individual (Keogh, 1990). The most prevalent early inferences were that learning disabilities were attributable to central nervous system dysfunction. This conceptualization has evolved with advances in neurosciences and neuropsychology (Knights & Bakker, 1976; Obrzut & Hynd, 1986; Rourke, 1985, 1991). Post mortem studies and, most recently, advances in sophisticated in vivo diagnostic imaging techniques, have supported the involvement of central nervous system dysfunction in learning disabilities (Gaddes, 1985; Galaburda & Eidelberg, 1982; Galaburda, Sherman, Rosen, Aboitz, & Geshwind, 1985; Hynd, Marshall, & Goncalez, 1991). Swanson (1988), however, working from an information processing paradigm, emphasizes functional processing problems, rather than structural ones. Regardless of differences in causal hypotheses for learning disabilities, the field tends to be dominated by exclusionary criteria or definitions (Fletcher & Morris, 1986; Keogh, 1990) that rule out putative factors such as social or economic disadvantage, lack of educational opportunities, and physical handicaps as being primary causative factors.

Second, it is assumed that individuals with learning disabilities do not function at levels consistent with their intellectual potential (Keogh, 1990). Thus, inherent in definitions of learning disabilities is the concept of a discrepancy between academic achievement and academic aptitude (Fletcher & Morris, 1986; Hooper & Willis, 1989). "Intra-individual differences can thus be considered the hallmark of learning disabilities" (Holmes, 1986, p. 304). The emphasis on discrepancies arose from attempts to operationalize learning disability diagnosis (Sattler, 1988) and is one of the indicators that separates learning disabled individuals from slow learners or retarded learners (Keogh, 1990).

A third common assumption is the notion of specificity introduced by Stanovich (1986). It is argued that learning disabled subjects exhibit unexpected deficits in "certain, but not all" areas of academic and educational functioning (Keogh, 1990, p. 16). The notion of specificity thus provides a further means of differentiating learning disabled subjects from other groups of poor achievers.

The extremely heterogeneous nature of the learning disabled population, however, defies easy categorization and makes the search for an appropriate definition challenging (Winzer, 1993). A universal definition of a learning disability has been elusive (Winzer, 1993), which in turn affects prevalence estimates, considerations of etiology, as well as confounding research. Despite the difficulties establishing a universal definition, many definitions have emerged, each with their respective merits as well as limitations. A number of these definitions are reviewed.

In Alberta, the following definition of learning disabilities was developed in consultation with Alberta Education, representatives of school jurisdictions, private schools, and the Alberta Association for Children and Adults with Learning Disabilities:

> The term "learning disability" refers to any one of a heterogeneous group of chronic disorders that may have as its basis an identifiable or inferred central nervous system dysfunction. These disorders may be manifested by difficulties in one or more processes such as attention and planning. This results in demonstrable weaknesses in language arts, mathematics, and/or social acceptance.
>
> Learning disabilities may affect anyone. [illegible]ever, if a student is underachieving relative to his or her learning potential and has no se[illegible]pairment, no motor impairment, adequate motivational and learning opportunities [illegible] adequate learning environment, then learning disabilities are considered the primary d[illegible]ing condition
>
> For the majority of students with learning disabilities, modification of the instructional process and/or the learning environment is required to meet their unique learning needs. In some instances, the use of a specific curriculum, directed to a student's need and abilities, may be required. (Alberta Education, 1986, p. 1-5, 1988, 1992)

Hammill (1990) argued, perhaps optimistically, that there is an emerging consensus in the area of learning disability definition. The definition of learning disabilities adopted in Alberta (Alberta Education, 1986, 1988, 1992) shares common features with four definitions that Hammill (1990) suggested "are professionally viable" (p. 82). These are definitions proposed by: the Interagency Committee on Learning Disabilities (ICLD) (1987); the Learning Disabilities Association of America (LDA) (1986); the National Joint Committee on Learning Disabilities (NJCLD) (Hammill, Leigh, McNutt, & Larsen, 1981; NJCLD, 1988); and the United States Office of Education (USOE) (1977), and are presented below.

<u>The Interagency Committee on Learning Disabilities Definition</u>

> Learning disabilities is a generic term that refers to heterogeneous groups of disorders manifested by significant difficulties in the acquisition and use of listening, speaking, reading, writing, reasoning, or mathematical abilities, or of social skills. These disorders are intrinsic to the individual and presumed to be due to central nervous system dysfunction. Even though a learning disability may occur concomitantly with other handicapping conditions (e.g., sensory impairment, mental retardation, social and emotional disturbance), with socioenvironmental influences (e.g., cultural differences, insufficient or inappropriate instruction, pyschogenic factors), and especially attention deficit disorder, all of which may cause learning problems, a learning disability is not the direct cause of those conditions or influences. (ICLD, 1987, p. 222)

<u>The Learning Disabilities Association of America Definition</u>

> Specific Learning Disabilities is a chronic condition of presumed neurological origin which selectively interferes with the development, integration, and/or demonstration of verbal and/or nonverbal abilities. Specific Learning Disabilities exist as a distinct handicapping

condition and varies in its manifestations and in degree of severity. Throughout life, the condition can effect self-esteem, education, vocation, socialization, and/or daily living activities. (LDA, 1986, p. 15)

The 1977 U. S. Office of Education Definition

The term "specific learning disability" means a disorder in one or more of the basic logical processes involved in understanding or in using language, spoken or written, which may manifest itself in an imperfect ability to listen, speak, write, spell, or to do mathematical calculations. The term includes such conditions as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. The term does not include children who have learning disabilities which are primarily the result of visual hearing or motor handicaps, or mental retardation, or emotional disturbance, or of environmental, cultural, or economic disadvantage. (USOE, 1977, p. 65083)

The National Joint Committee on Learning Disabilities Definition

Learning disabilities is a general term that refers to a heterogeneous group of disorders manifested by significant difficulties in the acquisition and use of listening, speaking, reading, writing, reasoning, or mathematical abilities. These disorders are intrinsic to the individual, presumed to be due to central nervous system dysfunction, and may occur across the life span. Problems in self-regulatory behaviors, social perception, and social interaction may exist with learning disabilities, but do not by themselves constitute a learning disability. Although learning disabilities may occur concomitantly with other handicapping conditions (for example, sensory impairment, mental retardation, serious emotional disturbance) or with extrinsic influences (such as cultural differences, insufficient or inappropriate instruction), they are not the result of these conditions or influences. (NJCLD, 1988, p. 1)

Of these four definitions, Hammill (1990) maintained the NJCLD definition is most likely the one to gain widespread acceptance. Winzer (1993) concurred. Hammill (1990) also stressed that current definitions cannot be considered perfect, but subject to revision and improvement, and maintained that there was an evolving consensus on the definition of learning disabilities. The NJCLD (1988) definition defines the syndrome of learning disabilities, proposes etiologies and explains developmental consequences (Winzer, 1993).

In Canada, the Learning Disabilities Association of Canada (LDAC) (1981) stated:

Learning disabilities is a generic term that refers to a heterogeneous group of disorders due to identifiable or inferred central nervous system dysfunction. Such disorders may be manifested by delays in early development and/or difficulties in any of the following areas: attention, memory, reasoning, co-ordination, communicating, reading, writing, spelling, calculation, social competence, and emotional interaction.

Learning disabilities are intrinsic to the individual and may affect learning and behaviour in any individual, including those with potentially average or above average intelligence.

Learning disabilities are not due primarily to visual, hearing or motor handicaps; to mental retardation, educational or environmental disadvantage; although these may occur concurrently with any of these. Learning disabilities may arise from genetic variation, biochemical factors, events in the pre- to post-natal period, or any other subsequent event resulting in neurological impairment.

As can be seen, the definitions cited are similar in many respects. Contemporary definitions of learning disabilities contain many components of the common assumptions identified by Keogh (1990) and, in addition, mirror Kirk's original conceptualization (Kirk, 1962, 1963), although differences in definitions tend to reflect differing philosophical bases which have variously focused on aspects such as developmental problems, neurological dysfunctions, psychological processes, language difficulties, task/environmental factors, and social and emotional issues. One difference relates to the emphasis placed on learning disabilities being directly related to problems in school learning as opposed to social, emotional, and motor difficulties. The LDAC definition, for example, allows difficulties in social competence and emotional interaction to be considered a learning disability. The NJCLD definition expressly does not. Other definitions expressly state that at least average learning potential is required for diagnosis of a learning disability, while in a few definitions, it is maintained that students having learning disabilities are found at all ability levels. Given the many differences among definitions of learning disabilities, it is important to reinforce the common elements in the majority of definitions most relevant to this study: assumed neurological dysfunction; irregular or uneven development; difficulty in academic or learning tasks; discrepancy between potential and performance; exclusion of other causes; and average or above average intelligence (Winzer, 1993).

As Sattler (1992) pointed out, current definitions of learning disabilities are difficult to operationalize. Central to all notions of learning disability, is the belief that an individual with a learning disability is an underachiever (Hammill, 1990; Winzer, 1993), either reflected in intra-individual ability differences, or in the presence of aptitude-achievement discrepancies. The latter, is according to Hammill (1990), a special application of the intra-individual ability approach and this conception of discrepancy has become firmly entrenched in methods for identifying learning disabilities (Sattler, 1988). The indicator most frequently used to identify learning disabilities is an ability-achievement discrepancy (Sattler, 1992, Winzer, 1993) and thus one of the first steps in the formal diagnosis of a learning disability is assessment of intellectual potential to rule out the possibility of mental retardation and establish the fact of normal intelligence (Winzer, 1993). Once this is judged normal, academic performance is assessed (Winzer, 1993), with a view to establishing academic underachievement or a discrepancy between ability and achievement.

Numerous formulas have been proposed to define a severe discrepancy or deficit (Sattler, 1988). According to Sattler (1992), academic underachievement has been operationally defined in terms of a deviation from grade or age level, an expectancy formula, a difference between standard scores, or a regression equation. Underachievement is most simply defined as a discrepancy between the child's grade equivalent score on an achievement test and his or her grade placement. Definitions with a constant deviation specify a value that must be present in order for the discrepancy to be severe. This approach is particularly flawed. Grade scores have few acceptable psychometric properties (Sattler, 1988). In addition, constant deviation criteria may mean something different at different grade levels because the dispersion of achievement test scores increases with advancing age (Spreen, 1976). Grade equivalent scores are based on the assumption that a constant rate of learning occurs across the entire school year (Sattler, 1988). The determination of the constant deviation criteria is also extremely subjective (Francis et al., 1991).

Expectancy formulas which calculate a child's expected grade placement based on his or her mental and chronological age can also be used in the determination of discrepancies. Sattler (1988) criticized this approach for relying on the outdated notion of mental age, and for its assumption of a perfect correlation between ability tests and achievement tests. Age based discrepancies represent unstandardized metrics that vary with age (Fletcher & Morris, 1986; Reynolds, 1984).

The difference between standard scores procedure involves comparison of standard scores on two tests, typically tests of cognitive ability (aptitude) and academic achievement. A criterion score for a meaningful discrepancy may be set, for example, at one standard deviation. Problems associated with this procedure relate to the subjectivity inherent in establishing criterion scores, failure to take into account the regression of IQ on achievement, and the assumption of a perfect correlation between ability and achievement. Phillips and Clarizio (1988) maintained standard scores possess many of the flaws that grade equivalent scores do, and may lead to unusual growth patterns in academic achievement. In addition, each test used for decision making must be based on the same standard score distributions (Sattler, 1988). Further, inaccuracies may arise from sampling differences in test norms.

Generally, these procedures have been criticized on methodological, statistical, psychometric, and logical grounds (Shepard, 1980, 1983, 1989). Failure to take into account regression artifacts leads to an overidentification of higher ability children as learning disabled and underidentification of lower ability children (Reynolds, 1984). The regression method of discrepancy determination has been cited as the least subject to error and misinterpretation (Bennett & Clarizio, 1988; Hooper & Willis, 1989; Reynolds, 1981, 1984; Sattler, 1988). This method uses a regression equation to determine expected scores. The regression equation takes into account possible regression-to-the-mean effects which occur when the correlation between two measures is less than perfect, as well as the standard error of measurement of the difference score (Sattler, 1988).

Bennett and Clarizio (1988) compared four methods used to determine severe discrepancies between ability and achievement. The methods included: z-score discrepancies, estimated true score discrepancies, unadjusted regression procedures, and adjusted regression procedures. The researchers concluded that the regression procedures were the most statistically sound approach to discrepancy determination (Bennett & Clarizio, 1988, p. 368). The adjusted regression formula takes into account regression to the mean and corrects for attenuation for less than perfectly reliable tests.

Reynolds (1981, 1984) provided a most comprehensive treatment of discrepancy determination (Shepard, 1989). Reynolds (1981, 1984) recommended the simple prediction or regression model be used to compute discrepancy scores. Achievement (the Y variable) is regressed on IQ (the X variable). The simple difference between X and Y is not found, but the discrepancy is between observed achievement and expected achievement ($\hat{Y}$), based on X. A more refined model takes into account the reliability of $Y-\hat{Y}$. However, Shepard (1989) noted that as long as relatively extreme cases are being sought, the discrepancy will automatically be large enough to be considered reliable. Shepard (1989) suggested the only serious alternative to the simple prediction model has been the true-score regression model. This model takes into account errors of measurement in both X and Y, but does not take the regression between achievement and IQ into account and is thus conceptually flawed.

Cronbach and Furby (1970) developed a formula to estimate true residual change or true residual difference that corrects for measurement error and regression to the mean. Shepard (1989) does not recommend using Cronbach and Furby's (1970) formula, given its complexity and the probable lack of availability of appropriate reliability coefficients. As well, Shepard (1989) maintained that the true score prediction model and the simple prediction model will identify the same cases, because the scores and slope are adjusted by the same linear transformation (p. 561).

Learning Disabilities in Historical Perspective

As noted, Kirk (1962, 1963) is generally recognized as the person who introduced the term "learning disability" (Hynd, 1988; Winzer, 1993), although the study of children with learning problems can be traced to Kussmaul (1877). However, until 1963, the learning disability field tended to focus on the etiology of specific learning disorders. Neurologist and physicians explored the causes and consequences of learning problems, while educators and psychologists applied many of the theories to develop diagnostic and remedial programs (Winzer, 1993). In 1963, after the term "learning disabilities" was created by Kirk, a phase of integration and development of the definitions of learning disabilities began, with an emphasis on educational rather than medical orientations (Winzer, 1993).

Early work in the area of learning disabilities was based on case studies and focused on single-factor explanations for learning problems (Hooper & Willis, 1989). Hooper and Willis (1989) discussed five of the most prominent single-factor explanations. These include: Delayed cerebral dominance (Bakker, 1973, 1979a, 1979b; Orton, 1928, 1937; Satz, Rardin, & Ross, 1971); visual perceptual deficits (Bender, 1956, 1957; Frostig, 1964; Hermann, 1959; Kephart, 1971; Lyle, 1969; Lyle & Goyen, 1968, 1975); auditory-perceptual deficits (Clarke, 1970; de Hirsch, Jansky, & Langford, 1966; Denckla & Rudel, 1976a, 1976b; Goldberg & Schiffman, 1972; Henry 1975; Silver, 1968; Valtin 1973; Vellutino, 1979; Wepman, 1960); intersensory deficits (Birch & Belmont, 1964, 1965; Senf, 1969); and attention or memory deficits (Cohen & Netley, 1978, 1981; Dykman, Ackerman, Clements, & Peters, 1971; Ross, 1976; Thomson & Wilsher, 1978).

Single-factor models could not explain the full range of learning problems. They addressed specific dysfunctional components in the learning process, but were unable to account for all behavioural variance within the learning disabled populations. However, they provided the foundations for multidimensional investigations of learning difficulties (Hooper & Willis, 1989; van der Vlugt, 1991) that are closely linked with comprehensive psychoeducational assessment of learning problems. They served to generate a large number of variables that appeared capable of differentiating between normal controls and learning disabled children, but caused "considerable controversy concerning underlying cause(s) of learning problems" (van der Vlugt, 1991, p. 140). The conceptualization of learning disabilities as a heterogeneous classification is becoming increasingly accepted (Adelman & Taylor, 1986; Rourke, 1983, 1985, 1991; Winzer, 1993). Concomitant with this acceptance, a wide variety of approaches have been used for conducting classification research on disabled learners (Fletcher & Morris, 1986).

Skinner's (1981) Theoretical Framework for Classification Research

Skinner's (1981) theoretical framework for conceptualizing and evaluating attempts to classify psychiatric diagnoses is important to this study. Skinner (1981) identified three components. The first, theory formulation, involves decisions concerning variables and models used to depict subtypes. The second, internal validation, relates to the reliability, coverage, and replicability of the subtypes. The third, external validation, refers to the issue of whether subtypes are truly distinct through systematic comparisons of subgroups on variables not used to derive the typology or subgroups (Fletcher, 1985; van der Vlugt, 1991). Such validity must be demonstrated on variables outside the original measurement domain, i.e., involving different modalities or processes (Fletcher, 1985). Thus, external validation with achievement measures becomes important in assessment of subtypes derived from cognitive measures.

Two of the major problems facing research in the area of learning disabilities relate to the lack of a universally accepted definition of learning disability and sampling difficulties (Algozzine & Ysseldyke, 1986). Epps, Ysseldyke, and Algozzine (1983, 1985) maintained that any child may be diagnosed as learning disabled, given the widely variant definitions extant. Many studies rely upon subjects obtained from school identified populations. Moreover, Shepard, Smith, and Vojir (1983) noted that many children actually identified and served in school based learning disability placements may not be learning disabled. Epps et al. (1983, 1985) provided similar evidence. Hooper and Willis (1989) stated that research is unable to assure that the characteristics of samples used in one study are the same as those of samples used in other studies.

One approach to the difficult with respect to sampling is to use large undifferentiated groups of children and then empirically determine the subgroups that can be identified as learning disabled. Johnston, Fennell, and Satz (1987), Morris, Blashfield, and Satz (1986), Satz and Morris (1981), and van der Vlugt and Satz (1985) adopted this approach.

Learning Disability Subtyping

Rejection of single-factor models of learning disabilities in favour of multidimensional conceptualizations has led to number of learning disability subtype models. Hooper and Willis (1989) and Fletcher (1985) identified two major classification models. Within these two major models, different classification criteria such as academic, intellectual, cognitive, neurocognitive, and neurolinguistic are then adopted to delineate subtypes (Fletcher, 1985; Hooper & Willis, 1989). Combined models utilizing a range of criteria also exist.

The first major model is the clinical-inferential model. This model represented attempts to group individuals into homogeneous clusters by identifying similarities in performance profiles, largely through the application of clinical judgment or statistical discrepancy criteria. "These are largely post hoc models in which the investigator typically uses measures of achievement or cognition as a basis for group separation" (Hooper & Willis, 1989, p. 41). Hooper and Willis (1989) noted clinical-inferential models have become less popular in recent years, but were "seminal in the subtyping literature" (p. 41) and provided a "foundation for the acceptance of the concept of heterogeneity of learning disabilities" (p. 60). While clinical inferential models make "intuitive sense" they are prone to "methodological weaknesses, limited data reduction strategies, and questionable validity" (Hooper & Willis, 1989, p. 62).

The second model is the empirical classification or multivariate model (Fletcher, 1985; Hooper & Willis, 1989). This approach relies upon statistical procedures such as Q type factor analysis (Burt & Stephenson, 1939; Stephenson, 1935) or cluster analysis (Sneath, 1957; Sokal & Sneath, 1963), which are used to empirically group subjects on the basis of profile similarities. Empirical classification models address a number of the problems that arise from use of clinical-inferential models, but not without creating others. Empirical techniques are intended to organize data into sets, but can even organize random numbers into relatively homogeneous groups. Such procedures require the addition of sound clinical judgment and decision making (Hooper & Willis, 1989) and are particularly prone to "naive empiricism" (Aldenderfer & Blashfield, 1984).

There are advantages and disadvantages to both approaches. Fletcher and Satz (1985) argued in favour of the use of empirical approaches and maintained that clinical inferential-models are inconsistent with the objective of classification research which is to identify naturally occurring subtypes. Willis (1988) maintained that integration of the two methods holds the most promise, with the aim of evaluating empirical approaches in terms of objective clinical judgment. To this end, a major goal of this proposed study will be to validate empirically derived subtypes with clinically derived subtypes in investigating the validity of the SB:FE GPAB in educational diagnosis.

Clinical Classification Models

Clinical Classification Models Using Academic Achievement Criteria

Working within a clinical-inferential framework, a series of studies, often conducted with Canadian subjects, by Rourke (1975, 1978, 1982, 1983, 1985, 1989), Rourke and Finlayson (1978), Rourke, Fiske, and Strang (1986), Rourke and Fuerst (1991), and Rourke and Strang

(1978) found differences on a wide variety of neuropsychological tests measuring language, visual spatial, and tactile-perceptual skills according to patterns on the Reading, Spelling, and Arithmetic subtests of the Wide Range Achievement Test (WRAT) (Jastak & Jastak, 1965, 1978) and WRAT-R (Jastak & Wilkinson, 1984).

In early studies, Rourke and Strang (1978) classified learning disabled children on the basis of WRAT scores. Three groups of learning disabled students were formed. Subjects in the first group evidenced uniformly deficient performance on all three measures (reading-, spelling-, and arithmetic-disabled group). A second group was formed with subjects disabled in reading- and spelling with better performance in arithmetic (reading-, spelling-disabled group). A third group showed average or better reading and spelling scores, but were impaired in arithmetic (arithmetic-disabled group). These groups were compared on a variety of neuropsychological tests measuring language, visual spatial, motor, and tactile-perceptual skills. The arithmetic disabled group displayed well developed verbal skills, but poorer visual-spatial, psychomotor and tactile-perceptual skills. The reading- and spelling-disabled group had a variety of problems associated with auditory linguistic skills, but better visual spatial skills. The reading-, spelling-, and arithmetic-disabled group had predominant problems processing verbal material. In addition, the groups displayed significant differences in terms of WISC-R Verbal (VIQ) and Performance (PIQ) scale scores, with the arithmetic disabled groups scoring significantly lower than the reading- and spelling-disabled group on the WISC-R PIQ. Moreover, the reading- and spelling-disabled group obtained lower WISC-R VIQ's than the arithmetic disabled group.

Related studies by Rourke and Finlayson (1978), Rourke, Fisk, and Strang (1986), Rourke and Strang (1978), and Strang and Rourke (1983) supported these typologies. Fletcher (1985) indicated that the findings have been replicated using different types of measures (Fletcher, 1983; Siegel & Linder, 1984). Breen (1986) replicated these subtypes using the Woodcock-Johnson Psycho-Educational Battery. Lennox and Siegel (1988, 1993) and Nolan, Hammeke, and Barkley (1983) provided additional support for the model. The relationship between reading/arithmetic patterns and tasks emphasizing verbal/nonverbal processing has been demonstrated repeatedly (Badian, 1983; Fletcher, 1985; Francis et al., 1991; Ozols & Rourke, 1991).

Boder (1970, 1971, 1973), relying on clinical observations focused on reading subtypes, proposed three dyslexic reading patterns: dysphonetic dyslexics; dyseidetic dyslexics; and alexics. It was proposed that dysphonetic dyslexics possessed auditory-linguistic deficits, with relatively intact visual spatial abilities. Dyseidetic dyslexics displayed deficits in visual-spatial processing with intact auditory linguistic functions. Alexics displayed deficits in both auditory-linguistic and visual spatial abilities. The Boder Test of Reading Spelling Patterns (Boder & Jarricho, 1982) was developed to classify children in a systematic manner along these lines. Hooper and Willis (1989) noted results supporting Boder's findings are mixed. Distinctions among subtypes often blur. Differentiation between subtypes has not been achieved on a variety of neurophysiological, neuropsychological, cognitive, and educational tasks. The model may thus have questionable utility (Hooper & Willis, 1989). Whilst similar clinical conceptualizations have been derived (Ingram, Mason, & Blackburn, 1970; Mitterer, 1982; Thomson, 1982), Hooper and Willis (1989) maintained that the model needed refinement "if it is to prove truly useful in classification" (p. 47). Conte, Samuels, and Zirk (1983) questioned the use of Boder's classifications as a means of subtyping, and suggested that the classification system may be sensitive to auditory-linguistic (dysphonetic) processing deficits, and not to visual-spatial (dyseidetic) processing deficits, therefore providing information on only one type of disabled reader. They also reported age related differences and misclassifications using Boder's system. Moreover, Boder's conceptualization is limited in terms of its applicability to different areas of academic functioning.

Denckla (1979) identified six reading disability subtypes based on clinical observation. These groups included: a global-mixed language disorder type; an articulation-graphomotor type; an anomic-repetition disorder; a dysphonetic sequencing disordered group; a verbal learning and memorization disordered type; and a correlational disorder group with relatively normal reading skills which were significantly lower than measured cognitive abilities.

Spelling disability subtypes have typically been identified in relation to reading patterns, although Carpenter and Miller (1982), Frauentreim (1978), Naidoo (1972), Nelson and Warrington (1978), and Sweeney and Rourke (1978) identified distinct patterns of spelling disability or dyspraxia that are useful for classification purposes. Arithmetic disabled subjects, identified by Rourke (1989) have also been the subject of considerable research as they appear to be at risk for development of socio-emotional disturbance (Rourke & Fuerst, 1991).

Academic subtyping approaches have proved to be successful in revealing differences in learning disabled children (Ozols & Rourke, 1991). Fletcher (1985) concurred stating:

> The classification of children according to patterns of academic deficiency has not always been conceptualized as classification research. However, many studies support the external validity of separating disabled learners according to differences in reading, spelling and arithmetic performance. Regardless of whether children are classified according to academic performance or processing deficiencies, virtually every study has identified at least one verbal subtype, one spatial subtype, and a mixed verbal spatial subtype. (p. 192)

Brodie (1988), Lyon (1983, 1985), and Satz and Morris (1981) in reviewing the literature pertaining to subtyping, agreed.

Clinical Classification Models Utilizing Intellectual Variables

Kinsbourne and Warrington (1963), in one of the first attempts to group learning disabled children in a homogeneous fashion using clinical inferential techniques, identified two clinical subtypes based on WISC (Wechsler, 1949) Verbal IQ (VIQ) and Performance IQ (PIQ) scores in a group of children referred due to reading difficulties. Subjects in the first group evidenced a VIQ-PIQ discrepancy of at least 20 points in favor of the PIQ, while subjects in the second group evidenced the opposite IQ discrepancy of 20 points. Members of the former group displayed deficits in speech acquisition, and receptive and expressive language, while members of the latter group displayed signs of finger agnosia, left right confusion, constructional difficulties, and arithmetic deficits. The former group was labelled language retarded, and the latter the group labelled the Gerstmann group.

Quiros (1964) identified two groups which were termed central auditory processing dyslexics and visual-perceptual dyslexics. Bannatyne (1966) identified two basic patterns based on neurological functioning and genetics. The genetic dyslexia group displayed deficits in auditory processing with relatively lower verbal abilities, while the second minimal neurological dyslexia group displayed deficits in all processing domains. Johnson and Myklebust (1967) identified two learning disabled groups: audiophonic dyslexics and visuospatial dyslexics. Bateman (1968), focusing on short term memory, classified children into three groups: one with auditory deficits; one with visual deficits; and one with a combination of visual and auditory deficits. Smith, Coleman, Dolecki, and Davis (1977) identified two subtypes of learning disabled based on the WISC. The first group had high IQ's and the second had low IQ's. Bakker (1973, 1979a, 1979b), working from a developmental perspective based on lateralized brain deficiencies, identified two subtypes of reading disabled individuals, focusing on hemispheric functioning. The first group was termed P-type dyslexics and the second L-type dyslexics. The former group displayed perceptual deficits, while the latter displayed linguistic deficits. Pirozzolo (1979) identified an auditory-linguistic and a visual-spatial subtype using the WISC-R. The auditory linguistic subtype displayed low VIQ's relative to PIQ's, while the visual-spatial subtype evidenced opposite patterns.

Smith (1970) separated retarded readers into three subtypes on the basis of WISC profiles. The first subtype displayed poor auditory-sequential skills, the second subtype displayed poor visual-spatial skills, while the third group displayed a profile of mixed deficits. Rourke, Young, and Flewelling (1971) separated a group of disabled learners into three groups based on WISC profiles. The first subtype had VIQ's > PIQ's, the second subtype had VIQ's < PIQ's, while the third group had VIQ's = PIQ's. These subtypes were validated through external criteria, the Halstead Reitan Battery (HNRB) (Halstead, 1947; Reitan & Davidson, 1974).

Empirical Classification Models

Empirical Classification Procedures Using Achievement Measures

Few studies have been conducted using empirical classification techniques with achievement data. Satz and Morris (1981) used hierarchical agglomerative cluster analysis of the WRAT to identify nine achievement clusters in a sample of 236 white males. Seven of the clusters represented variants of normal profiles. Two clusters displayed significantly low achievement (one in reading and one in arithmetic). The individuals in these two groups could be diagnosed as having learning problems. Moreover, the two groups displayed significantly poorer performance on the Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 1959) and on a number of neuropsychological tests. The methods were replicated by van der Vlugt and Satz (1985) in a Dutch sample with similar results. Johnston, Fennell, and Satz (1987) identified six stable subtypes, four normal learning patterns and two evidencing distinct achievement deficits on the basis of WRAT (Jastak & Jastak, 1978) scores. Deficits in cognitive and neuropsychological variables were also noted.

DeLuca, Del Dotto, and Rourke (1987) used clustering techniques to group children with arithmetic deficiencies into four homogeneous subtypes. All children obtained WISC-R Full Scale IQ scores that fell within the average range. The first group displayed deficits in tactile-perception, conceptual flexibility, and expressive language. Weaknesses on tasks measuring attention, auditory memory, and visual-spatial output were noted. The second subtype displayed deficits in higher order tactile perception, visual motor speed and coordination, verbal fluency, and verbal memory. The third subtype evidenced deficits in nonverbal problem solving and processing, psychomotor skills, conceptual flexibility, and finger localization. The fourth subtype evidenced deficits in verbal expression, conceptual flexibility, and manipulation of visual symbolic materials.

Empirical Classification Procedures Using Intellectual Measures

Empirical classification techniques using neurocognitive variables often use standard intellectual assessment measures in identification of learning disabilities. A vast number of studies have used the WISC-R (Wechsler, 1974) in an empirical manner to identify subtypes of learning disabled (Hooper & Willis, 1989). Vance, Wallbrown, and Blaha (1978) were some of the first researchers to use the WISC-R exclusively in classification of learning disabled subjects. Q-type factor analysis was conducted on 158 elementary school children identified as reading disabled. Five subtypes were identified. Three of the subgroups evidenced impairments in language functioning, the fourth group displayed difficulties with attention, and the fifth displayed deficits in visual-perceptual abilities.

Hale and Sax (1983) used modal profile analysis (Skinner & Lei, 1980) with WISC-R data to classify a heterogeneous group of learning disabled subjects. Four subtypes of problem learners were identified. Of note, one group was characterized by relatively intact visual spatial abilities with impaired sequential processing skills.

Snow, Cohen, and Holliman (1985) derived WISC-R factor scores from a sample of 106 learning disabled children, which were then used in a hierarchical group cluster analysis. A six group solution was found to be the most appropriate. Group I demonstrated relative weakness across all factors, particularly in perceptual-organizational and attention abilities. Group 2 was slightly low on verbal comprehension, but high on perceptual organizational abilities. Group 3 was high on both verbal comprehension and perceptual organizational factors. Group 4 demonstrated impaired performance on verbal comprehension. Group 5 evidenced evenly developed abilities, whilst Group 6 was high in verbal comprehension and attention abilities. An important group to emerge was that of Group 1, which was conceptualized as a borderline IQ, or slow learner group. Results indicated that subgroups of learning disabled can be identified on the basis of WISC-R performances. A number of the subgroups appear to display similar profiles to learning disabled subgroups identified in previous research studies. The absence of a group with patterns of visual spatial deficits may have been a reflection of the limited sample used in the study.

Snow, Koller, and Roberts (1987) administered the WAIS-R (Wechsler, 1981) to 115 learning disabled adolescents and adults. Ward's minimum variance cluster analysis was then applied to the derived factor scores. A seven group solution appeared optimal. Group 1 was characterized by low verbal-comprehension and attention abilities, relative to perceptual-organizational abilities. Group 3 evidenced deficits in verbal-comprehension abilities, with higher perceptual-organizational skills and attention abilities. Group 5 appeared to have no deficits with relative strengths in attention abilities. Groups 2, 6, and 7 appeared to have evenly developed abilities. Snow et al. (1987) identified similarities between profiles obtained in this study and a previous study using the WISC-R (Snow et al., 1985). No subgroups emerged with deficits primarily in the area of visual spatial or perceptual-organizational skills. The researchers suggest that the WISC-R and WAIS-R may thus lack sensitivity in these particular areas. However, limitations in sampling may be responsible as Vance et al. (1978) identified a subtype with deficits in visual spatial abilities among a group of reading disabled individuals using the WISC-R.

A number of studies applied Q type factor analysis (Burt & Stephenson, 1939; Stephenson, 1935) to a combination of classification variables. Doehring and Hoshko (1977) administered 31 reading related tasks measuring visual matching, auditory visual matching, visual scanning, and oral reading to groups of reading disabled, learning disabled, language disordered, and mentally retarded. Subtypes with linguistic deficits, phonological deficits and intersensory integration deficits were identified. Doehring, Hoshko, and Bryans (1979) included a sample of normal readers and replicated the study. Petrauskas and Rourke (1979) applied a number of cognitive, motor and tactile skills tests to identify three replicable subtypes. Snow and Hynd (1985a, 1985b) identified subtypes with expressive and receptive deficits, reading and spelling deficits, and tactile perceptual problems. Studies like these tended to provide support for the hypothesis of multiple causes for specific broad-band subtypes of disabled children (Hooper & Willis, 1989).

Empirical subtyping models developed by Del Dotto and Rourke (1985), Fisk and Rourke (1979), Johnston, Fennell, and Satz (1987), Joschko and Rourke (1985), Lyon, Stewart, and Freedman (1982), Lyon and Watson (1981), Morris, Blashfield, and Satz (1986), Petrauskas and Rourke (1979), Satz and Morris (1981), and van der Vlugt and Satz (1985) have contributed significantly to neurocognitive subtyping efforts. Models derived from these studies have created a theoretical base for subtyping and have laid foundations for brain-behaviour relationships within the framework of multidimensional approaches to learning disabilities. Results of these studies have generally led to subtyping models comprising two to five different subtypes reflecting various degrees and combinations of auditory-linguistic processing, visual-spatial processing, and academic dysfunctions. From a theoretical and methodological perspective, these studies evidenced a commitment to the need to demonstrate the external validity of the subtypes derived and also provide tentative findings pertaining to developmental trends in subtyping. However, as Satz and Morris (1981) cautioned, contemporary classification models are preliminary and clearly require refinement. Application of empirical clustering techniques to the SB:FE may contribute to this refinement.

Comment About the Validity of Subtyping Research

Subtyping research does have limitations. Hooper and Willis (1989) were critical of models or subtyping approaches derived from the Wechsler scales, maintaining that the models essentially reflect the factor structure of the measurement instrument and this has implications for the external validity of the models. Such a point of view reflects the classification model proposed by Skinner (1981) introduced earlier, and reinforces the need to address the issue of external validity. Satz and Morris (1981) identified a need for normal populations to be included in studies in addition to special populations. Sp[illegible]990) noted "sample definition problems in learning disabilities are legion" (p. 203). Spee[illegible]ticized cluster analytic investigations for relying on purely clinical populations, noti[illegible]ception to this are the studies carried out by Satz and Morris (1981) and Morris, Blashfield, and Satz (1986). Most subtype studies have clustered only learning disabled children and have used performance of normals as baseline data (Speece, 1990).

Meta-analytic research (Glass, 1976, 1977), conducted by Mueller, Dash, Matheson, and Short (1984) and Mueller et al. (1986) has important ramifications for the literature pertaining to subtyping on the basis of intelligence test scores. Mueller et al. (1984) completed a meta-analysis of 29 samples of WISC-R subtest performance patterns of below average, average, and above average children. Ten subtests were included, with the Mazes and Digit Span subtest excluded. Standard literature search procedures identified studies using the WISC-R with the target populations. Scaled score data obtained from the identified studies were subjected to a K-means (Dixon, 1981) clustering procedures. Euclidean distance was the similarity measure used. A three cluster solution was considered the most optimal, resulting in distinct WISC-R subtest profiles emerging for the three groups. Multivariate Analysis of Variance procedures indicated a significant main effect for profile elevation and parallelism. Mueller et al. (1984) concluded that WISC-R subtest performances are strongly related to overall intellectual level. In addition, they noted that it is possible that the profile differences reflected intercorrelations among subtests.

In a related study, Mueller et al. (1986) conducted a meta-analytic exploration of WISC-R factor scores as a function of educational diagnosis and intellectual level using multivariate clustering and profile analysis procedures. The studies included in the meta-analysis comprised 119 samples of children with various psychoeducational diagnoses (unimpaired; learning disabled; clinic referred; mentally handicapped; emotionally disturbed; behaviour disordered, and other impaired), intellectual levels (based on the six intellectual level categories (Wechsler, 1974), and demographic backgrounds. Euclidean distance was used as the similarity measure. K-means cluster analysis (Dixon, 1981) was performed on Kaufman recategorized factor scores (Kaufman, 1975). A three cluster solution was considered optimal. Cross-tabulation of cluster membership by diagnostic and intellectual codes indicated that group membership, although determined by profile similarity, was more related to sample mean Full Scale IQ than to the diagnostic classification.

Mueller et al. (1986) conceded that "distinct WISC R profiles do indeed appear to exist", but noted that these are "strongly associated with the individual's global level of intellect" (p. 37). They cautioned against using the WISC-R as a diagnostic test, except to determine global intellect. Mueller et al. (1986) tended to be critical of attempts to classify individuals on the basis of group profiles, but suggested that multivariate exploratory data analysis may be of use in profile analysis with the WISC-R, particularly when used in conjunction with meta-analysis. They also stressed the need to include samples of normal children in studies that investigate intellectual profiles.

Meta-analytic research is not without weaknesses. Meta-analysis relies heavily on published articles (Kraemer & Andrew, 1982; Slavin, 1983, 1984; Wolf, 1986). Use of non-parametric statistics, which make no assumptions about the underlying normality of distributions of effects sizes, has also been advocated in meta-analysis (Kraemer & Andrew, 1982; Slavin, 1983, 1984).

Meta-analytic studies can be uninterpretable because results of poorly designed studies are addressed along with results of better studies (Glass, McGaw, & Smith, 1981). The meta-analytic studies discussed in this brief review, treated learning disabled children as a homogeneous group, which may be untenable in light of evidence supporting the heterogeneity of learning disabilities.

These concerns notwithstanding, some important ramifications emerge from the meta-analytic studies. Mueller et al. (1984) cautioned against the use of profile analysis for classification of children and emphasized the need for future research to include samples of normal children as well as exceptional children. Further, a cogent reminder is provided that group profiles should not be applied to "differential diagnosis of individual children; rather, such group-profiles are more likely to be useful at a more molar level in suggesting broad cognitive differences between levels of intellectual functioning" (Mueller et al., 1984, p. 81). The implications of this latter observation, in particular, are addressed in the interpretation of the research findings arising from this study. The results of Mueller et al.'s. (1984, 1986) studies have implications for the use of the SB:FE in this present study and given the high saturation of g in the SB:FE (Thorndike et al., 1986b), it would be important to determine whether the findings documented by Mueller et al. (1984, 1986) with the WISC-R also apply to the SB:FE. Finally, Sattler (1992) noted that there is no WISC-R profile that is reliably diagnostic of learning disabilities, although this is in part attributed to the heterogeneous nature of learning disabilities.

The SB:FE and Subtyping

Rourke, Fisk, and Strang (1986) criticized traditional psychoeducational approaches to the identification of learning disabilities maintaining that they do not cover a broad enough range of abilities for the purposes of remediation. They also noted that the psychoeducational test results are interpreted primarily from "a level of performance perspective" (p. 160). However, neuropsychological models (of which Rourke and colleagues are strong proponents) for use with children must be viewed with caution (Shepard, 1989). Brain-behaviour relationships, on which such models are based, are derived largely from adult models and fail to take into account developmental variables (Holmes, 1986, p. 305). Since 1975, a number of criticisms have been levelled at neuropsychological approaches to learning disabilities (Gaddes, 1981). Neuropsychological measures used with children have inadequate norm bases, limited content validity, homogeneous scales and measures, and generally tend to have inadequate coverage of all domains of neuropsychological functioning (Hynd, 1988, p. 84). As well, existing traditional neuropsychological measures are confounded by high saturations of general ability (Hynd, 1988) and similar factor structures (Sutter, Bishop, & Battin, 1986). Hynd (1988) maintained "there is no convincing evidence that these batteries offer any more discrete measures of neuropsychological functioning than traditional clinical or psychoeducational evaluations" (p. 85). Shepard (1989) noted that, regardless of the current enthusiasm for neuropsychological approaches, "they do not have established validity for making diagnoses in a milder and more ambiguously defined school population" (p. 561) and further questioned whether these instruments "should be disseminated for use in the schools" (p. 562).

Added to these limitations are the constraints imposed by time and levels of clinical expertise that are required to conduct neuropsychological assessments. It is questionable whether such approaches will become practical from a clinical and educational perspective. Moreover, recent applications of the WAIS-R as a neuropsychological index (WAIS-R NI) (Kaplan, Fein, Morris, & Delis, 1991), focusing on process oriented interpretations, suggest intellectual measures that tap higher order cognitive functioning, may be of increasing utility in understanding brain-behaviour relationships. Thus, it is necessary to emphasize the importance of using traditional cognitive instruments in the identification of learning problems because these are consistently and regularly used in clinical practice. These factors make research into learning profiles and subtyping with contemporary cognitive assessment scales like the SB:FE a priority.

The need for logically consistent and specific criteria for diagnosing the presence of learning disabilities is paramount. Traditionally, such diagnoses have been made on the basis of exclusionary criteria. Learning disabled individuals possess normal intellect, intact sensory/perceptual functions, cultural and educational exposure, but still display deficits in academic achievement. There is a shift toward diagnosis of learning disabilities, not by exclusion, but by adoption of more inclusionary criteria. This has been accomplished through incorporation into standard definitions of learning disabilities statements concerning the presence of identifiable or presumed neurological dysfunction or processing deficits. In conjunction with this, there has arisen an emphasis on identifying valid and meaningful subtypes of learning disability. Research is documenting the power of subtyping models in relation to specific intervention plans (Fiedorowcz & Trites, 1991; Hooper & Willis, 1989; Lyon & Flynn, 1991; Rourke, 1991). It is these trends that make it vital to assess the validity of current psychoeducational instruments such as the SB:FE in the identification of individual differences.

Rationale for the use of the SB:FE and SB:FE GPAB in Multivariate Classification Research

The SB:FE and SB:FE GPAB have been introduced in this chapter. The psychometric properties of both instruments were reviewed. In addition, the need to develop accurate and valid diagnostic measures in education in order to facilitate educational planning was highlighted in this chapter. The SB:FE was designed to be a cornerstone in psychoeducational assessment (Thorndike et al., 1986b). Thorndike et al. (1986a) stressed the importance of SB:FE results in formulating recommendations for educational interventions. The four cognitive ability area scores of the SB:FE appear to have the potential to yield valuable diagnostic and remedial information in educational programming. However, claims concerning the utility of the instrument need to be established through research.

In this study, the relationship of the SB:FE GPAB to the complete test battery was investigated. Use of the SB:FE GPAB reduces the time needed for test administration and ensures that a uniform test battery is administered to all ages and abilities. The discriminant validity of the SB:FE GPAB was investigated in the identification of special populations. Mild handicaps such as learning disabilities, pose a particular challenge to psychoeducational assessment and there is a need to develop adequate subtyping nosologies within these populations. This is because of their value in the applications aimed at understanding the etiology, assessment and treatment of these conditions. Thus, the SB:FE GPAB was subjected to multivariate clustering procedures in order to investigate the viability of subtyping individuals into groups based on similarity or distinctiveness of cognitive profiles. Skinner's (1981) theoretical framework, emphasizing internal and external validity, provided a guide for this aspect of the research. The agreement between empirical and a priori classification schema was explored in the final external validation stage.

In the next chapter, the research methods and procedures used in this dissertation are presented. This is followed by a presentation of the results in Chapters Four, Five, Six, and Seven. In Chapter Eight, the results of the study are summarized, and the conclusions and implications for future research and practice discussed.

CHAPTER THREE: METHODS AND PROCEDURES

Overview

In this chapter the research purposes and hypotheses arising from the literature reviewed in the previous chapter are outlined. For convenience, the full study has been separated into four main sections, corresponding, respectively, to the main objectives of the study stated previously in Chapter One. In the first part of the study, descriptive information about the SB:FE when used with a Canadian population is examined. Hypotheses designed to address the last three parts of the study are presented next, followed by a description of the sample, test instruments, test administration procedures, and age categorizations to be used in the study. The chapter concludes with a brief overview of the statistical procedures used to test the hypotheses.

As noted previously, in describing the research procedures, short hand notation has been used to designate the age groupings. For example, 2-23-11 means 2 years through 23 years, 11 months; 2-6-11 means 2 years through 6 years, 11 months; 7-11-11 means 7 years through 11 years, 11 months; and 12-23-11 means 12 years through 23 years, 11 months.

Part One

In Part One of the study, a comprehensive description of the SB:FE when used with a Canadian clinic sample was determined. The demographic characteristics of the sample were considered, and SB:FE descriptive statistics, subtest and Reasoning Area and Composite SAS means, standard deviations, and intercorrelations were calculated. In addition, WRAT-R data, available for part of the sample, were also examined.

Part Two

In Part Two of the study the relationship between the SB:FE full battery and the SB:FE GPAB was explored using the research hypotheses stated below.

Hypothesis 1:1

There will be no significant differences between the four Reasoning Area and Composite SAS's means based on the full battery of the SB:FE and the four respective Reasoning Area and

Composite SAS's means derived from the SB:FE GPAB at each age level 2-6-11, 7-11-11, and 12-23-11.

Hypothesis 1:2

There will be no significant differences between the four Reasoning Area and Composite score variances based on the full battery of the SB:FE and the four respective Reasoning Area and Composite score variances based on the SB:FE GPAB at each age level 2-6-11, 7-11-11, and 12-23-11.

Hypothesis 1:3

Significant correlations, approaching unity, will be obtained between the four Reasoning Area and Composite SAS's based on the SB:FE full battery and the four respective Reasoning Area and Composite SAS's derived from the SB:FE GPAB at each age level 2-6-11, 7-11-11, and 12-23-11.

Part Three

Due to the methodological difficulties inherent in multivariate work with the SB:FE described in the previous chapter, the third part of the study focused on analyses based on the abbreviated version of the SB:FE, the SB:FE GPAB. The purpose of this stage of the study was to apply multivariate cluster analytic procedures to SB:FE GPAB data in order to derive reliable and replicable (internally valid) groups of individuals with distinct cognitive profiles. The variables used in the cluster analysis were the four reasoning area SAS's of the SB:FE GPAB: the Verbal Reasoning SAS; the Abstract/Visual Reasoning SAS; the Quantitative Reasoning SAS; and the Short Term Memory SAS. Note, for this part of the study, the age group 2-6-11 was separated into two subsamples: age 2-4-11 and 5-6-11. This is because external validation procedures applied to the samples in Part Four of this study utilize WRAT-R achievement data. Subjects below the age of 5-00 are not routinely administered the WRAT-R, and the youngest subsample, age 2-4-11 will be excluded from this part of the study. Inclusion of this group in this part of the analysis is warranted as this may reveal developmental patterns or trends, that would otherwise be overlooked.

Hypothesis 2:1

Application of hierarchical agglomerative (HAP) and iterative partitioning (IPP) cluster analytic procedures to the SB:FE GPAB data will result in the identification of subgroups within the sample age ranges 2-4-11, 5-6-11, 7-11-11, and 12-23-11 years that display significant differences with respect to profile elevation and/or shape.

Part Four

The purpose of this part of the study was to explore the external validity of the cluster solutions derived in the previous section. First, the empirically derived subgroups' mean scores on the external criteria of Wide Range Achievement Test-Revised scores were compared. Next, the empirically derived cluster solutions were compared with a priori derived clinical inferential groups. Analyses were conducted for the age samples 5-6-11, 7-11-11, and 12-23-11.

Hypothesis 3:1

Subgroups, within the age samples 5-6-11, 7-11, and 12-23-11, identified on the basis of iterative partitioning cluster analytic procedures applied to SB:FE data, will demonstrate significant mean differences on the external criteria of WRAT-R achievement scores.

Hypothesis 3:2

There will be congruence between a priori clinically determined subgroups and subgroups empirically derived through application of iterative partitioning cluster analytic procedures to the subsamples, age 5-6-11, 7-11-11, and 12-23-11, for whom both SB:FE data and WRAT-R data were available.

Method

Subjects

SB:FE data obtained from psychoeducational assessments of subjects tested at the University of Alberta, Faculty of Education Clinic from 1986 to 1993 were analyzed to provide comprehensive descriptive data for Part One of the study and to test each of the hypotheses stated for Parts Two, Three, and Four of the study. The sample comprised 1220 individuals tested at the University of Alberta, Faculty of Education Clinic between 1986 to 1993. The sample of psychoeducational assessments included: SB:FE data; comprehensive demographic data for subjects age 2 years, 0 months through 23 years, 11 months and WRAT-R achievement data for certain subjects (n=607) age 5-00 and older. The University of Alberta, Education Clinic is an established service operated under the auspices of the Department of Educational Psychology, Faculty of Education. Its primary function is to train graduate students in School Psychology and Counseling Psychology programs. Referrals are accepted from a range of sources including medical doctors, community service agencies, schools, parents, and individuals.

Spelliscy (1991) identified several reasons which support conducting research with this clinic sample. First, the logistic effort to obtain protocols is onerous; the sample available through the Education Clinic is readily accessible. Test administration and scoring are overseen by qualified experts. Within this sample, the reasons for referral or presenting concerns are diverse, suggesting the sample may be ideal for the empirical analysis proposed. Services are provided to a wide range of individuals, and the sample thus closely approximates the general population and individuals routinely seen by educational psychologists. For example, the sample used by Spelliscy (1991) presented with a diverse range of referral concerns: parental interest (approximately 43%); learning concerns (approximately 30%); school placement (approximately 15%); gifted populations (approximately 8%); and mentally handicapped (approximately 4%).

Spelliscy (1991, p. 220) reported findings "consistent with those proposed by Thorndike et al. (1986b)" based on analysis of descriptive statistics and intercorrelations among subtests of the SB:FE. Individual subtest means and standard deviation scores typically fell within two points of those reported by Thorndike et al. (1986b). The means and standard deviations of the four cognitive ability areas were slightly higher than those reported by Thorndike et al. (1986b) within the sample age 2-6-11, were generally consistent in the sample age 7-11-11, and slightly lower than those reported by Thorndike et al. (1986b) in the sample age 12-23-11. Overall, the clinic sample used by Spelliscy (1991) typically scored approximately three to six points higher or lower and ranged 3 points more than the reported SB:FE values (p. 220).

Instruments

In addition to the SB:FE and the SB:FE GPAB, which have been extensively reviewed in Chapter 2, subjects comprising the subsamples used in Part Three and Four of the full study were also administered the WRAT-R (Jastak & Wilkinson, 1984).

Wide Range Achievement Test-Revised (WRAT-R)

The Wide Range Achievement Test (WRAT) was originally published in 1936 (Jastak, 1936). Since then it has been revised six times (1946, 1965, 1976, 1978, 1984, & 1993) culminating with the WRAT-3 (Wilkinson, 1993). This study used the WRAT-R (Jastak & Wilkinson, 1984). The WRAT-R is a brief individually administered achievement test designed to provide a standardized assessment of educational skills. It measures three major areas of educational achievement: Reading (word recognition), Spelling, and Arithmetic (mechanical arithmetic). The first two subtests are untimed. The Rasch Model (Rasch, 1980) was used in the development of the WRAT-R. Items were calibrated and persons were measured using the unconditional maximum likelihood estimation method (Wright & Meade, 1978). Standardized on 6,000 subjects, the test has two forms: Level I for subjects aged 5-00 to 11-11, and Level II for subjects aged 12-00 to 74-11. Standard scores with a mean of 100 and a standard deviation of 15 are available, as well as percentiles and grade equivalents.

Internal consistency estimates were based on statistics derived from the Rasch Model and within this model Jastak and Wilkinson (1984) note that reliability refers to the separation of persons or the ability to distinguish among a sample of persons on the basis of their measures (p. 61). Thus, WRAT-R Level I internal consistency reliability estimates based on this statistic ranged from .86 to .98, with a median of .94 for the Reading subtest, from .88 to .94, with a median of .93 for the Spelling subtest, and from .78 to .87, with a median of .82, for the Arithmetic subtest. WRAT-R Level II internal consistency reliability estimates ranged from .93 to .98, with a median of .96 for the Reading subtest, from .92 to .97, with a median of .93 for the Spelling subtest, and from .81 to .97, with a median of .91 for the Arithmetic subtest.

Test-retest coefficients for the WRAT-R Level I Reading, Spelling, and Arithmetic tests were reported to be .96, .97, and .94, respectively (Jastak & Wilkinson, 1984, p. 64). The WRAT-R Level II stability coefficients for Reading, Spelling, and Arithmetic were .90, .89, and .79, respectively. Stability coefficients for Reading, Spelling, and Arithmetic for the combined sample were .94, .95, and .92, respectively. Concurrent validity studies cited in the WRAT-R test manual (Jastak & Wilkinson, 1984) compared the WRAT-R with the Peabody Individual Achievement Test (PIAT) (Dunn & Markwardt, 1970). Twenty different studies involving the WRAT-R Reading subtest yielded an average correlation of .87 with the PIAT Reading recognition subtests (Jastak & Wilkinson, 1984). The average correlation between WRAT Reading subtest scores and the California Achievement Test (CAT) (CTB/McGraw Hill, 1979) Total Reading scores across several studies was .81. Twenty different studies involving CAT Spelling and WRAT-R Spelling scores yielded an average correlation of .85. WRAT-R Reading subtests scores correlated .82 with the Stanford Achievement Test (SAT) (Gardner, Rudman, Karlson, & Merwin, 1982) Word Reading subtest. Sattler (1988) reported significant correlations between WRAT-R Reading, Spelling, and Arithmetic scores and relevant subtests of the Woodcock-Johnson Psycho-Educational Battery (Woodcock, 1977) and the Kaufman Test of Educational Achievement (K-TEA) (Kaufman & Kaufman, 1985). Most recently, Wilkinson (1993) reported significant correlations between subtests of both forms of the WRAT-3 and respective WRAT-R subtests in a sample of 77 subjects with an average age of just over 10 years. Uncorrected correlations of .83 and .87 were obtained between Reading Subtests, of .87 and .92 between Spelling Subtests, and of .79 and .85 between Arithmetic subtests (Wilkinson, 1993).

Despite its extremely widespread use, the WRAT-R has been criticized for its restricted coverage, particularly with respect to the absence of reading comprehension measures. However, according to Fletcher (1985), WRAT-R "scores correlate highly with more specific academic measures" (p. 197). Moreover, in a number of studies where WRAT-R measures have been used to develop typologies, the addition of reading comprehension measures did not alter subgroups (Fletcher & Satz, 1985). In addition, Siegel (1984, 1986, 1990), Siegel and Heaven

(1986), and Siegel and Ryan (1984) argued that word recognition tasks such as the reading subtest of the WRAT-R importantly measure impaired decoding skills that may constitute the basis of reading problems. They are thus less problematic than other types of reading measurement such as reading comprehension. Reading comprehension measures may confound the measurement of reading skills, especially if speed of performance is a factor. Questions on reading comprehension tests may require previous knowledge, speed of processing, and different inferential processes, and may tap confounding visual processing dimensions (Biemiller & Siegel, 1988).

The WRAT-R and earlier versions of the test have been consistently used in a vast number of studies in the area of learning problems. In this study, use of the WRAT-R provided an important base for comparative studies, in addition to providing continuity with past research endeavours of a similar nature, but utilizing different cognitive ability measures such as the WISC-R (Wechsler, 1974).

Age Categorizations

The adaptive format of the SB:FE poses particular problems in clinical use (Sattler, 1988, 1992) and in multivariate research (Fritzke, 1988; Spelliscy, 1991) that are compounded by differences in test format at various age levels and, more importantly, across individuals of the same age. The testing format of the SB:FE was designed to be flexible, in that Reasoning Area and Composite SAS's may be based on a number of different combinations of tests, in addition to the six core tests which have to be administered to all age levels (Glutting, 1989; Thorndike et al., 1986a, 1986b, 1986c). Age groupings are routinely employed prior to statistical analysis to overcome concerns relating to this difficulty. Each age group is typically treated as a separate group. Thorndike et al. (1986b) used the following age groups: 2 through 6 years; 7 through 11 years; and 12 through 18-23 years, because these were the groups that had typically taken the same tests from among the 15 subtests included in the SB:FE during standardization. Keith (1987) maintained that using age groupings enabled researchers to observe developmental changes in intelligence if they are present. Addressing the difficulty posed by the fact different individuals of the same age may be administered different subtests is more complex. However, to minimize this particular difficulty, the SB:FE GPAB, which covers the entire age range from 2 to 23-11, was used in the analyses designed to test the hypotheses in Part Three and Part Four of this study. This ensured similar constructs were measured at each age level throughout the SB:FE's age range and across all subjects within the same age group. In Part Two of this study, the relationship between the SB:FE full battery and SB:FE GPAB was documented. Throughout this study, the same age groupings used by Thorndike et al. (1986b) were adopted in order to facilitate comparisons with the test publishers findings.

Test Administrators

All test administrators were students enrolled in graduate level psychoeducational assessment courses. Test procedures taught in these courses follow guidelines provided by Thorndike et al. (1986a, 1986c). Test protocols are checked for administration and scoring accuracy by University of Alberta, Department of Educational Psychology faculty or trained assistants.

Statistical Analyses

Part One

Descriptive and Demographic Data

All subjects (N=1220) administered the SB:FE comprised the sample used in Part One of the study. The total sample age 2-23-11 and the three age groupings, 2-6-11, 7-11-11, and 12-23-11 used by Keith et al. (1988a, 1988b), Spelliscy (1991) and Thorndike et al (1986a, 1986b) were used. Comprehensive demographic data was presented. Chi square analysis was conducted to test for differences in frequencies among demographic characteristics across the three age samples. Means, standard deviations, and Pearson product-moment correlations for subtest, Reasoning Area, and Composite SAS's across the three age groups and the total sample were also calculated. Finally, differences in performance on SB:FE and on WRAT-R variables across the three age groups sampled were tested using Univariate Analysis of Variance and Multivariate Analysis of Variance followed by univariate F tests for independent groups respectively.

Part Two

Evaluation of the Relationship Between the SB:FE and the SB:FE GPAB

All subjects (N=1220) comprised the sample used in Part Two of the study. Differences in means and variances between Verbal, Abstract/Visual, Quantitative, and Short Term Memory Reasoning Areas, and Composite SAS's derived from the SB:FE full battery and the SB:FE GPAB were tested using respectively, student's t-test, and the F test test for dependent groups. Next, the relationship (Pearson product-moment correlation) between these variables was examined. The nature of the hypothesis with respect to this relationship called for correlations corrected for attenuation. To determine the corrected correlations required estimates of the internal consistency of the SB:FE and SB:FE GPAB Reasoning Area and Composite scores. Unfortunately these estimates were not available. Consequently, the uncorrected coefficients, which provide a lower bound estimate of the degree of correlation were examined. The analyses were completed for each of the three age groups, 2-6-11, 7-11-11, and 12-23-11.

Creation and Internal Validation of [illegible]

In the third part of the stu [illegible] AD to identify homogeneous subgroups or subtypes of indi [illegible] y profiles was examined. The cluster analyses used relied la [illegible] s supportive of the construct validity of the four Reasoning Area score [illegible] s, 1990; Keith et al., 1988a, 1988b; Spelliscy, 1991; Thorndike et al., 1986b). Standard Age Scores (SAS's) derived for each of the four Reasoning Areas of the SB:FE GPAB were first clustered using hierarchical agglomerative (HAP) clustering procedures (SPSSx, 1990) and then validated using iterative partitioning (IPP) techniques (BMDP-KM) (Dixon, 1981). The total sample was divided into two main groups for analyses. The first group included subjects for whom both SB:FE GPAB data and WRAT-R data were available and analyses were conducted for the following age groups; 5-6-11 (n=120), 7-11-11 (n=301), and 12-23-11 (n=186). Subjects for whom only SB:FE GPAB data were available were used to validate the cluster solutions derived using the sample with both SB:FE GPAB and

WRAT-R achievement data. Analyses were conducted for the following age groups; 2-4-11 (n=110), 5-6-11 (n=164), 7-11-11 (n=226), and 12-23-11 (n=113).

Cluster Analysis

Cluster analytic procedures are multivariate statistical procedures designed to find groups or clusters of similar entities in a sample of data by minimizing profile differences within groups and maximizing differences among groups. They are intended for classification purposes (Aldenderfer & Blashfield, 1984). Aldenderfer and Blashfield (1984) identifed five steps in cluster analysis: sample selection; definition of variables on which to measure the subjects in the sample; computation of the similarities among the entities; use of a cluster method to create groups of similar entities; and validation of the resulting cluster solution (p. 12). Issues relating to the first two of these five steps, the sample and definition of variables used to measure the subjects have been addressed.

Similarity index Correlation coefficients and distance measures are two measures of similarity commonly used in the social sciences (Aldenderfer & Blashfield, 1984). Squared Euclidean distance (D^2) was used as the similarity measure in this study. This measure takes into account relative profile elevation, shape, and scatter of scores (Blashfield, 1980; Cronbach & Gleser, 1953; Fleiss & Zubin, 1969; Skinner, 1978). Although correlation coefficients are the most frequently used measures of similarity in the social sciences, they are insensitive to differences in elevation and dispersion and results that do not take these factors into account may be misleading (Aldenderfer & Blashfield, 1984).

Cluster method. The next step in cluster analysis involves the choice of clustering technique. Several clustering methods have been developed (Aldenderfer & Blashfield, 1984). Of these, hierarchical agglomerative methods and iterative partitioning techniques are the most popular in the social sciences, with the latter more typically used (Aldenderfer & Blashfield, 1984; Everitt, 1979). Both procedures were used in the present study.

Hierarchical agglomerative methods search a similarity matrix and sequentially merge the most similar cases which are typically represented visually by tree diagrams or dendograms.

Although conceptually simple, hierarchical agglomerative techniques are single stage methods; only one pass is made through the data. Thus, it is impossible to relocate poor partitions of the data set in subsequent steps of the clustering process. The first stage of the cluster analytic procedures adopted in this study involved use of the hierarchical agglomerative procedure (SPSSx, 1990). Subjects in each of the subsamples were subjected to a stepwise clustering process, using a hierarchical agglomerative algorithm that defines clusters on the basis of maximizing between group distances. This procedure produces successively larger cluster groups by joining together individual cases that have the highest similarity between their profiles, and which are maximally dissimilar to other individuals or already formed clusters. The process continues until all clusters are combined into one group. As noted, Squared Euclidean distance (D^2) was used as the distance (similarity) measure. Ward's minimum variance was selected as the clustering algorithm (Ward, 1963). Studies have suggested that Ward's method is highly accurate in recovering clusters (Blashfield, 1976; Edelbrock, 1979; Gross, 1972; Kuiper & Fisher, 1975; Mojena, 1977; Milligan, 1981). Speece (1990) and Lorr (1983) maintain that Ward's method performs best in combination with Euclidean distance measures.

A k-means iterative clustering procedure (BMDP-KM) was then used to partition the data set (Dixon, 1981). Iterative partitioning approaches are two-stage methods which allow for relocation of entities that have been poorly classified in earlier stages (Aldenderfer & Blashfield, 1984; DeLuca. Adams, & Rourke, 1991). They are based on estimates of the number of clusters in a data set. For any given number of clusters, each cluster's centroid is estimated. Centroids may be chosen randomly, based on the distance between centroids, or based on estimates established by the researcher. The BMDP-KM procedure (Dixon, 1981) partitions a set of cases, each consisting of variable scores into clusters using an algorithm outlined by Hartigan (1975). Squared Euclidean distance (D^2) was again used as the similarity measure. The program begins with all the cases in one cluster and then splits one cluster into two clusters at each step and continues until a user specified number of clusters is reached. The program allows clusters to be reallocated at each step. An optimal solution is achieved when the clusters possess maximum homogeneity, while the distance between clusters is simultaneously maximized.

Analysis was performed for 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 11- and 12-cluster solutions. Cluster group means derived from the SPSSx hierarchical agglomerative procedure were used as initial seed values for each of the specified iterative partitioning cluster solutions. Scheibler & Schneider (1985) reported that k-means iterative procedures yielded accurate solutions when non-random starting seeds are used, but they noted "evidence from different Monte Carlo procedures leads to the conclusion that the use of random starting points will generally result in suboptimal classifications" (p. 299). Milligan (1981) concurred. Once seed values are determined, subjects are then assigned to the clusters with the most similar centroids. Importantly, iterative techniques allow for reassignment to more appropriate groups (Morris, Blashfield, & Satz, 1986).

Number of clusters. One of the most difficult issues in cluster analysis is the determination of the number of clusters (Aldenderfer & Blashfield, 1984; Blashfield, 1980; Everitt, 1979). No adequate techniques have been proposed for solving the number of clusters problem (Everitt, 1979). However, Everitt (1979) argued that given the lack of a universally acceptable definition of the term cluster, interpretability and simplicity may be important factors in determining the number of clusters. Heuristic procedures tend to be the most commonly used methods to determine the number of groups (Aldenderfer & Blashfield, 1984). These involve subjective inspection of the general levels to determine the number of clusters.

Cluster Analytic Procedures Used in This Study

In this study, determination of number of clusters was guided by a combination of approaches. The clustering coefficient metric or fusion coefficient (which reflects the overall level of distance differences within the clusters) was calculated for each stage of the clustering process. The SPSSx hierarchical agglomerative clustering program output reports clustering coefficients for all steps in the clustering process (SPSSx, 1990). The optimal cluster solution lies at a point just before the clustering coefficient displays a sudden increase that is out of proportion to previous increments (Aldenderfer & Blashfield, 1984; Everitt, 1974, 1979; Mojena, 1977; Wishart, 1975). A sudden increase in the value of the clustering coefficient indicates two dissimilar groups have been forced together by the agglomeration process, resulting in a new group with significantly reduced internal homogeneity.

With the iterative partitioning procedures a different method was used to determine the number of clusters. Mueller et al. (1984, 1986) developed a modification of Lee and MacQueen's (1980) Next-Nearest-Neighbour Ratio (NNR) in order to determine the optimal number of solutions and this modification is used in this study. The NNR provides information about the distribution of points within each cluster and the relationship of the distributions between clusters. It simply provides a ratio of the distance of the member from its nearest centroid to the distance to the next nearest centroid (Lee & MacQueen, 1980). In Mueller et al's. (1984, 1986) modification, the greatest within-cluster mean D^2 was divided by the smallest between-cluster mean D^2. This ratio becomes greater as the shortest distance between any two clusters becomes smaller relative to the mean distance of samples from their own cluster centre within the least homogeneous cluster and ideally should not exceed .6. The best cluster solution is thus that which defines the most clusters with the greatest homogeneity while maximizing the separation between the clusters.

In the present study, examination of the fusion coefficient and application of the modification of the NNR were used to determine the optimal number of clusters.

Internal Validation of Cluster Solution. Results of the various solutions at each age level were examined for internal validity using procedures suggested by Aldenderfer and Blashfield (1984), Blashfield (1980), Everitt (1974), Joschko and Rourke (1985), Rourke (1991), and Skinner (1981). Clustering should be attempted using various methods and algorithms and groupings replicated across clustering techniques should be adopted (Joschko & Rourke, 1985, Wishart, 1975). First, solutions obtained from hierarchical agglomerative methods and iterative partitioning techniques were compared. The level of congruence between results achieved through these two different procedures reflects the replicability and internal consistency of cluster solutions.

Second, the cluster solutions derived from the data set comprising subjects with both SB:FE and WRAT-R data were compared to the cluster solutions obtained with the sample comprising subjects for whom SB:FE data only was available. To the extent that similar solutions were discovered across different samples from the same general population this provided evidence supportive of the stability and generality of the solutions obtained (Blashfield, 1980; Everitt, 1974, 1980).

Part Four

External Validation of Cluster Solutions

The sample used in the fourth part of the study comprised only those individuals administered both the SB:FE GPAB and the WRAT-R (n=607). Three groups were used: ages 5-6-11, 7-11-11, and 12-23-11. This is because the external validation procedures involved comparisons of cluster analytic groups with external measures (e.g., WRAT-R data) not used in the delineation of the original subtypes. The clusters were formed using SAS's from the four Reasoning Areas of the SB:FE GPAB. Solutions derived from the iterative partitioning procedures were selected. This is because iterative partitioning procedures allow reallocation of subjects poorly classified in earlier stages, and may thus yield better results than procedures like hierarchical agglomerative techniques, which do not allow this (Aldenderfer & Blashfield, 1984). WRAT-R achievement test scores which were not used in the creation of the cluster solutions were used as the external measure. Conducting significance tests on external variables such as WRAT-R achievement data is the most powerful means of validating clustering solutions and involves performing significance tests on variables that are not used to generate the original cluster solution, directly testing the generality of a cluster solution against relevant criteria (Aldenderfer & Blashfield, 1984). Subjects below the age of five are not routinely administered achievement tests and thus achievement data were not available for this particular age group.

Evidence pertaining to the external validity of the cluster solutions was collected in two separate steps. First, the external validity of various cluster solutions was tested by comparing mean WRAT-R achievement scores of each group derived through application of empirical clustering procedures to SB:FE GPAB data. Multivariate Analysis of Variance (MANOVA) was followed by univariate F tests. Second, cross classification tables were used to investigate the relationship between subtypes empirically derived through the cluster analytic procedures applied to SB:FE GPAB data and clinically derived a priori determined groups. The cross-classification table is one of the most useful tools for the bivariate or multivariate analysis of nominal data (Agresti & Finlay, 1986). Cross-classifications appear in two-dimensional tables, each row representing a category of one variable and each column representing a category of another variable. Each unique combination of categories of the variable is represented by a cell in the table. Four a priori clinical inferential models, described in the next subsection, were considered.

The results of the procedures used in Part Three of the study generally did not support the existence of more than three clusters in the age samples 5-6-11, 7-11-11, and 12-23-11. Therefore, the cross-classification data for the three cluster solutions were considered first. Based on the cross-classification tables, the agreements obtained between the empirically derived and clinically derived models were not perfect. The question then arose whether the agreement between the empirically derived solution and the clinically driven classification systems would improve if other than a three cluster solution was used. Therefore to test this conjecture alternate numbers of empirically driven solutions were examined. Alternate values for k were considered: 4; 5; 6; and 7. Speece (1990) maintained there is no single set of procedures that has evolved as the superior approach to cluster analysis" (p. 209). Speece (1990) stated that researchers have been overly concerned with deriving the correct number of clusters in early stages of research design and suggested that it may be useful to entertain several possible solutions and evaluate them against external criteria. Speece (1990) noted that classification efforts within the area of learning disabilities are primarily exploratory in nature and thus such an approach is warranted. To the extent that results of comparisons between various cluster solutions are not be not meaningful, the three cluster solution gains more credibility. Thus, in light of Speece's (1990) observations and in light of the relatively poor fits obtained between the three cluster empirical solutions and the clinical classification models, the four a priori models were then tested against a range of cluster solutions from 2 through to 7 clusters.

Clinically Derived a priori Groups

Subjects were allocated to a priori determined groups based on clinical-inferential classification procedures modified and adapted from Mueller et al. (1984, 1986), Reynolds (1984), Rourke (1975, 1978, 1982, 1983, 1985, 1989), Rourke and Finlayson (1978), Rourke et al. (1986), Rourke and Fuerst (1991), Rourke and Strang (1978), and Thorndike et al. (1986b).

Model One: Intellectual classification groups. Each subject in the study was assigned to one of seven a priori determined groups according to SB:FE GPAB Composite SAS scores, based on the classifications used by Thorndike et al. (1986a, 1986b). These classifications are summarized in Table 5. In the ability classifications provided in the SB:FE Technical Manual (Thorndike et al., 1986b, p. 127), the Slow Learner classification replaces the traditionally used Borderline Defective classification.

Table 5

Classification Ratings for Composite SAS's Scores for the SB:FE

Classification	SB:FE GPAB Composite Range
Very Superior	132 and above
Superior	121-131
High Average	111-120
Average or Normal	89-110
Low Average	79-88
Slow Learner	68-78
Mentally Retarded	67 and below

Adapted from Thorndike et al. (1986b) and Sattler (1988)

The purpose of this comparison was to explore the effect profile elevation has primarily on cluster solutions consistent with analysis performed by Mueller et al. (1986). Individuals who obtain Superior and Very Superior SB:FE GPAB Composite SAS's could typically be classified as gifted depending on pre-determined cut off scores.

In Alberta, gifted children are described as:

> (T)hose who by virtue of outstanding abilities are capable of high performance. These are children who require differentiated educational programs and/or services beyond those normally provided by the regular school program in order to realize their contribution to self and society. (Wilde & Sillito, 1986, p. 7)

Typically, criteria for inclusion into "gifted programs" include superior intellectual ability as obtained on an individually administered intelligence test (e.g., IQ's above 120 or 125). Other factors that may be taken into account include superior academic achievement, adequate social and emotional maturity, ability to work independently, enjoyment of intellectual challenges, and maintenance of satisfactory levels of achievement in the regular classroom. These latter factors are difficult to measure objectively. Zigler and Farber (1985) maintained specific IQ levels were the most adequate index of giftedness. Sattler (1988) concurred, noting the "single best method for the identification of children with superior cognitive abilities is a standardized, individually administered test of intelligence" (p. 671).

Individuals who score at 67 or below on the SB:FE GPAB Composite SAS were classified as falling in the Mentally Retarded category. Mental retardation refers to "significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behaviour, and manifested during the developmental period" (Grossman, 1983, p. 11). Significantly subaverage refers to performance on an individually administered intelligence test that is two standard deviations below the population mean (Sattler, 1988). Adaptive functioning refers to the effectiveness with which individuals meet the standards of personal independence and social responsibility in terms of culture and age (Sattler, 1988). In this study, the only criteria for assignment to the Mentally Retarded group was psychometric, based on SB:FE GPAB Composite SAS's. Although less acceptable for clinical purposes, this is acceptable for research purposes where the valence attached to identification/misidentification is different (Shepard, 1989).

Model Two: Normal learners compared with homogeneous Learning Disabled subjects. In this model, subjects were assigned to one of five *a priori* determined groups. This model separated subjects broadly into groups classified according to SB:FE GPAB Composite SAS scores with academic achievement (WRAT-R) commensurate with cognitive ability, in addition to a group of homogeneous Learning Disabled (LD) subjects. Assignment to groups was based on information obtained from original clinic assessment data. The following groups were identified:

Gifted Group: Individuals who obtain SB:FE Composite SAS's of 120 or above were included in the gifted group. No subjects with academic deficits or delays were present in this group.

Normal Learner Group: The Normal Learner Group comprised individuals with SB:FE GPAB Composite SAS's between 79 and 124. Thus, the normal group comprised subjects with Composite SAS's in the Low Average, Average, and High Average range defined by the classification guide to the SB:FE (Thorndike et al., 1986b, p. 127). No subjects with academic deficits or delays were present in this group.

Slow Learner: The Slow Learner Group consisted of individuals with SB:FE Composite SAS's between 68 and 78.

Mentally Retarded Group: Individuals who score 67 or below on the SB:FE GPAB Composite SAS's were classified as falling in the Mentally Retarded category.

Learning Disabled Group (LD): Criteria consistent with current learning disability definitions were applied in identification of the learning disabled sample (e.g., Alberta Education, 1986, 1988, 1992; LDAC, 1981). Subjects demonstrated at least average levels of psychometric intelligence (e.g., SB:FE Composite scores of 79 and above). English must have been the first language spoken at home. On the basis of information derived from assessment records none of the subjects showed evidence of the following: inadequate visual or auditory acuity; neurological disease; severe cultural, educational or environmental deprivation; or significant emotional disturbance. Subjects with evidence of these difficulties were excluded from the study. The determination of the learning disabled groups was based primarily on the use of discrepancies between ability as measured by the SB:FE GPAB and achievement as measured by the WRAT-R. The simple prediction model (Reynolds, 1984; Shepard, 1989) was used to determine a severe discrepancy and one homogeneous LD group was formed.

<u>Model Three: Normal Learners compared with heterogeneous Learning Disabled groups</u> This model is based on the model proposed above. However, Learning Disabled subjects were treated as heterogeneous groups and classified according to academic criteria. Thus, in addition to Gifted (Very Superior and Superior), Normal Learners, Slow Learners, and Mentally Retarded groups, Learning Disabled subjects were placed in groups according to performance on specific WRAT-R achievement subtests. Again, the simple prediction model (Reynolds, 1984; Shepard, 1989) was used to determine a severe discrepancy within each of the academic achievement subject areas. The following groups of LD subjects [illegible]ated: Spelling Disabled; Math Disabled; Re[illegible]ing Disabled; Reading and Spelling [illegible]ied; Reading and Math Disabled; Spelling and Math Disabled; and Reading, Spelling, and Math Disabled.

<u>Model Four: Normal Learners compared with Learning Disabled subjects based on Rourke's classification system</u>. Procedures modified from Rourke (1975, 1978, 1982, 1983, 1985, 1989), Rourke and Finlayson (1978), Rourke et al. (1986), Rourke and Fuerst (1991), and Rourke and Strang (1978) were used in determination of Learning Disability subgroups. The exclusionary criteria adopted by Rourke and colleagues are consistent with currrent learning disability definitions cited earlier in this study (e.g., Alberta Education, 1986, 1988, 1992; LDAC, 1981). Again subjects demonstrated at least average levels of psychometric intelligence as measured by SB:FE GPAB Composite scores. English must have been the first language. None of the subjects used in the study showed evidence of the following: inadequate visual or auditory acuity; neurological disease; severe cultural, educational, or environmental deprivation; or significant emotional disturbance.

In addition to Gifted (Very Superior and Superior), Normal Learners, Slow Learners, and Mentally Retarded groups, the LD subjects were broken into the following groups based on WRAT-R achievement scores and Rourke's research, which was reviewed in Chapter Two.

The first LD group comprised individuals who exhibited uniformly deficient performances on the Reading and Arithmetic subtests of the WRAT-R. The second LD group consisted of individuals with Reading Disabilities, with significantly higher Arithmetic achievement. The third

LD group comprised individuals with Arithmetic Disabilities with relatively higher and normal Reading achievement. This latter group is analogous to the Non-Verbal learning disabled group identified by Rourke (1978) and Rourke and Fuerst (1991).

It is important to reiterate that Rourke and colleagues emphasized patterns of reading and arithmetic scores, and their subtypes were not based purely on levels of performance. This reflects an underlying assumption that the bases for impaired performance in learning disabled populations are markedly different and related to cognitive ability profiles. These assumptions were explored in the present study.

CHAPTER FOUR: RESULTS

Part One

Presentation of the Results

The results of this dissertation are presented in Chapters Four, Five, Six, and Seven. In this chapter, the results of Part One are presented. Part One, which corresponds to the first objective of this study, provides comprehensive descriptive data about the SB:FE when used with a Canadian clinic sample. In Chapter Five, the relationship between the SB:FE full battery and abbreviated version, the SB:FE GPAB, is explored. The results of Part Three, which correspond to the third objective of this study, and in which multivariate clustering procedures were applied to the SB:FE GPAB data, are presented in Chapter Six. Chapter Seven, which corresponds to the fourth objective of this study, contains the results of the external validation procedures applied to the clusters derived through application of empirical clustering procedures to SB:FE GPAB data.

Demographic Characteristics of the Sample

Age. The sample comprised 1220 individuals referred to the Education Clinic, University of Alberta, for psychoeducational assessments. The mean age of the total sample was 9.44 years, the standard deviation was 4.03, and the age range covered the entire age span covered by the SB:FE, from 2 years through 23-11. A breakdown of the subjects' age according to the age groups used by Thorndike et al. (1986a) is presented in Table 6. The frequency of the remaining characteristics are also presented in Table 6. For each of the three age groups, chi-square analyses were performed to assess the consistency of the distributions across the age groups for each characteristic: In the accompanying discussion, the results of the chi-square test are reported only when it revealed a significant difference or lack of consistency.

Table 6

Demographic Characteristics, Age Group and Total Sample

	Age Groups							
	2-6-11		7-11-11		12-23-11		2-23-11	
Demographic Variable	n	%	n	%	n	%	n	%
Age Groups								
Age Distribution	394	32.3	527	43.2	299	24.5	1220	100.0
Gender								
Female	167	42.4	190	36.1	119	39.8	476	39.0
Male	227	57.6	337	63.9	180	60.2	744	61.0
Ethnicity								
Caucasian	360	91.4	490	93.0	269	90.0	1119	91.7
Asian	13	3.3	16	3.0	12	4.0	41	3.4
Native	14	3.6	12	2.3	12	4.0	38	3.1
Black	3	0.8	6	1.1	1	0.3	10	0.8
Other	4	1.0	3	0.6	5	1.7	12	1.0
Community Size								
Above 299,999	243	61.7	287	54.5	180	60.2	710	58.2
25,000 to 99,999	64	16.2	64	12.1	33	11.0	161	13.2
2,500 to 24,999	47	11.9	90	17.1	46	15.4	183	15.0
Less than 2,500	40	10.2	86	16.3	40	13.4	166	13.6
Primary Language								
English	386	98.0	509	96.6	273	91.3	1168	95.7
Bilingual	6	1.5	12	2.3	14	4.0	30	2.5
Non English	2	0.5	6	1.1	12	4.7	22	1.8
Parental Occupation								
Managerial/ Professional	186	47.2	143	27.1	77	25.8	406	33.3
Technical/Sales	44	11.2	88	16.7	39	13.0	171	14.0
Service	46	11.6	80	15.2	41	13.7	167	13.7
Farming/Forestry	14	3.6	35	6.6	20	6.7	69	5.7
Production	12	3.0	16	3.1	10	3.3	38	3.1
Operator/Fabricator	35	8.9	58	11.0	43	14.4	136	11.1
Unemployed	3	.8	10	1.9	7	2.3	20	1.6
Other	29	7.4	40	7.6	22	7.4	91	7.5
Not Known	25	6.3	57	10.8	40	13.4	122	10.0

continued

(Table 6 continued)

Parental Education								
College Graduate or more	174	44.2	131	24.8	64	21.7	369	30.2
1 to 3 years college	91	23.1	145	27.5	59	20.5	295	24.2
High School Graduate	49	12.4	91	17.3	50	16.7	190	15.6
Less Than High School	35	8.9	59	11.2	61	19.7	155	12.7
Not Known	45	11.4	101	19.2	65	21.4	211	17.3
Referral Reasons								
Self/Parental Interest	184	46.7	145	27.4	81	27.1	410	33.6
Giftedness	85	21.6	27	5.1	22	7.4	134	11.0
Learning Concerns	54	13.7	276	52.4	124	41.5	454	37.2
School Placement	42	10.7	23	4.4	22	7.4	87	7.1
Emotional Problems	27	6.8	33	6.3	17	5.6	77	6.3
Mentally Deficient	2	0.5	23	4.4	33	11.0	58	4.8

Gender Distribution. The sample consisted of 476 females and 744 males, approximately 39% female and 61% male. Throughout the three age groups identified by Thorndike et al. (1986a) and used in this study the ratio of males to females was approximately one and a half to one. This is not unreasonable, as typically males tend to be overrepresented in groups of school aged children referred for psychological assessments and learning difficulties (Brodie, 1988; MacGregor, Rosenbaum, & Skoutajan, 1982; Ozols & Rourke, 1991; Winzer, 1993). The gender distribution appears in Table 6.

Ethnicity. As shown in Table 6, the ethnic groups represented in the sample tend to be fairly diverse including not only Native Canadians and Caucasians, but also a few individuals of Black and Asian descent. In general, however, the sample is predominantly Caucasian.

Community size. As can be seen from Table 6, there appeared to be an adequate representation of individuals from various community sizes within Alberta, with the largest category comprising subjects from urban centres. A significant χ^2 (6, 1220) = 16.9, $p < .01$, was obtained, suggesting that the frequencies of subjects from different community sizes differed across age ranges. This difference, however, appears most prevalent between the age group 2-

6-11 and 7-11-11. Chi-square analyses conducted between the age groups 7-11-11 and 12-23-11, and 2-6-11 and 12-23-11, were not significant.

Primary language. As seen in Table 6, the largest group of individuals were from families that reported English as their primary language.

Parental occupation. The parental occupation of subjects is reported in Table 6. The highest occupational level of either parent was recorded, consistent with the procedure adopted in the standardization of the SB:FE (Thorndike et al., 1986a). In general, there appears to be a slight overrepresentation of managerial/professional occupations. This trend is consistent with the demographic data obtained in the original SB:FE standardization (Thorndike et al., 1986b). A weighting system was employed to overcome this discrepancy in the norming of the test (Thorndike et al., 1986a). In addition, the present sample is consistent with that obtained by Fritzke (1988) and Spelliscy (1991) in their study of a similar clinic population. The "unknown" category is fairly large, but this reflects a lack of available data. A significant χ^2 (16, 1220) = 62.5, $p < .001$, was obtained, suggesting that the frequencies of parental occupations differed across age ranges. Most noticeably, a larger percentage of parents in the age range 2-6-11 appeared to come from managerial or professional occupations.

Parental education levels. The parental educational level of subjects is reported in Table 6. The highest educational level of either parent was recorded, consistent with the procedure adopted in the standardization of the SB:FE (Thorndike et al., 1986a). With respect to education levels there is possibly an overrepresentation of subjects whose parents were college graduates or had one to three years of college education. This is consistent with the trends noted earlier with respect to occupational status. A significant χ^2 (8, 1220) = 78.5, $p < .001$, was obtained, suggesting that the frequencies of highest level of parental education differed across age ranges. Most noticeably, a larger frequency of parents in the age range 2-6-11 appeared to be college graduates or more, and fewer frequencies were reported in the less than high school category. Interestingly the major referral concern in this age group related to parental interest and giftedness.

Reason for referral. The reasons for referral within the sample is listed in Table 6. The reasons for referral are diverse. As shown, with the exception the youngest age, 2-6-11, where the most frequent reason for referral reason was Parental Interest, the most frequent reason for referral was concerns with learning by the subjects, highest within the age group 7-11-11 at 52.4%. The one exception, in the age group 2-6-11, is not surprising given that the majority of subjects in this age group have not yet entered school. Consequently the subjects have not yet had an opportunity to experience difficulty in learning. As the subjects within the age groupings get older, the number of referrals dealing with questions surrounding Mental Deficiency rise from .3% at age 2-6-11, to around 4% at ages 7-11-11 and 12-23-11. In general, these differences are statistically significant and a X^2 (8, 1220) = 235.3, $p < .001$, was obtained, indicating that the reasons for referral vary with the age group sampled. Overall, the sample presents with a diverse range of referral concerns suggesting that while this is indisputably a clinic population, the sample comprises individuals with a diversity of presenting concerns and likely represents a wide range of ability levels. Further, the frequencies of reasons for referral are similar to those reported by Spelliscy (1991) and Fritzke (1988).

Performance of Age Groups on the SB:FE

Presented in Table 7 are the means and standard deviations for each age group and the total sample age 2-23-11. Examination of the means for the entire sample suggests that overall the mean scores for the individual subtests, the four Reasoning Area SAS's, and Composite SAS tend to be somewhat higher than those reported in the standardization sample (Thorndike et al., 1986a). The Copying subtest appears to be a clear exception, particularly for the two older age groups. With respect to variability, the standard deviations of subtests and Reasoning Area and Composite SAS's tend to be slightly higher than those found in the standardization sample, which might reflect the composition of the sample given the wide range of referral concerns. Exceptions to this trend generally emerge on subtests that were administered to fewer subjects, for example Verbal Relations and Paper Folding and Cutting.

Table 7

Means and Standard Deviations for Each Age Group and the Total Sample

	Age Groups											
	2-6-11			7-11-11			12-23-11			2-23-11		
SB:FE Subtest	n	M	SD	n	M	SD	n	M	SD	n	M	SD
Vocabulary	394	55.0	7.1	527	50.3	7.9	299	49.1	9.1	1220	51.6	8.4
Comprehension	394	55.0	7.6	527	52.0	8.4	299	49.2	9.7	1220	52.3	8.8
Absurdities	389	55.2	6.3	511	51.2	8.5	208	48.4	7.8	1091	52.1	8.0
Verbal Relations	-	-	-	-	-	-	118	53.4	6.6	118	53.5	6.6
Pattern Analysis	394	54.0	9.3	527	51.1	8.7	299	49.3	8.8	1220	51.6	9.1
Copying	357	51.3	8.3	439	46.0	8.4	139	42.3	8.2	935	47.5	8.9
Matrices	48	56.9	3.8	322	52.7	7.1	257	49.2	8.4	627	51.6	7.8
PFC	-	-	-	-	-	-	101	55.3	7.2	101	55.4	7.2
Quantitative	394	54.1	7.2	527	48.6	7.4	299	46.5	10.4	1220	49.9	8.9
Number Series	16	57.6	5.2	283	52.7	7.4	223	50.2	7.8	522	51.8	7.7
Equation Building	-	-	-	-	-	-	77	57.7	9.4	77	57.7	9.5
Bead Memory	394	51.2	8.9	527	48.1	9.5	299	47.3	10.2	1220	48.9	9.6
MFS	394	52.3	8.4	527	48.9	8.6	299	45.9	9.5	1220	49.3	9.1
Memory for Digits	135	55.0	5.9	449	50.3	7.2	276	48.1	8.2	860	50.5	7.7
Memory for Objects	108	55.7	6.2	423	51.9	7.1	263	48.9	7.8	794	51.4	7.6
Verbal SAS	394	111.5	13.5	527	102.8	16.2	299	98.3	18.6	1220	108.0	16.8
A/V SAS	394	106.8	17.2	527	99.4	16.7	299	96.5	19.2	1220	101.1	17.9
Quantitative SAS	394	108.4	15.7	527	99.4	15.7	299	95.5	21.3	1220	101.4	17.6
STM SAS	394	105.6	16.2	527	98.9	16.9	299	93.6	19.3	1220	99.8	17.8
Composite SAS	394	109.6	14.9	527	100.2	16.4	299	95.4	20.2	1220	102.1	17.9

Note: Standardization Subtest SAS M=50, SD=8; Reasoning Area and Composite SAS M=100, SD=16.

It is acknowledged that the sample used in this study cannot be considered representative of a normal population and as such cannot be compared directly with the SB:FE normative sample which was designed to be a nationally representative sample based on U.S. 1980 census data (Thorndiket et al., 1986b). For this reason, statistical tests to compare means and standard deviations of this sample with the sample tested by Thorndike et al. (1986b) were not conducted.

Qualitative comparisons are made, however, to highlight differences and provide further understanding of this particular sample.

Comparability of Age Samples

Distinct and statistically significant differences emerge when the sample is examined according to the three age groups. Presented in Table 8, are the results of statistical tests conducted to compare the means of the SB:FE subtests, Reasoning Area and Composite SAS's across each of the three age samples. Ideally, multivariate comparisons should have been conducted. Due to the adaptive administration procedures (Thorndike et al., 1986c), however, the sample sizes across each of the subtests within each age group vary. It was decided not to eliminate subjects to achieve equal numbers in that it was deemed important to keep the sample intact across all analyses, thereby maintaining consistency. Consequently, a multivariate analysis was not performed. Within the univariate paradigm, there is a risk, given the number of tests, of finding a number of differences that exceed that which would be expected by chance. Thus, to help control this, a .01 level of significance was selected. At this level, 1 out of 100 tests would be expected to yield significant results by chance alone, given the null or test hypothesis was true. But, as can be seen from Table 8, the number of significant results obtained far exceeded the number expected by chance. Thus, these results suggest that there were reliable differences among the mean scores obtained by each of the three age groups.

Subsample Age 2-6-11. In the sample age 2-6-11, subtest, Reasoning Area, and Composite SAS's scores are typically higher than those reported by Thorndike et al. (1986b). Mean subtest scores are up to six points higher, while Reasoning Area scores range from five to 11 points higher. The Composite SAS score within this age range is approximately nine points higher than those reported by Thorndike et al. (1986a). Subtest standard deviations generally ranged from one to two points lower to just over one point higher, although these differences are perhaps most marked, with the exception of the Absurdities subtest, on those subtests administered to fewest subjects. This latter observation is particularly true for the Matrices and Number Series subtests. However, these two subtests are not routinely administered to subjects within this age range (Thorndike et al., 1986b) and data concerning these two subtests within this age range must be regarded with caution. Standard deviations within the four Reasoning Areas range from two points below the standardization mean to one point above. As noted, in this age group a large percentage of subjects are referred for assessment of possible giftedness which may account for the higher scores (cf., Table 6). This point will be addressed again in discussion results of the other age groups.

Table 8

ANOVA and Univariate Differences Between Means of Each Age Group

SB:FE Test	F Ratio (2,604)	Post hoc comparisons 1 vs 2	1 vs 3	2 vs 3
Vocabulary	56.41	*	*	NS
Comprehension	39.79	*	*	*
Absurdities	57.62	*	*	*
Verbal Relations	–	a	a	a
Pattern Analysis	24.04	*	*	NS
Copying	70.41	*	*	*
Matrices	28.80	*	*	*
PFC	–	a	a	a
Quantitative	80.66	*	*	*
Number Series	11.59	NS	*	*
Equation Building	–	a	a	a
Bead Memory	17.13	*	*	*
MFS	45.90	*	*	NS
Memory for Digits	38.96	*	*	NS
Memory for Objects	35.38	*	*	*
Verbal SAS	62.83	*	*	NS
A/V SAS	33.99	*	*	*
Quantitative SAS	17.14	*	*	*
STM SAS	42.24	*	*	*
Composite SAS	64.72	*	*	*

* denotes pairs of means significantly different at the $p < .01$ level; NS denotes pairs of means that were not significantly different; a denotes subtests administered to only one of these age groups; Group 1=Age Group 5-6-11; Group 2=Age Group 7-11-11; Group 3=Age Group 12-23-11.

Subsample Age 7-11-11. Subtest and Reasoning Area and Composite SAS scores for the sample age 7-11-11 tend to be somewhat lower than those reported for the younger age group, and generally within one to two points of the values reported by Thorndike et al. (1986b). Scores on the Comprehension and Matrices subtests were between two and three points above the standardization mean. The Copying subtest again was the lowest of the SB:FE subtests within this age range, with Bead Memory and Quantitative subtests also lower than the standardization mean reported by Thorndike et al. (1986b). The Verbal Reasoning Area was almost three points above the standardization mean.

Subsample Age 12-23-11. Subtest and Reasoning Area, and Composite SAS scores for the sample age 12-23-11 tend to be somewhat lower than those reported for the two previous age groups and moreover, generally lower than the values reported by Thorndike et al. (1986a). The Comprehension and Matrices subtests were between two and three points above the standardization mean. The Copying subtest again was the lowest of the SB:FE subtests within this age range, almost eight points below the standardization sample mean (Thorndike et al., 1986b), with Quantitative, Bead Memory, and Memory for Sentences subtests also lower than the standardization mean. Paper Folding and Cutting and Equation Building subtests were exceptions to this trend, and between five to almost 8 points over the standardization mean respectively. It is possible that only the more capable subjects in this sample were administered these subtests. Thorndike et al. (1986a) recommended these subtests be used in the assessment of students for gifted programs. Typically, the four Reasoning Area SAS's were almost two points lower than the standardization mean in the Verbal Reasoning Area, to just over six points lower than the normative sample mean in the Short Term Memory Area. The Composite SAS was approximately five points below the mean reported by Thorndike et al. (1986b).

Results of the total sample age 2-23-11, tend to be confounded by the between age differences scores each of the three age groups adopted by Thorndike et al. (1986b). They are presented in Table 7 in order to enable comparison with data presented by Thorndike et al. (1986b) in the SB:FE Technical Manual.

In sum, as presented in Table 8, there appears to be a noticeable decline in subtest scores and Reasoning Area and Composite SAS's as the age of this clinic sample increased. Of the 51 comparisons between means of these variables across the three age ranges, 45 were significant. In each case, the differences revealed a decline in SB:FE scores as age increased. In part this may reflect different referrals concerns at different ages and thus the sample composition (cf., Table 6), although it must be noted that referral differences were most marked between the age group 2-6-11 and the two older age groups. Differences in frequencies of reasons for referral were not as marked between the two older groups. Sattler (1992) noted that the SB:FE fails to provide the same range of scores throughout the age levels covered by the scales. For example, Sattler (1992) noted that Composite SAS scores of up to 160 can be obtained from age 2 years through to about age 13 years. However, after this age the maximum scores begin to drop. Sattler (1992) noted that maximum composite scores at age 12 years and age 18 years may differ by about 15 points even if performance on both occasions is flawless. Conversely, the SB:FE appears to have a limited floor, most marked at younger ages. For example, a Composite SAS score as low as 36 cannot be obtained at age 2-00, although this is possible in the older age ranges. Sattler (1992) reported that the SB:FE cannot be used to diagnose Mental Retardation at age two or three for example. Thus, the trend noted in this sample may not only reflect specific sample composition and referral concerns, but may also reflect inherent limitations in the SB:FE. This suggests that caution may be needed in interpreting SB:FE scores at different age groups, particularly when subjects at the higher and lower levels in the score range are being interpreted. This situation will most likely occur when educational placement decisions are being made.

Intercorrelations Among Subtests, Reasoning Areas, and Composite SAS's

As has been discussed in the review of the literature presented in Chapter Two, the construct validity of the SB:FE has been the subject of much debate. Intercorrelations provide information about the relationships between each of the SB:FE subtests and the Reasoning Area and Composite SAS's. These correlations are provided separately for each age group in Tables 9, 10, 11, and 12. Consistent with the results presented by Thorndike et al. (1986b), subtests at all ages generally correlated moderately to highly positively with their Composite scores. As expected, the correlations between a subtest and the Reasoning Area SAS of which it was a part, were stronger than the correlations with the other Reasoning Area SAS's. Correlations among Reasoning Areas were also highly positive, as were correlations between these variables and Composite SAS's. Taken together, the pattern of correlations is generally supportive of the hypothesized structure posited by Thorndike et al. (1986b).

Intercorrelations for the total sample 2-23-11. These are presented in Table 9. Intercorrelations between subtests in the sample 2-23-11 ranged from a low of .22 between Paper Folding and Cutting and Memory for Objects to a high of .72 between the Comprehension and Vocabulary subtests. Low correlations of .09 and .18 were noted between the Copying subtest and Equation Building (n=22), and Paper Folding and Cutting (n=34) respectively. However, subtest intercorrelations are more typically in the range .4 to .6 and the overall patterns observed are consistent with those reported by Thorndike et al. (1986b). Correlations between a subtest and the Reasoning Area SAS of which it was a part were stronger than the correlations with the other Reasoning Area SAS's. The correlations between subtests and Composite SAS 's ranged from .66 to .90. Correlations between the four Reasoning Areas SAS's ranged from .68 to .73. Correlations between Reasoning Area SAS's and the Composite SAS ranged from .87 to .88. The data presented for the total sample, age 2-23-11, may be of limited utility. However, they are presented in this study as many of the factor analytic studies supporting the constructs hypothesized by Thorndike et al. (1986b) are based on this sample age range. However, analyses of the correlation matrices for the three age groups are more meaningful, as results do not appear to be confounded by between age differences as in the total sample.

Table 9

Subtest, Reasoning Area, and Composite SAS Intercorrelations for the SB:FE Total Sample Age 2-23-11

	Verbal Reasoning Area				Abstract/Visual Reasoning Area				Quantitative Reasoning Area			Short Term Memory Area				Area and Composite SAS's				
Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Voc. (1)	–	.72	.61	.67	.51	.50	.59	.51	.59	.63	.49	.52	.66	.52	.46	.90	.61	.63	.68	.80
Comp.(2)	1220	–	.58	.51	.48	.47	.56	.47	.57	.57	.42	.49	.62	.49	.46	.89	.58	.60	.64	.77
Ab. (3)	1108	1108	–	.52	.49	.45	.55	.41	.51	.52	.29	.46	.50	.44	.47	.83	.56	.55	.57	.72
V.R. (4)	118	118	54	–	.37	.33	.55	.60	.55	.58	.53	.48	.47	.43	.27	.82	.54	.62	.55	.73
P.A. (5)	1220	1220	1108	118	–	.51	.54	.53	.56	.58	.47	.58	.46	.48	.46	.57	.87	.50	.61	.76
C. (6)	935	935	916	36	935	–	.43	.18	.52	.41	09	.43	.44	.48	.46	.55	.83	.53	.53	.70
Mat. (7)	627	627	525	116	405	405	–	.63	.60	.66	.53	.54	.56	.54	.48	.66	.80	.68	.67	.79
P.F.C. (8)	101	101	49	78	34	34	101	–	.66	.69	.65	.47	.30	.33	.22	.59	.84	.73	.44	.75
Qnt. (9)	1220	1220	1108	118	1220	935	627	101	–	.63	.67	.53	.53	.53	.45	.65	.64	.95	.63	.82
N.S. (10)	522	522	439	114	522	341	485	94	522	–	.68	.49	.50	.53	.44	.66	.70	.88	.64	.82
E.B.(11)	77	77	38	67	77	22	75	59	77	74	–	.36	.16	.32	.18	.52	.63	.88	.36	.74
B.M.(12)	1220	1220	1108	118	1220	935	627	101	1220	522	77	–	.50	.47	.47	.57	.61	.56	.81	.73
M.F.S. (13)	1220	1220	1108	118	1220	935	627	101	1220	522	77	1220	–	.63	.47	.68	.54	.55	.84	.74
M.F.D. (14)	860	860	771	115	860	620	586	98	860	503	74	860	860	–	.52	.56	.56	.57	.82	.71
M.F.O. (15)	794	794	715	110	794	578	558	95	794	484	72	794	794	762	–	.54	.53	.50	.76	.66
Ver. SAS (16)	1220	1220	1108	118	1220	935	627	101	1220	522	77	1220	1220	860	794	–	.68	.69	.73	.88
A/V. SAS (17)	1220	1220	1108	118	1220	935	627	101	1220	522	77	1220	1220	860	794	1220	–	.69	.68	.87
Qnt. SAS (18)	1220	1220	1108	118	1220	935	627	101	1220	522	77	1220	1220	860	794	1220	1220	–	.67	.87
STM. SAS (19)	1220	1220	1108	118	1220	935	627	101	1220	522	77	1220	1220	860	794	1220	1220	1220	–	.88
Com. SAS (20)	1220	1220	1108	118	1220	935	627	101	1220	522	77	1220	1220	860	794	1220	1220	1220	1220	–

Note: Upper triangular matrix contains correlation coefficients. Lower triangular matrix contains corresponding sample size.

Intercorrelations for the total sample 2-6-11. The correlation table for this age group appears in Table 10. Intercorrelations between subtests in the sample 2-6-11 ranged from a low of .02 between Absurdities and Matrices (n=46) to a high of .64 between the Comprehension and Vocabulary subtests. However, subtest intercorrelations were more typically in the range .3 to .4, with some around .5. The correlations between subtests and Composite SAS scores ranged from .51 to .78. Correlations between the four Reasoning Areas SAS's ranged from .45 to .61. Correlations between Reasoning Areas and the Composite SAS ranged from .77 to .83. Correlations among the three subtests in the Verbal Reasoning Area administered to subjects in this age range were moderate to moderately strong, with the correlation between the Vocabulary and Comprehension being the highest. Further, the correlations of these subtests with the Verbal Reasoning Area SAS were strong (.76-.85) and as expected, exceeded the correlations between these subtests and other Reasoning Areas which tend to be somewhat lower. Correlations among the three subtests within the Abstract/Visual Reasoning Area tended to be lower than those within the Verbal Reasoning Area with the correlation between Pattern Analysis and Copying being the highest. Correlations between these subtests and the Abstract/Visual Reasoning Area SAS were strong, particularly between Pattern Analysis and Copying subtests and this reasoning area (.54-.87) and as expected exceeded the correlations between these subtests and the remaining Reasoning Area scores. The correlation among the two subtests in the Quantitative Reasoning Area was weaker (.37) although this may be attributable to limited variability due very few subjects (n=16) completing the Number Series subtest. However, the correlations between these subtests and the Quantitative Reasoning Area SAS were strong (.83-.99) and higher than correlations between these subtests and the remaining Reasoning Area scores, as expected. Correlations among the four subtests in the Short Term Memory Area were moderate, with the highest correlation between Memory for Digits and Memory for Sentences. The correlation between these subtests and the Short Term Memory Reasoning Area SAS were moderate to strong (.67-.82) and higher than correlations between these subtests and the remaining Reasoning Area scores.

Table 10

Subtest, Reasoning Area, and Composite Score Intercorrelations for the SB:FE Sample Age 2-6-11

	Verbal Reasoning Area				Abstract/Visual Reasoning Area				Quantitative Reasoning Area			Short Term Memory Area				Area and Composite SAS's				
Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Voc. (1)	–	.64	.45	–	.32	.36	.14	–	.39	.53	–	.40	.56	.44	.24	.85	.40	.40	.56	.67
Comp.(2)	394	–	.46	–	.31	.35	.35	–	.38	.30	–	.33	.52	.41	.16	.87	.37	.38	.49	.64
Ab. (3)	389	389	–	–	.26	.31	.02	–	.37	.03	–	.36	.38	.34	.32	.76	.32	.37	.46	.59
V.R. (4)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
P.A. (5)	394	394	389	–	–	.47	.29	–	.48	.44	–	.45	.32	.43	.36	.37	.87	.48	.47	.71
C. (6)	357	357	353	–	357	–	.27	–	.45	.22	–	.36	.35	.40	.33	.42	.83	.45	.43	.68
Mat. (7)	48	48	46	–	48	43	–	–	.27	.42	–	.50	.51	.46	.35	.27	.54	.28	.55	.51
P.F.C. (8)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Qnt. (9)	394	394	394	–	394	394	48	–	–	.37	–	.47	.39	.34	.33	.47	.53	.99	.50	.78
N.S. (10)	16	16	15	–	16	14	16	–	16	–	–	.24	.38	.35	.28	.43	.51	.83	.43	.63
E.B.(11)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
B.M.(12)	394	394	389	–	394	357	48	–	394	16	–	–	.44	.37	.45	.45	.47	.47	.82	.69
M.F.S. (13)	394	394	389	–	394	394	48	–	394	16	–	394	–	.53	.36	.60	.40	.43	.82	.69
M.F.D. (14)	135	135	135	–	135	123	44	–	135	14	–	135	135	–	.29	.50	.34	.35	.69	.62
M.F.O. (15)	108	108	108	–	108	100	37	–	108	13	–	108	108	108	–	.30	.40	.33	.67	.55
Ver. SAS (16)	394	394	394	–	394	394	48	–	394	16	–	394	394	135	108	–	.45	.47	.61	.77
A/V. SAS (17)	394	394	394	–	394	394	48	–	394	16	–	394	394	135	108	394	–	.53	.53	.80
Qnt. SAS (18)	394	394	394	–	394	394	48	–	394	16	–	394	394	135	108	394	394	–	.50	.78
STM SAS (19)	394	394	394	–	394	394	48	–	394	16	–	394	394	135	108	394	394	394	–	.83
Com. SAS (20)	394	394	394	–	394	394	48	–	394	16	–	394	394	135	108	394	394	394	394	–

Note: Upper triangular matrix contains correlation coefficients. Lower triangular matrix contains corresponding sample size.

Intercorrelations among the sample 7-11-11. The correlation matrix is presented in Table 11. Intercorrelations between subtests in the sample 7-11-11 ranged from a low of .27 between Copying and Matrices to a high of .69 between the Comprehension and Vocabulary subtests. However, subtest intercorrelations were more typically in the range .3 to .5. Correlations between subtests and Composite SAS scores ranged from .61 to .78. Correlations between the four Reasoning Areas SAS's ranged from .66 to .72. Correlations between Reasoning Area SAS's and the Composite SAS ranged from .85 to .89. Correlations among the three subtests in the Verbal Reasoning Area administered to subjects in this age range were moderate to moderately strong, with the correlation between the Vocabulary and Comprehension being the highest (.69). Further, the correlations of these subtests with the Verbal Reasoning Area SAS were strong (.82-.88) and as expected, exceeded the correlations between these subtests and other Reasoning Areas. Correlations among the three subtests within the Abstract/Visual Reasoning Area tend to be lower than those within the Verbal Reasoning Area with the correlation between Pattern Analysis and Copying being the highest, although moderate only at .47. Correlations between the three subtests within this area and the Abstract/Visual Reasoning Area SAS were strong (.76-.86) and exceeded the correlations between these subtests and the remaining Reasoning Area scores. The correlation between the two subtests in the Quantitative Reasoning Area was moderate (.48). However, the correlation between these subtests and the Quantitative Reasoning Area SAS are strong (.86-.92) and exceeded the correlations between these subtests and the remaining Reasoning Area scores, as expected. Correlations among the four subtests in the Short Term Memory Area were moderate, with the highest correlation between Memory for Digits and Memory for Sentences (.58). The correlations between these subtests and the Short Term Memory Reasoning Area SAS were strong (.72-.81) and exceeded the correlations between these subtests and the remaining Reasoning Area scores.

Table 11

Subtest, Reasoning Area, and Composite Score Intercorrelations for the SBFE Sample Age 7-11-11

	Verbal Reasoning Area				Abstract/Visual Reasoning Area				Quantitative Reasoning Area			Short Term Memory Area				Area and Composite SAS's				
Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Voc. (1)	–	.69	.57	–	.53	.46	.51	–	.53	.51	–	.49	.63	.47	.39	.88	.62	.59	.66	.78
Comp.(2)	527	–	.52	–	.49	.44	.46	–	.49	.47	–	.47	.60	.42	.40	.87	.58	.56	.63	.76
Ab. (3)	511	511	–	–	.51	.35	.50	–	.44	.50	–	[illegible]	.46	.38	.41	.82	.57	.52	.54	.71
V.R. (4)	–	–	–	–	–	–	–	–	–	–			–	–	–	–	–	–	–	–
P.A. (5)	527	527	511	–	–	.47	.29	–	.52	.51	–	.58	.46	.41	.37	.60	.86	.59	.62	.76
C. (6)	439	439	428	–	439	–	.27	–	.45	.33	–	.38	.37	.38	.35	.49	.80	.48	.48	.64
Mat. (7)	322	322	309	–	322	255	–	–	.48	.60	–	.49	.46	.43	.45	.59	.76	.60	.61	.74
P.F.C. (8)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Qnt. (9)	527	527	511	–	527	439	322	–	–	.48	–	.52	.48	.45	.36	.58	.60	.92	.60	.76
N.S. (10)	283	283	277	–	283	227	254	–	283	–	–	.40	.42	.52	.39	.60	.62	.86	.58	.77
E.B.(11)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
B.M.(12)	527	527	511	–	527	439	322	–	527	283	–	–	.45	.43	.38	.55	.61	.54	.79	.72
M.F.S. (13)	527	527	511	–	527	439	322	–	527	283	–	527	–	.58	.42	.66	.52	.52	.81	.72
M.F.D. (14)	449	449	441	–	449	367	300	–	449	274	–	449	449	–	.46	.50	.50	.53	.79	.67
M.F.O. (15)	423	423	418	–	423	349	292	–	423	264	–	423	423	404	–	.48	.46	.44	.72	.61
Ver. SAS (16)	527	527	511	–	527	439	322	–	527	283	–	527	527	449	423	–	.69	.66	.72	.88
A/V. SAS (17)	527	527	511	–	527	439	322	–	527	283	–	527	527	449	423	527	–	.68	.69	.87
Qnt. SAS (18)	527	527	511	–	527	439	322	–	527	283	–	527	527	449	423	527	527	–	.67	.85
STM SAS (19)	527	527	511	–	527	439	322	–	527	283	–	527	527	449	423	527	527	527	–	.89
Com. SAS (20)	527	527	511	–	527	439	322	–	527	283	–	527	527	449	423	527	527	527	527	–

Note. Upper triangular matrix contains correlation coefficients. Lower triangular matrix contains corresponding sample size.

Intercorrelations among the sample 12-23-11. The correlation matrix is presented in Table 12. The lowest intercorrelations between subtests in the sample 12-23-11 were .09 and .18 between the Copying subtest and Equation Building (n=22) and Paper Folding and Cutting (n=34) respectively. The highest intercorrelation was .77, between the Comprehension and Vocabulary subtests. However, subtest intercorrelations were more typically in the range .5 to .6. The correlations between subtest and Composite SAS scores ranged from .67 to .86. Correlations between the four Reasoning Areas SAS's ranged from .77 to .81. Correlations between Reasoning Area SAS's and the Composite SAS ranged from .89 to .92. Correlations among the three subtests in the Verbal Reasoning Area administered to subjects in this age range were moderate to moderately strong, with the correlation between the Vocabulary and Comprehension being the highest (.77). Further, the correlations of these subtests with the Verbal Reasoning Area SAS were strong (.82-.93) and as expected, exceeded the correlations between these subtests and other Reasoning Areas. Correlations among the three subtests within the Abstract/Visual Reasoning Area tended to be slightly lower than those within the Verbal Reasoning Area, with the correlation between Matrices and Paper Folding and Cutting being the highest at .63. However, the intercorrelation between Copying and Paper Folding and Cutting was weak at .18. Correlations between the four subtests within this area and the Abstract/Visual Reasoning Area SAS were generally strong (.81-.88) and exceeded the correlations between these subtests and the remaining Reasoning Area scores. The correlations among the three subtests in the Quantitative Reasoning Area were moderately strong (.67-.75). The correlation between these subtests and the Quantitative Reasoning Area SAS were strong (.89-.95) and exceeded the correlations between these subtests and the remaining Reasoning Area scores, as expected. Correlations among the four subtests in the Short Term Memory Area are moderate, with the highest correlation between Memory for Digits and Memory for Sentences (.65). The correlations between these subtests and the Short Term Memory Reasoning Area SAS were strong (.77-.84) and exceeded the correlations between these subtests and the remaining Reasoning Area scores.

Table 12

Subtest, Reasoning Area, and Composite Score Intercorrelations for the SB:FE Sample Age 12-23-11

	Verbal Reasoning Area				Abstract/Visual Reasoning Area				Quantitative Reasoning Area			Short Term Memory Area				Area and Composite SAS's				
Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Voc. (1)	–	.77	.72	.67	.62	.54	.66	.51	.71	.64	.49	.61	.70	.53	.53	.93	.72	.75	.73	.86
Comp.(2)	299	–	.68	.51	.58	.50	.62	.41	.70	.64	.42	.59	.64	.51	.52	.91	.68	.74	.71	.83
Ab. (3)	208	208	–	.52	.63	.58	.62	.60	.60	.52	.29	.53	.55	.43	.46	.87	.71	.63	.61	.78
V.R. (4)	118	118	54	–	.37	.33	.55	.59	.55	.58	.53	.48	.47	.43	.28	.82	.54	.61	.55	.73
P.A. (5)	299	299	208	118	–	.57	.60	.53	.63	.63	.47	.66	.52	.52	.53	.68	.88	.68	.69	.80
C. (6)	139	139	135	36	439	–	.46	.18	.47	.48	.09	.53	.47	.51	.49	.59	.81	.51	.59	.68
Mat. (7)	257	257	170	116	257	107	–	.63	.67	.69	.53	.63	.60	.58	.45	.72	.86	.74	.70	.83
P.F.C. (8)	101	101	49	78	101	34	101	–	.66	.69	.65	.47	.30	.33	.22	.59	.84	.73	.44	.75
Qnt. (9)	299	299	208	118	299	139	257	101	–	.75	.67	.56	.61	.58	.46	.76	.72	.95	.69	.86
N.S. (10)	223	223	147	114	223	100	216	94	223	–	.68	.60	.56	.51	.44	.72	.77	.91	.67	.87
E.B.(11)	77	77	36	67	77	22	75	59	77	74	–	.36	.16	.32	.18	.52	.63	.89	.36	.74
B.M.(12)	299	299	208	118	299	139	257	101	299	223	77	–	.57	.52	.55	.66	.71	.63	.83	.77
M.F.S. (13)	299	299	208	118	299	139	257	101	299	223	77	299	–	.65	.50	.72	.61	.63	.84	.74
M.F.D. (14)	276	276	197	115	276	130	242	98	276	215	74	276	449	–	.56	.57	.60	.60	.83	.71
M.F.O. (15)	263	263	189	110	263	129	229	95	263	207	72	263	423	258	–	.57	.57	.51	.77	.67
Ver. SAS (16)	299	299	299	299	299	139	257	101	299	223	77	299	527	276	263	–	.77	.81	.78	.92
A/V. SAS (17)	299	299	299	299	299	139	257	101	299	223	77	299	527	276	263	299	–	.79	.77	.91
Qnt. SAS (18)	299	299	299	299	299	139	257	101	299	223	77	299	527	276	263	299	299	–	.74	.92
STM SAS (19)	299	299	299	299	299	139	257	101	299	223	77	299	527	276	263	299	299	299	–	.89
Comp. SAS (20)	299	299	299	299	299	139	257	101	299	223	77	299	527	276	263	299	299	299	299	–

Note: Upper triangular matrix contains correlation coefficients. Lower triangular matrix contains corresponding sample size.

WRAT-R Achievement Scores

The data available for the purposes of this study are somewhat complicated. Not all of the 1220 subjects for whom SB:FE data had been administered the WRAT-R achievement test. The WRAT-R is not administered to subjects below age 5-00. Further, it is not always appropriate to administer the WRAT-R to all subjects over the age of 5 years given the presenting reasons for referral are not the same (cf., Table 6). Consequently as can be seen in Table 13, the proportion of subjects at each age group with both SB:FE and WRAT-R data increased with age: from 0.0% (2-4-11) to 42.2% (5-6-11) to 57% (7-11-11) to 62.2% (12-23-11). The means and standard deviations of the three WRAT-R subtests for each of the three age groups for whom WRAT-R data were available are reported in Table 14.

Table 13

Breakdown of Subjects With and Without WRAT-R Achievement Data by Age Category

	SB:FE Data Only		SB:FE/WRAT-R Data		Total	
Age Groups	n	%	n	%	n	%
2-4-11	110	100.0	N/A	0.0	110	100.0
5-6-11	164	57.8	120	42.2	284	100.0
7-11-11	226	42.9	301	57.1	527	100.0
12-23-11	113	37.8	186	62.2	299	100.0
2-23-11	613	50.2	607	49.8	1220	100.0

Note: The WRAT-R is only administered to subjects age 5-00 and above.

WRAT-R data for the sample 5-6-11. As can be seen in Table 14, the WRAT-R scores for the age group 5-6-11 display greater variability than the standardization sample (Jastak & Wilkinson, 1984). The Reading and Arithmetic subtest means are between three to five points higher than the standardization sample (Jastak & Wilkinson, 1984). These results are consistent with the earlier finding (cf., Table 7) that the means on the corresponding SB:FE Reasoning Area and Composite SAS scores exceeded the normed mean. The mean scores for the Spelling subtest are consistent with the standardization sample mean. In addition, the Reading and Arithmetic subtest scores in this age group are in keeping with the SB:FE Reasoning area and Composite SAS's, which were also above the standardization mean.

Table 14

WRAT-R Means and Standard Deviations for Each Age Group

	Age Groups								
	5-6-11			7-11-11			12-23-11		
WRAT-R Subtests	n	M	SD	n	M	SD	n	M	SD
Reading	120	103.3	22.3	301	93.1	19.8	186	93.1	18.5
Spelling	120	100.5	19.7	301	89.4	18.4	186	89.4	17.9
Arithmetic	120	105.5	20.4	301	91.0	15.7	186	88.3	21.2

Note: WRAT-R standardization M=100, SD=15 (Jastak & Wilkinson, 1984).

WRAT-R data for the sample 7-11-11. For this age group, the mean scores for the Reading, Spelling, and Arithmetic subtests are clearly below the mean of the standardization sample (see Table 14). Further, they are typically below the means for the corresponding norm values of SB:FE Reasoning Area and Composite SAS's (cf., Table 7). For the Reading and Spelling subtests on the WRAT-R, the variability of scores within this age group exceeds the norm values reported by Jastak and Wilkinson (1984).

WRAT-R data for the sample 12-23-11. For this age, mean scores for the Reading, Spelling, and Arithmetic subtests are again clearly less than the mean of the standardization sample and in addition are below the means for the corresponding norm values of SB:FE Reasoning Area and Composite SAS's (cf., Table 7). However, for all three WRAT-R subtests, the variability of scores within this age group exceeds the norm values reported by Jastak and Wilkinson (1984).

Comparisons among the three age groups. Scores in the age group 12-23-11 are in keeping with the scores obtained by the subjects in the age group 7-11-11, but both are lower than the scores obtained by the subjects in the age group, 5-6-11. However, this trend toward lower scores may reflect the sample composition and the relatively greater frequency of referral concerns specifically related to learning difficulties at school within these older samples (cf., Table 6). Distinct and statistically significant differences emerge when the sample is examined according to the three age groups. Presented in Table 15, are the results of statistical tests conducted to compare the means of the WRAT-R subtests. Multivariate tests followed by post hoc comparisons were conducted. A .01 level of significance was selected. Out of the nine post hoc comparisons, six were statistically significant, which exceeded the number expected by chance. These results suggest that there were reliable differences among the mean scores obtained by the age group 5-6-11 and 7-11-11, and the age group 5-6-11 and 12-23-11. There were, however, no significant differences between the scores obtained by the samples age 7-11-11 and 12-23-11, which suggested the clear separation between all three age groups observed on the SB:FE may not be present in the subjects that completed the WRAT-R.

Table 15

<u>MANOVA, Univariate F Ratios, and Significance of Differences Between WRAT-R Means for Each Age Group</u>

		Age Groups		
WRAT-R Subtests	F ratio	1 vs 2	1 vs 3	2 vs 3
Reading	12.65	**	**	NS
Spelling	17.66	**	**	NS
Arithmetic	35.22	**	**	NS

Multivariate Main Effect:
$F(6, 1204) = 12.17, p < .001$

* denotes pairs of means significantly different at the $p < .05$ level; ** denotes pairs of means significantly different at the $p < .01$ level; NS denotes pairs of means that were not significantly different; Group 1=Age Group 5-6-11; Group 2=Age Group 7-11-11; Group 3=Age Group 12-23-11.

Summary of the Results of Chapter Four

In this chapter, comprehensive demographic and descriptive data about the SB:FE when used with a Canadian clinic sample were presented for each of the age samples 2-6-11, 7-11-11, 12-23-11, and the total sample. The ratio of males to females in the total sample (N=1220) was approximately 3 to 2 across the age levels sampled. The sample was predominantly Caucasian. English was overwhelmingly the primary language spoken by subjects in the sample. Most notably, subjects in the age sample 2-6-11 generally had higher frequencies of parents from managerial/professional occupations and with a college education or more. The most frequent reason for referral within this age group was parental interest. In contrast, the number of referrals dealing with mental retardation and learning problems increased with age. With respect to performance on the SB:FE subtests and Reasoning Area and Composite SAS's, means generally decreased with increasing age. While the respective means for the age sample 2-6-11 were generally higher than the corresponding standardization means (Thorndike et al., 1986b) the SB:FE scores in the age sample 12-23-11 were slightly below the standardization means. The differences in means across age levels sampled may be attributable either to floor effects in the lower age range and ceiling effects in the higher age range of the SB:FE, the different frequencies of reasons for referral across the sample, or to a combination of these factors.

The intercorrelations among the SB:FE subtest and Reasoning Area and Composite SAS's in this sample were reasonably consistent across the age groups sampled and appear to be similar to those reported by Thorndike et al. (1986b). The results for the total sample are not as clear as the results obtained for the three age groups, but in general, the intercorrelations tend to be supportive of the hypothetical constructs underlying the SB:FE posited by Thorndike et al. (1986b), specifically with respect to the four proposed reasoning or cognitive ability areas that form the second level of the SB:FE. As expected the correlations between a subtest and the Reasoning Area SAS of which it was a part were stronger than the correlations with the other Reasoning Area SAS's. Evidence also supported the SB:FE as a measure of g or overall ability.

Results do not support the existence of the Fluid-Analytic and Crystallized ability dimensions posited by Thorndike et al. (1986b).

Finally, the performance of the age samples 5-6-11, 7-11-11, and 12-23-11 on the WRAT-R were examined. WRAT-R Reading, Spelling, and Arithmetic scores in the sample age 5-6-11 were consistent with the standardization means (Jastak & Wilkinson, 1984), but slightly lower than the SB:FE scores obtained by this sample. WRAT-R scores in the age samples 7-11-11 and 12-23-11 were lower than the standardization means and, in addition, were significantly lower than the respective WRAT-R scores obtained by the sample age 5-6-11.

In sum, however, the sample presented with a wide range of demographic characteristics and ability levels, and likely not only approximates a reasonably typical clinical population routinely seen by educational psychologists, but also contains a number of normal subjects. Thus, the sample appears ideally suited to the multivariate research purposes and objectives of this research, and in the next part of the study, the relationship between the SB:FE full battery and the GPAB in this sample will be examined.

CHAPTER FIVE

Part Two

Comparison Between SB:FE Full Battery and GPAB Reasoning Area and Composite SAS's Means, Variances, and Correlation Coefficients

Presented in this chapter are the results of the examination of the relationship between the Reasoning Area and Composite SAS's obtained from the SB:FE Full Battery and the SB:FE GPAB respectively. In addition, correlations between these two versions of the SB:FE and the WRAT-R are examined.

Reported in Table 16 are the means, standard deviations, and correlations for each of the four Reasoning Area and Composite SAS scores for the SB:FE Full Battery and the SB:FE GPAB for each of the age groups. Significant differences between pairs of means and pairs of variances identified from the statistical analyses conducted to test the statistical hypotheses stated at the beginning of Chapter Three are indicated next to the statistics for the SB:FE GPAB. Statistically significant correlations are indicated next to the coefficient.

Hypothesis 1:1

It was hypothesized that there would be no significant differences between Reasoning Area and Composite SAS means based on the full battery of the SB:FE and Reasoning Area and Composite SAS means derived from the SB:FE GPAB at all three age levels. This hypothesis generally did not receive support; employing the *t*-test for dependent groups, 11 of the 15 comparisons were significant. This number exceeds the number expected only by chance.

Table 16

Means, Standard Deviations, and Correlations of SB:FE Full Battery and GPAB Reasoning and Composite SAS scores for Each Age Group

SB:FE Version		Ver. SAS	A/V SAS	Qnt. SAS	STM SAS	Com. SAS
Age 2-6-11						
(n=394)						
Full Battery	M	111.5	106.8	108.4	105.6	109.6
	SD	13.5	17.1	15.7	16.2	14.9
GPAB	M	111.0*	108.0*	108.3	104.1*	109.4
	SD	14.6**	18.5**	15.6	17.1**	15.3**
Correlations		.95***	.88***	.99***	.97***	.97***
Ages 7-11-11						
(n=527)						
Full Battery	M	102.8	99.4	99.4	98.9	100.2
	SD	16.2	16.7	14.7	16.9	16.4
GPAB	M	102.6	102.3*	97.3*	96.6*	99.8*
	SD	16.5	17.5**	14.7	18.0**	16.2
Correlations		.95***	.86***	.96***	.94***	.97***
Age 12-23-11						
(n=299)						
Full Battery	M	98.3	96.4	95.5	93.5	95.4
	SD	18.7	19.2	21.3	19.3	20.4
GPAB	M	98.2	98.8*	93.1*	92.2*	94.9*
	SD	19.5**	17.7**	20.9	20.4**	19.4**
Correlations		.98***	.88***	.95***	.95***	98***

* denotes pairs of means significantly different at $p < .05$; ** denotes pairs of variances significantly different at $p < .05$; *** denotes correlations significant at $p < .05$.

Significant differences were found between mean SB:FE full battery and GPAB Composite SAS's in the age sample 7-11-11 [t (526) = 2.6, $p < .05$], and in the age sample 12-23-11 [t (298) = 2.04, $p < .05$]. (See Table 16). No significant difference was noted in the age sample 2-6-11.

Within the age sample 2-6-11, significant differences were noted between the SB:FE and SB:FE GPAB areas scores in the Verbal Reasoning Area [t (393) = 2.00, $p < .05$]; the Abstract/Visual Reasoning Area [t (393) = 2.70, $p < .05$]; and in the Short Term Memory Area [t

(393) = 6.84, $p < .05$]. No significant difference was noted between SB:FE full battery and GPAB Quantitative Reasoning Area SAS's.

Within the age sample 7-11-11, significant differences were noted between the SB:FE and SB:FE GPAB areas scores in the Abstract/Visual Reasoning Area [t (526) = 7.29, $p < .05$]; the Quantitative Reasoning Area [t (526) = 8.05, $p < .05$]; and the Short Term Memory Area [t (526) = 6.59, $p < .05$]. No significant difference was noted between SB:FE full battery and GPAB Verbal Reasoning Area SAS's.

Finally, within the age sample 12-23-11, significant differences were noted between the SB:FE and SB:FE GPAB areas scores in the Abstract/Visual Reasoning Area [t (298) = -4.36, $p < .05$]; the Quantitative Reasoning Area [t (298) = 6.59, $p < .05$]; and the Short Term Memory Area [t (298) = 3.56, $p < .05$]. No significant difference was noted between SB:FE full battery Verbal SAS's and GPAB SAS's.

Hypothesis 1:2

It was hypothesized that there would be no significant differences between Reasoning Area and Composite SAS variances based on the SB:FE full battery and the respective Reasoning Area and Composite SAS variances obtained from the SB:FE GPAB at all three age levels. The t-test for homogeneity of variances for paired observations was used to test these hypotheses (Glass & Hopkins, 1984, p. 269). This hypothesis generally did not receive support. As shown in Table 16, 10 of the 15 tests revealed significant differences, which is far in excess of what would be expected by chance alone.

There was a significant difference between the variances of the Composite SAS's obtained from each version of the test in the age samples 2-6-11 [t (2, 393) = 2.21, $p < .05$] and 12-23-11 [t (2, 298) = 3.66, $p < .05$]. However, in the age sample 7-11-11, the difference between the two Composite score variances was not significant [t (2, 526) = .93, nsd].

Within the age sample 2-6-11, significant differences between variances were noted in the Verbal Reasoning Area [t (2, 393) = 5.11, $p < .05$]; the Abstract/Visual Reasoning Area [t (2, 393)

= 3.27, $p < .05$]; and the Short Term Memory Area [t (2, 393) = 4.47, $p < .05$]. No significant difference was noted between the Quantitative Reasoning Area variances.

Within the age sample 7-11-11, significant differences between variances were noted in the Abstract/Visual Reasoning Area [t (2,526) = 2.09, $p < .05$] and the Short Term Memory Area [t (2, 526) = 4.34, $p < .05$]. No significant difference was noted between the variances in the Verbal and Quantitative Reasoning Areas.

Lastly, within the age sample 12-23-11, significant differences between variances were noted in the Verbal Reasoning Area [t (2,298) = 3.43, $p < .05$]; the Abstract/Visual Area [t (2,298) = 2.75, $p < .05$]; and the Short Term Memory Area [t (2, 298) = 2.89, $p < .05$]. No significant difference was noted between the Quantitative Reasoning Area variances.

<u>Hypothesis 1:3</u>

It was hypothesized that there would be significant correlations, approaching unity, between the Reasoning Area and Composite SAS's based on the SB:FE full battery and the Reasoning Area and Composite SAS's derived from the SB:FE GPAB at all three age levels. As pointed out in Chapter Three, it was not possible to use correlations corrected for attenuation. Consequently, the results of the correlation coefficients provide lower bound estimates. The lack of internal consistency estimates also prevented testing statistically whether the correlation coefficients obtained differed significantly from one, which would have been the strongest test of this particular hypothesis. However, inspection of the correlation coefficients presented in Table 16, generally provided support for the hypothesis. All 15 correlations conducted were statistically significant ($p < .01$).

Correlations between the Composite SAS's under consideration ranged from .97 to .98 across the three age groups.

Within the age group 2-6-11, correlations between the SB:FE full battery and SB:FE GPAB Verbal, Quantitative, and Short Term Memory Reasoning Area SAS's ranged from .9[illegible] to .99. The correlation between SB:FE full battery and SB:FE GPAB Abstract/Visual Reasoning Area SAS's was lower at .88, but still significant.

Within the age group 7-11-11, correlations between the SB:FE and SB:FE GPAB Verbal, Quantitative, and Short Term Memory Reasoning Area SAS's ranged from .95 to .96. The correlation between SB:FE full battery and GPAB Abstract/Visual Reasoning Area SAS's was lower at 86, but still significant.

Within the age group 12-23-11, correlations between the SB:FE and SB:FE GPAB Verbal, Quantitative, and Short Term Memory Reasoning Area SAS's ranged from .95 to .98. The correlation between SB:FE full battery GPAB Abstract/Visual Reasoning Area SAS's was lower at .88, but still significant.

Comparison Among SB:FE Full Battery and GPAB Reasoning Area and Composite SAS's and External Criteria

As a final means of investigating the relationship between the SB:FE and the SB:FE GPAB, the respective correlations among these two versions of the tests and an external criterion, the WRAT-R, were examined. These results appear in Table 17. Glutting (1989) criticized Thorndike et al. (1986b) for being remiss in not providing evidence of correlations between the abbreviated batteries and valid external criteria. As can be seen from Table 17, the correlations between the SB:FE full battery and the WRAT-R and the SB:FE GPAB and the WRAT-R are generally similar, with the slight differences noted, likely due to attenuation due to unreliability. The largest discrepancies between the correlations obtained by the two forms with the WRAT-R involved the Abstract/Visual Reasoning Area. As noted previously, the correlations between the SB:FE full battery and GPAB were also lowest in the Abstract/Visual Reasoning Area (cf., Table 16). Overall, however, the correlations of both versions of the test with the external criteria provides further support for the validity of the SB:FE GPAB as a substitute for the full battery.

Table 17

Correlations Among SB:FE Full Battery, GPAB, and WRAT-R for Each Age Group

SB:FE Version	Ver. SAS	A/V SAS	Qnt. SAS	STM SAS	Com. SAS
Age 5-6-11					
(*n*=120)					
SB:FE Full Battery					
Reading	.49	.43	.43	.59	.60
Spelling	.57	.52	.48	.62	.67
Arithmetic	.54	.50	.47	.64	.67
SB:FE GPAB					
Reading	.44	.39	.40	.56	.57
Spelling	.52	.40	.46	.59	.62
Arithmetic	.49	.40	.46	.63	.63
Age 7-11-11					
(*n*=301)					
SB:FE Full Battery					
Reading	.60	.53	.58	.58	.66
Spelling	.56	.51	.59	.56	.64
Arithmetic	.62	.63	.67	.64	.73
SB:FE GPAB					
Reading	.60	.46	.49	.54	.64
Spelling	.57	.43	.50	.53	.62
Arithmetic	.59	.55	.58	.60	.71
Age 12-23-11					
(*n*=186)					
SB:FE Full Battery					
Reading	.75	.67	.69	.73	.78
Spelling	.65	.62	.67	.70	.73
Arithmetic	.73	.71	.82	.72	.82
SB:FE GPAB					
Reading	.74	.51	.65	.69	.75
Spelling	.65	.46	.61	.63	.69
Arithmetic	.71	.59	.77	.69	.80

Summary of the Results of Chapter Five

In this chapter, the relationship between the SB:FE full battery and the SB:FE GPAB was explored for each age group, 2-6-11, 7-11-11, and 12-23-11. First, differences in means and variances between the two versions of the test were compared. In general, statistically significant differences between full battery and abbreviated battery means and variances in Reasoning Area and Composite SAS's were observed. However, the differences observed may largely be attributable to the sample sizes. Glass and Hopkins (1984) noted that "when n is very large, even a trivial difference may be large enough to be highly statistically significant" (p. 214). Glass and Hopkins (1984) distinguished between practical and statistically significant results. This distinction is relevant to this dissertation, as the actual differences in absolute terms between the scores obtained from the SB:FE full battery and the SB:FE GPAB were minimal. For example, a difference of half an SAS point in Reasoning Area and Composite SAS's with means typically around 100 were large enough to produce statistically significant results (cf., Table 16). Next, the correlations between the two versions of the test were examined. These correlation coefficients were not corrected for attenuation. However, the coefficients obtained were all statistically significant and generally close to unity, particularly among the Composite SAS's and three of the four Reasoning Area SAS's: the Verbal, Quantitative, and Short Term Memory Areas respectively. Correlations were consistently lower between the Abstract/Visual Reasoning Area scores obtained from the two versions. Finally, correlations between the SB:FE full battery and SB:FE GPAB and an external criteria were examined. Similar correlations coefficients, albeit slightly lower, were observed among the two versions of the SB:FE and the WRAT-R. The slightly lower correlations coefficients that were obtained between the SB:FE GPAB and WRAT-R likely reflect attenuation due to unreliability; however lack of available internal consistency estimates precluded calculation of disattenuated correlation coefficients for any of the variables considered.

Thus, from a practical point of view, the results, when taken together, suggest that when working with Reasoning Area and Composite SAS's, the SB:FE full battery and the SB:FE GPAB may, for certain purposes be interchangeable, and that it may be possible to use the abbreviated version without significant loss of information or practical differences in scores. Use of the SB:FE GPAB is advantageous in that it provides a uniform battery that is administered across all ages, 2-23-11. Consequently, in light of the documented relationship between the full and abbreviated versions and in light of the benefits accruing from use of the GPAB, this version of the SB:FE was used in subsequent analyses conducted in this study.

CHAPTER SIX

Part Three

Presented in this chapter are the results of the application of the multivariate cluster analysis of the SB:FE GPAB data used to identify reliable and replicable (internally valid) groups of individuals with distinct cognitive profiles with the age groups considered. As noted, the adaptive testing and age scale formats of the SB:FE pose particular problems from a clinical and research perspective because different subjects may be given different subtests. Because of the difficulties inherent in the use of the SB:FE full battery and due to the multivariate nature of the study the Area SAS's derived from the SB:FE GPAB were used in this part. The relationship between the two versions of this test has been documented in Chapter Five. As pointed out in Chapter Three, the initial analyses were performed on the subsample of subjects at each age level for whom both SB:FE GPAB data and WRAT-R date were available. The subsamples for whom SB:FE GPAB data only were available were then used to validate the initial cluster solutions at each age group.

The following hypothesis was tested in this section:

Hypothesis 2. Application of hierarchical agglomerative (HAP) and iterative partitioning (IPP) cluster analytic procedures to the SB:FE GPAB data will result in the identification of subgroups within the sample age ranges 2-4-11, 5-6-11, 7-11-11, and 12-23-11 years that display significant differences with respect to profile elevation and/or shape.

As pointed out, the WRAT-R is administered to subjects age 5-00 and older. No subjects in the age group 2-4-11 were therefore administered the WRAT-R. However, SB:FE GPAB scores were available in the for 110 subjects within this age group. Clustering procedures were therefore applied to this subsample, and results are presented after the results of the cluster analysis with the age samples 5-6-11, 7-11-11, and 12-23-11 are presented.

Cluster Analytic Procedures Applied to the Subsample Age 5-6-11

The subsample of subjects for whom both SB:FE GPAB and WRAT-R data were available in the age group 5-6-11 comprised 120 subjects.

Determination of the Number of Clusters

Presented in Table 18 are the values of the fusion coefficients corresponding to the last 10 (out of 120) steps of the hierarchical agglomerative algorithm (HAP) used to identify the clusters or "subgroups" within this sample which display differences between profile elevation and/or shape. Inspection of the results reveals a sudden increase in fusion values when going from 3 to 2 clusters, followed by a further extreme increase as the last two clusters are forced together. Thus, it would appear that there are 3 subgroups within the 5-6-11 age group.

Table 18

HAP Fusion Coefficients: SB:FE GPAB and WRAT-R Data Subsample 5-6-11

Number of Clusters	Clustering Coefficient	Percentage Change From Previous Step
10	36812.79	6.08
9	39256.75	6.64
8	41862.46	6.64
7	44906.82	7.27
6	48231.08	7.40
5	54700.01	13.41
4	64080.66	17.15
3	75253.00	17.43
2	95988.00	27.55
1	138047.43	43.82

To further confirm that the three cluster solution optimally represented this data set, a *k*-means iterative procedure (IPP) was performed for 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 11- and 12- clusters. The results of this analysis, along with the number of subjects and means for each group yielded by this procedure are reported in Table 19. For parsimony, only the results of the analysis 2 through 7 solutions are reported. Examination of the triangular matrices of sample mean D^2 distances within clusters across clustering solutions presented in Table 19, generally suggests that the 3 cluster solution is optimal. A two cluster solution is acceptable, however, the three cluster solution maintains the homogeneity within groups, and also maintains the separation between groups. The four cluster solution does not appear to increase cluster homogeneity, and has the disadvantage that there appears to be relatively poor separation between the first and fourth groups. The NNR for the two cluster solution was .54 and for the three cluster solution was .59 which is also acceptable. The NNR for the four cluster solution increased to .87 which is unacceptable. The NNR's for the solutions, 5, 6, and 7 are .76, .81, and .81 respectively. Not shown in Table 19 are the NNR's for the cluster solutions 8 through to 12. These were all above .9, and are thus unacceptable. As can be seen from the means reported in columns 3-6, Table 19, and the corresponding graph of this solution displayed in Figure 2, the three cluster solution groups the subjects into a below average group, an average group, and an above average group.

Table 19

IPP Data for 2-7 Cluster Solutions: SB:FE GPAB and WRAT-R Data Subsample Age 5-6-11

		Reasoning Area Means				Mean D^2 Between and Within Clusters								
k	n	Ver. SAS	A/V SAS	Qnt. SAS	STM SAS	1	2	3	4	5	6	7	Range of Distances From Centre	NNR
1	91	115.8	116.3	114.9	109.7	25.1							5.0-59.3	
2	29	96.1	89.4	93.3	85.0	46.7	22.3						6.2-44.7	.54
1	12	130.9	139.5	121.3	133.9	24.9							10.7-44.0	
2	80	113.2	113.0	113.7	105.7	43.2	21.9						4.1-43.3	
3	28	96.5	88.0	93.1	85.0	83.8	41.8	21.5					4.9-44.7	.59
1	46	118.6	103.4	113.8	108.9	19.5							8.9-44.6	
2	8	135.0	141.3	126.3	141.3	53.8	22.3						9.6-36.4	
3	27	95.9	88.1	92.9	84.4	42.2	93.2	21.5					5.4-44.4	
4	39	107.8	125.9	113.1	103.3	25.5	50.9	48.3	19.7				4.7-33.3	.87
1	44	120.3	104.6	113.1	108.5	19.0							7.9-42.9	
2	5	141.2	141.6	131.6	150.4	62.4	19.3						14.5-29.1	
3	22	98.6	110.6	112.0	94.0	46.9	107.4	20.2					5.7-46.4	
4	22	95.0	85.6	88.9	83.8	26.7	79.6	35.7	17.6				7.6-31.8	
5	27	113.8	131.3	115.0	111.4	26.6	51.5	62.3	31.1	18.6			5.9-31.4	.76
1	33	122.5	114.7	116.2	106.9	15.9							6.7-37.1	
2	5	141.2	141.6	131.6	150.4	56.5	19.3						14.5-29.1	
3	24	112.3	92.3	110.5	110.0	25.4	73.0	18.5					10.4-30.8	
4	20	94.4	86.0	88.2	81.3	55.3	109.3	41.0	19.3				5.2-45.8	
5	19	111.4	134.9	112.9	112.5	28.0	79.0	31.4	38.7	15.9			6.9-30.0	
6	19	98.9	114.9	110.8	92.7	24.0	52.1	42.7	65.3	30.7	18.8		5.2-30.3	.81
1	33	122.5	114.7	116.2	106.9	15.9							6.7-37.1	
2	5	141.2	141.6	131.6	150.4	56.5	19.3						14.5-29.1	
3	22	113.5	93.1	110.5	111.2	24.4	71.4	18.2					9.6-31.1	
4	7	90.9	80.0	107.1	84.1	52.9	106.4	37.7	19.7				12.3-31.3	
5	19	111.4	134.9	112.9	112.5	28.0	79.0	32.0	37.0	15.9			6.9-30.5	
6	19	98.9	114.9	110.8	92.7	24.0	52.1	41.9	65.3	30.7	18.8		5.2-30.3	
7	15	96.7	88.5	82.3	82.0	55.8	109.1	44.2	27.0	40.4	65.1	15.8	7.5-33.7	.81

Note: Within each triangular cluster matrix mean distances within clusters are read along the diagonal. Distances between clusters are read at the intersection of the vertical and horizontal of respective clusters.

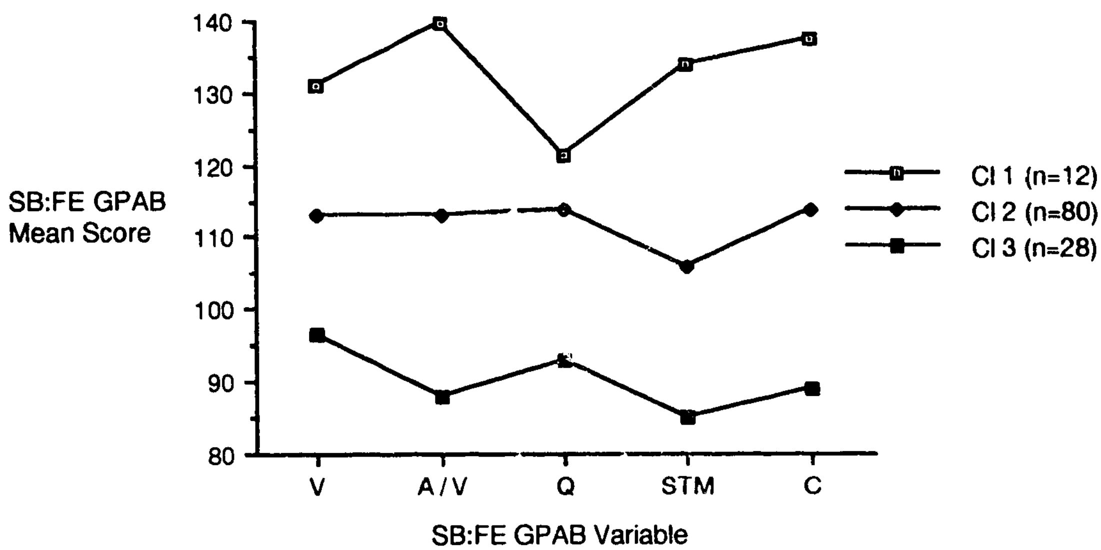

Figure 2. Graph displaying cluster scores on SB:FE GPAB Reasoning Area and Composite SAS's: SB:FE GPAB and WRAT-R data subsample age 5-6-11.

Agreement Between Hierarchical Agglomerative and Iterative Procedures Assignment of Subjects to Clusters

Comparison of the hierarchical agglomerative and iterative clustering procedures, together with the rules for determining the number of clusters, generally pointed to a three cluster solution as being the most optimal representation of this data set. Therefore the level of agreement between the assignment of individuals to cluster groups by the hierarchical agglomerative solution and the assignment of individuals to the cluster groups by the iterative partitioning procedure was examined next. The classification matrix for this comparison appears in Table 20. Groups were matched according to SB:FE GPAB Area scores; in other words the high average group from the first procedure was matched with the high average group from the second procedure and so on. From Table 20, the percentage agreement between the two sets identified, 88%, was significant [X^2 (4, 120) = 139.9: $p < .001$]. This finding tends to support the internal validity of the obtained solutions.

Table 20

HAP vs. IPP Assignment of Subjects: SB:FE GPAB and WRAT-R Data Subsample Age 5-6-11

		Hierarchical Agglomerative Procedures			
		Group 1	Group 2	Group 3	n
	Group 1	7	0	0	7
Iterative	Group 2	5	80	9	94
Procedures	Group 3	0	0	19	19
	n	12	80	28	120

Multivariate Profile Analysis

The groups derived through the iterative partitioning procedure were subjected to a multivariate profile analysis. There was a significant multivariate main effect for profile elevation [$F(8, 228) = 34.77$, $p < .001$] when the three profiles were tested simultaneously. This test is analogous to simultaneous MANOVA exploring differences in group means of the four SB:FE GPAB Reasoning Area SAS's. This statistic was significant and therefore univariate F tests for each SB:FE GPAB Reasoning Area SAS were conducted next. The results are presented in Table 21. This analysis cannot be considered a validation procedure, as by definition the clusters were formed so as to maximize between groups difference and minimize within group differences, but it enables examination of the differences between the groups on the SB:FE variables used in the clustering procedure. The results indicate that each of the group means for the SB:FE GPAB variables in the three cluster solution are significantly different, emphasizing the clear separation between the groups. There was also a significant multivariate main effect for the three group solution when the profile line segments were tested for parallelism [$F(6, 230) = 5.12$, $p < .001$]. This statistic examines whether the group-mean profiles are similar, in the sense that the line segments of adjacent variables are parallel. Univariate statistical comparisons for each pair of group differences indicated that groups 1 and 3 differed with respect to all three line segments, Groups 1 and 2 differed with respect to the Abstract/Visual - Quantitative and Quantitative - Short Term Memory SAS line segments. Groups 2 and 3 did not differ with respect to any of the three line segments. Results are presented in Table 22.

Table 21

Differences Between SB:FE Reasoning Area Group Means for IPP Cluster Solution: SB:FE GPAB and WRAT-R Data Subsample Age 5-6-11

	Means of Iterative Clusters			Univariate F tests (2, 298)			
SB:FE Variable	1 (n=12)	2 (n=80)	3 (n=28)	F ratio	1 vs 2	1 vs 3	2 vs 3
Verbal SAS	130.9	113.2	96.5	36.4	*	*	*
A/V SAS	139.5	112.9	88.0	68.5	*	*	*
Quantitative SAS	121.3	113.7	93.1	41.1	*	*	*
STM SAS	133.9	105.7	85.0	73.5	*	*	*
Multivariate Main Effect for Elevation: $F(8, 228) = 34.77, p < .001$							

*denotes pairs of means significantly different at the $p < .01$ level.

Table 22

Tests for Profile Parallelism on IPP Clusters: SB:FE GPAB and WRAT-R Data Subsample Age 5-6-11

	Iterative Clusters		
SB:FE Segment	1 vs 2	1 vs 3	2 vs 3
Verbal - A/V SAS	NS	*	NS
A/V - Quantitative SAS	*	*	NS
Quantitative - STM SAS	*	*	NS

Multivariate Main effect for Parallelism:
$F(6, 230)$ 5.122, $p < .001$

*denotes pairs of line segments significantly different at the $p < .05$ level; NS refers to non significant differences between pairs of line segments.

Cross-validation of Cluster Solutions for the Subsample Age 5-6-11

As a final means of testing the validity of the solutions derived in the previous stage, identical clustering techniques (hierarchical agglomerative and iterative partitioning procedures) were applied to a different subsample of subjects with SB:FE GPAB data of the same age (n=164). This sample comprised the subjects included in Part One of the study for whom WRAT-R data were not available.

Determination of the Number of Clusters

Presented in Table 23 are the values of the fusion coefficients corresponding to the last 10 (out of 164) steps of the hierarchical agglomerative algorithm used to identify the clusters within this sample which display differences between profile elevation and/or shape. Inspection of the results indicates that between the 3 and 2 cluster solutions, there was an increase in the percentage of change in the fusion coefficient, followed by a further extreme increase as the last two clusters are forced together. This pattern strongly points again to the presence of a three cluster solution as being the optimal solution.

Table 23

HAP Fusion Coefficients: SB:FE Data Only Subsample Age 5-6-11

Number of Clusters	Clustering Coefficient	Percentage Change From Previous Step
10	47527.47	5.97
9	51341.67	8.03
8	55638.65	8.37
7	61559.13	10.64
6	67505.81	9.66
5	74871.18	10.91
4	83662.75	11.74
3	93010.63	11.17
2	116488.38	25.24
1	187947.87	61.34

To further confirm that the three cluster solutions optimally represented this data set, a k-means iterative procedure was performed for 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 11- and 12-clusters. The results of this analysis, along with the number of subjects and means for each group within the cluster solution are reported in Table 24. For parsimony, only the results of the analysis 2 through 7 solutions are reported. Examination of the triangular matrices of sample mean Squared Euclidean distances within and between clusters, across the clustering solutions, supports a three cluster interpretation (see Table 24). The three cluster solution maintains homogeneity within groups, as well as the separation between groups. The NNR for the two cluster solution was .53, which is lower than the three cluster solution. However, the NNR for the three cluster solution was .59 which is within acceptable limits. The four cluster solution does appear to have increased cluster homogeneity, but the NNR for the four cluster solution increased to .69, which is unacceptable. The NNR's for the solutions, 5, 6, and 7 are .83, .77, and .91 respectively. Not shown in Table 24 are the NNR for the cluster solutions 8 through to 12. These were all above .90, and are thus unacceptable. As can be seen from the means presented in the graph of this solution displayed in Figure 3, and in columns 3-6, Table 24, the three cluster solution groups the subjects into a below average group, an average group, and an above average group.

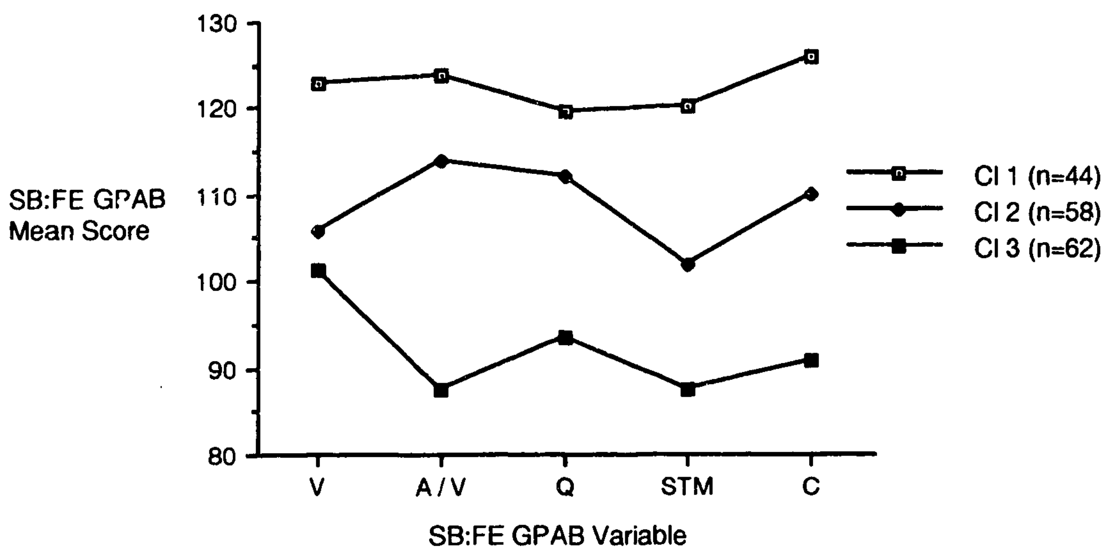

Figure 3. Graph displaying cluster scores on SB:FE GPAB Reasoning Area and Composite SAS's: SB:FE only subsample age 5-6-11.

Table 24

<u>IPP Data for 2-7 Cluster Solutions: SB:FE GPAB Only Subsample Age 5-6-11</u>

		Reasoning Area Means				Mean D^2 Between and Within Clusters								
k	n	Ver. SAS	A/V SAS	Qnt. SAS	STM SAS	1	2	3	4	5	6	7	Range of Distances From Centre	NNR
1	96	114.0	118.8	116.2	110.7	23.5							6.8-56.2	
2	68	101.2	89.5	94.0	88.0	44.8	23.6						6.2-55.8	.53
1	44	122.9	123.7	119.6	120.1	20.1							2.7-43.6	
2	58	105.8	113.8	111.9	101.8	38.7	19.9						7.2-43.6	
3	62	101.3	87.4	93.5	87.4	59.4	35.5	22.9					4.3-55.1	.59
1	40	111.2	91.9	109.1	100.5	19.7							9.7-41.2	
2	39	124.1	120.9	120.8	121.3	39.7	19.2						3.3-30.6	
3	41	96.95	88.7	87.7	82.8	31.3	65.9	21.7					4.7-67.9	
4	44	103.8	123.6	110.9	101.5	32.6	30.1	46.4	18.2				6.9-33.6	.69
1	33	115.9	97.2	113.7	112.9	18.6							11.6-37.4	
2	10	127.1	130.2	118.6	136.6	42.3	24.4						8.1-28.5	
3	49	99.9	86.1	91.2	84.4	41.2	78.5	18.6					4.1-70.4	
4	37	100.4	115.5	107.4	96.4	29.7	51.6	35.6	18.2				8.1-36.9	
5	35	117.7	127.5	119.0	109.1	31.0	29.2	58.4	27.1	15.7			8.9-29.5	.83
1	25	120.6	106.7	123.5	118.5	18.7							11.3-32.8	
2	9	126.4	131.1	116.4	137.8	32.4	14.6						9.8-28.3	
3	29	110.4	85.1	103.7	95.4	38.6	65.8	17.6					4.9-41.2	
4	32	95.4	87.6	84.8	82.0	61.8	83.4	27.7	21.4				6.4-64.8	
5	40	101.5	114.4	107.3	97.7	33.4	50.9	30.8	38.7	17.5			6.8-36.6	
6	29	115.9	130.1	117.2	107.3	27.0	32.2	48.7	62.5	25.4	15.79		7.2-27.9	.77
1	23	122.0	112.9	122.8	119.0	16.2							7.1-24.9	
2	9	126.4	131.1	116.4	137.8	27.2	14.6						5.6-28.3	
3	21	112.7	91.4	109.2	109.8	28.6	50.9	18.9					8.1-34.7	
4	26	105.1	85.6	103.5	86.7	49.4	72.7	25.5	14.4				6.7-19.8	
5	35	100.1	115.8	107.5	96.9	34.8	51.8	30.3	32.4	17.1			7.8-37.0	
6	25	115.6	131.3	116.8	106.0	24.1	33.5	40.8	52.4	25.4	15.5		7.3-27.8	
7	25	95.6	88.2	80.5	81.4	67.1	85.2	43.9	25.5	41.7	64.5	22.6	5.1-63.4	.91

<u>Note</u>: Within each triangular cluster matrix mean distances within clusters are read along the diagonal. Distances between clusters are read at the intersection of the vertical and horizontal of respective clusters.

Agreement Between Hierarchical Agglomerative and Iterative Procedures Assignment of Subjects to Clusters

The level of agreement between the assignment of individuals to cluster groups by the hierarchical agglomerative solution and the assignment of individuals to cluster groups by the iterative partitioning solution was examined next. The classification matrix for this comparison appears in Table 25. From Table 25, the percentage agreement between the two sets identified, 87%, was significant [X^2 (4, 164) = 223.4: $p < .001$]. This finding tends to support the internal validity of the obtained solutions.

Table 25

HAP vs. IPP Assignment of Subjects: SB:FE GPAB Data Only Subsample Age 5-6-11

		Hierarchical Agglomerative Procedures			
		Group 1	Group 2	Group 3	n
	Group 1	43	12	0	55
Iterative	Group 2	1	46	8	55
Procedures	Group 3	0	0	54	54
	n	44	58	62	164

Multivariate Profile Analysis

The groups derived through the iterative partitioning solution were then subjected to multivariate profile analysis. There was a significant multivariate main effect for elevation [$F(8, 316) = 59.10$, $p < .001$]. The results, along with the univariate comparisons, presented in Table 26, indicate that each of the group means for the SB:FE GPAB variables in the three cluster solution are significantly different from each other. A graph showing these scores is presented in Figure 3, emphasizing the clear separation between the groups. There was also a significant multivariate main effect for the three group solution when the profile line segments were tested for parallelism [$F(6, 318) = 2.66$, $p < .05$]. Univariate comparisons indicated that Groups 1 and 2 differed with respect to the profile line segment defined by the Quantitative - Short Term Memory line segment. Groups 1 and 3 differed with respect to the shape of their Verbal - Abstract/Visual and Abstract/Visual - Quantitative Reasoning Area line segments. Groups 2 and 3 were different to each other in the shape (segment) of the profile defined by their Verbal - Abstract/Visual Reasoning Area line segment. The results are presented in Table 27.

Table 26

Differences Between SB:FE Reasoning Area Group Means for IPP Cluster Solution: SB:FE GPAB Data Only Subsample Age 5-6-11

	Means of Iterative Clusters			Univariate F tests (2, 223)			
SB:FE Variable	1 (n=44)	2 (n=58)	3 (n=62)	F ratio	1 vs 2	1 vs 3	2 vs 3
Verbal SAS	116.8	100.4	76.8	137.7	*	*	*
A/V SAS	118.7	98.8	72.1	148.3	*	*	*
Quantitative SAS	111.8	93.5	76.2	97.2	*	*	*
STM SAS	113.5	94.4	65.8	176.8	*	*	*

Multivariate Main Effect for Elevation:
$F(8, 316) = 59.10, p < .001$

*denotes pairs of means significantly different at the $p < .05$ level.

Table 27

Tests for Profile Parallelism on IPP Clusters: SB:FE GPAB Data Only Subsample Age 5-6-11

	Iterative Clusters		
SB:FE Segments	1 vs 2	1 vs 3	2 vs 3
Verbal - A/V SAS	NS	*	*
A/V - Quantitative SAS	NS	*	NS
Quantitative - STM SAS	*	NS	NS

Multivariate Main effect for Parallelism:
$F(6, 318) = 10.79$, $p < .001$

*denotes pairs of line segments significantly different at the $p < .05$ level; NS refers to non significant differences between pairs of line segments.

Comparison between the solutions derived from the application of hierarchical agglomerative and iterative partitioning procedures to the two subsamples. Overall the results suggest some similarity between the cluster solutions applied to the two samples. The second sample has slightly lower means across all Reasoning Area and Composite SAS's (see Table 28). However, a three cluster solution optimally represented the data set in both samples.

Table 28

SB:FE GPAB Reasoning Area and Composite SAS Cluster Means for the Two Samples Age 5-6-11

Subsample	n	Ver. SAS	A/V SAS	Qnt. SAS	STM SAS	Com. SAS
SB:FE with WRAT-R data						
Cluster 1	12	130.9	139.5	121.3	133.9	137.5
Cluster 2	80	113.2	112.9	113.7	105.7	113.7
Cluster 3	28	96.5	88.0	93.1	85.0	88.9
Total	120	111.1	109.8	109.7	103.7	110.3
SB:FE Data only						
Cluster 1	44	122.9	123.7	119.6	120.1	125.9
Cluster 2	58	105.8	113.8	111.9	101.8	110.0
Cluster 3	62	101.5	87.4	93.5	87.4	90.9
Total	164	108.7	106.5	107.0	101.3	107.1

Cluster Analytic Procedures Applied to the Subsample Age 7-11-11

The subsample of subjects for whom both SB:FE GPAB and WRAT-R data were available in the age group 7-11-11 comprised 301 subjects.

Determination of the Number of Clusters

Presented in Table 29 are the values of the fusion coefficients corresponding to the last 10 (out of 301) steps of the hierarchical agglomerative algorithm used to identify the clusters or within this sample which display differences between profile elevation and/or shape. Inspection of the results reveals a sudden increase in fusion values when going from 3 to 2 clusters, followed by a further extreme increase as the last two clusters are forced together. Thus, it would appear that there are 3 subgroups within the 7-11-11 age group.

Table 29

HAP Fusion Coefficients: SB:FE GPAB and WRAT-R Data Subsample Age 7-11-11

Number of Clusters	Clustering Coefficient	Percentage Change From Previous Step
10	86666.1	5.19
9	92894.5	7.19
8	99766.7	7.40
7	108868.9	9.12
6	118115.3	8.49
5	129023.0	9.23
4	145542.7	12.80
3	165890.2	13.98
2	202371.5	21.99
1	329086.2	62.61

To further confirm that the three cluster solutions optimally represented this data set, a k-means iterative procedure was performed for 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 11- and 12-clusters. The results of this analysis, along with the number of subjects and means for each group yielded by this procedure are reported in Table 30. For parsimony, only the results of the analysis 2 through 7 solutions are reported. Examination of the triangular matrices, presented in Table 30, of sample mean D^2 distances within and between clusters across clustering solutions generally suggests that the 3 cluster solution is optimal, maintaining the homogeneity within groups, whilst maintaining the separation between groups. The four cluster solution does not appear to increase cluster homogeneity, and has the disadvantage that there appears to be relatively poor separation between the first and second groups. The NNR for the two cluster solution was .56, which is lower than the three cluster solution. The NNR for the three cluster solution was .66 which, although approaching the upper limits, is acceptable. The NNR for the four cluster solution increased to .90 which is unacceptable. The NNR solutions, 5, 6, and 7 are .88, .94, and .90 respectively. Not shown in Table 30 are the NNRs for the cluster solutions 8 through to 12. These were all above .9, and are thus unacceptable. As can be seen from the means reported in columns 3-6, Table 30, and the corresponding graph of this solution displayed in Figure 4, the three cluster solution groups the subjects into a below average group, an average group, and an above average group.

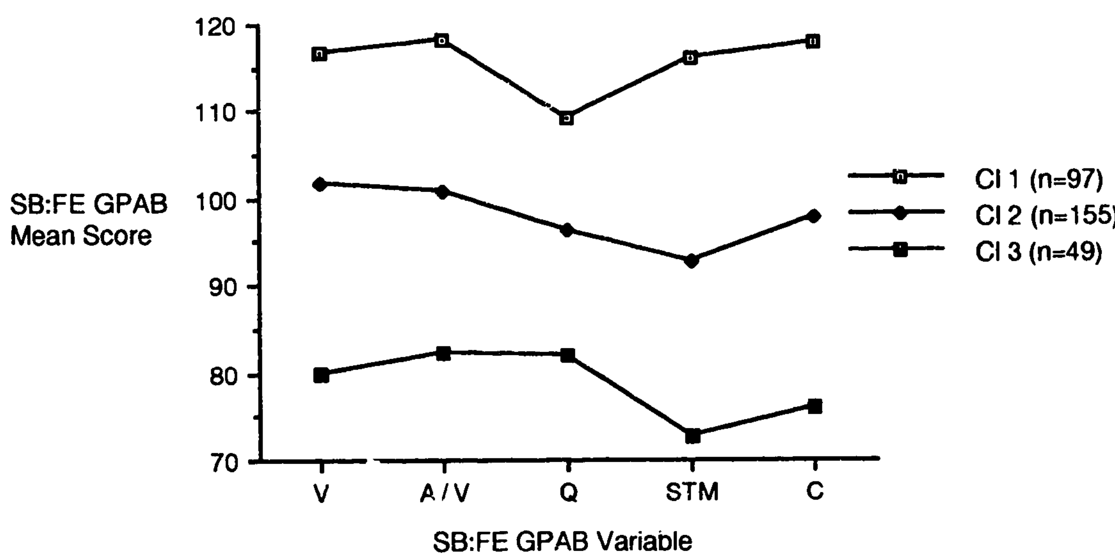

Figure 4. Graph displaying cluster scores on SB:FE GPAB Reasoning Area and Composite SAS's: SB:FE GPAB and WRAT-R data subsample age 7-11-11.

Table 30

IPP Data for 2-7 Cluster Solutions: SB:FE GPAB and WRAT-R Data Subsample Age 7-11-11

		Reasoning Area Means				Mean D^2 Between and Within Clusters								
k	n	Ver. SAS	A/V SAS	Qnt. SAS	STM SAS	1	2	3	4	5	6	7	Range of Distances From Centre	NNR
1	166	93.5	93.9	91.0	85.4	22.7							6.3-45.7	
2	135	114.5	115.1	107.2	111.3	42.5	23.9						5.0-62.2	.56
1	155	101.7	100.9	96.4	92.8	18.9							4.5-38.7	
2	97	116.7	118.3	109.3	116.0	35.0	23.1						6.8-47.5	
3	49	79.9	82.2	82.0	72.8	37.7	72.5	20.4					6.7-50.8	.66
1	120	98.7	96.7	94.1	89.8	17.7							4.7-39.3	
2	91	107.8	113.1	101.5	103.9	24.3	19.2						6.4-38.3	
3	52	124.5	118.7	115.0	121.7	50.9	28.8	21.9					5.7-47.2	
4	38	77.2	80.8	80.2	69.7	36.2	59.7	87.1	20.0				7.8-45.6	.90
1	63	88.8	97.1	90.7	81.3	16.6							4.1-33.8	
2	54	124.1	118.6	113.8	121.8	87.9	21.4						6.7-50.9	
3	91	107.2	94.9	96.4	95.9	24.3	42.6	16.8					2.3-41.4	
4	71	103.5	117.1	102.3	103.5	35.3	29.8	24.4	17.8				5.0-29.7	
5	22	72.9	75.6	75.6	64.8	34.8	95.7	54.3	69.7	20.7			8.8-36.6	.88
1	61	88.9	92.7	90.0	81.5	15.8							5.3-29.4	
2	42	124.9	116.1	118.4	121.9	65.4	22.0						2.1-52.3	
3	77	107.5	94.7	94.2	96.1	24.0	44.8	15.6					2.9-35.9	
4	54	98.3	113.1	105.3	94.8	30.1	40.3	23.3	17.1				4.0-32.5	
5	48	112.4	120.2	99.2	112.6	48.5	25.1	31.1	24.5	16.9			6.5-27.6	
6	19	70.9	76.0	73.9	62.7	34.8	100.6	56.6	64.2	82.4	20.3		8.7-35.1	.94
1	59	88.6	92.7	90.1	81.3	15.7							5.1-29.5	
2	43	120.3	106.8	111.6	115.6	53.3	18.2						5.2-42.4	
3	78	106.9	94.7	93.6	95.9	23.8	32.2	15.6					3.5-36.3	
4	55	98.7	113.2	103.9	94.9	29.9	31.5	22.6	16.6				2.7-32.2	
5	33	111.3	125.8	100.0	113.2	52.2	24.1	36.3	25.8	16.1			7.3-25.1	
6	14	128.9	133.4	123.1	129.9	82.0	33.4	63.3	53.9	34.3	18.5		6.6-31.6	
7	19	70.9	76.0	73.9	62.7	34.6	87.2	56.0	63.9	85.6	16.1	20.3	8.7-35.1	.90

Note: Within each triangular cluster matrix mean distances within clusters are read along the diagonal. Distances between clusters are read at the intersection of the vertical and horizontal of respective clusters.

Agreement Between Hierarchical Agglomerative and Iterative Procedures Assignment of Subjects to Clusters

Comparison of the hierarchical agglomerative and iterative clustering procedures, together with the rules for determining the number of clusters, generally pointed to a three cluster solution as being the most optimal representation of this data set. Therefore, the level of agreement between the assignment of individuals to cluster groups by the hierarchical agglomerative solution and the assignment of individuals to the cluster groups by the iterative partitioning procedure was examined next. The classification matrix for this comparison appears in Table 31. Groups were matched according to SB:FE GPAB Area score. From Table 31, the percentage agreement between the two sets identified, 77%, was significant [X^2 (4, 301) = 268.9: $p < .001$].

Table 31

HAP vs. IPP Assignment of Subjects: SB:FE GPAB and WRAT-R Subsample Age 7-11-11

		Hierarchical Agglomerative Procedures			
		Group 1	Group 2	Group 3	n
	Group 1	97	39	0	136
Iterative	Group 2	0	116	30	146
Procedures	Group 3	0	0	19	19
	n	97	155	49	301

Multivariate Profile Analysis

The groups derived through the iterative partitioning procedure were subjected to multivariate profile analysis. There was a significant multivariate main effect for profile elevation [$F(8, 590) = 91.81$, $p < .001$] when the three profiles were tested simultaneously. Univariate F tests for each SB:FE GPAB Reasoning Area SAS were conducted next. The results are presented in Table 32. As noted, this analysis cannot be considered a validation procedure, but enables examination of the differences between the groups on the SB:FE variables used in the clustering procedure. Each of the group means for the SB:FE GPAB variables in the three cluster solution are significantly different. There was also a significant multivariate main effect for the three group solution when the profile line segments were tested for parallelism [$F(6, 592) = 10.9$, $p < .001$]. All three groups were different to each other in the shape of their Quantitative - Short Term Memory SAS line segments. Groups 2 and 3 were also different in the shape of their profiles defined by the Abstract/Visual - Quantitative SAS segments. No other differences were noted. Results are presented in Table 33.

Table 32

Differences Between SB:FE Reasoning Area Group Means for IPP Cluster Solution: SB:FE GPAB and WRAT-R Data Subsample Age 7-11-11

	Means of Iterative Clusters			Univariate F tests (2, 298)			
SB:FE Variable	1 (n=97)	2 (n=155)	3 (n=49)	F ratio	1 vs 2	1 vs 3	2 vs 3
Verbal SAS	116.8	101.7	79.9	165.8	*	*	*
A/V SAS	118.3	100.9	82.2	158.4	*	*	*
Quantitative SAS	109.3	96.4	82.0	115.6	*	*	*
STM SAS	116.0	92.8	72.8	289.3	*	*	*

Multivariate Main Effect for Elevation:
$F(8, 590) = 91.81$, $p < .001$

*denotes pairs of means significantly different at the $p < .01$ level.

Table 33

Tests for Profile Parallelism on IPP Clusters: SB:FE GPAB and WRAT-R Data Subsample Age 7-11-11

	Iterative Clusters		
SB:FE Segments	1 vs 2	1 vs 3	2 vs 3
Verbal - A/V SAS	NS	NS	NS
A/V - Quantitative SAS	NS	NS	*
Quantitative - STM SAS	*	*	*

Multivariate Main effect for Parallelism:
$F(6, 592) = 10.9$, $p < .001$

*denotes pairs of line segments significantly different at the $p < .05$ level; NS denotes pairs of line segments not significantly different.

Cross-validation of Cluster Solutions for the Subsample Age 7-11-11

As a final means of testing the validity of the solutions derived in the previous stage, identical clustering techniques (hierarchical agglomerative and iterative partitioning procedures) were applied to a different subsample of subjects with SB:FE GPAB data of the same age (n=226). This sample comprised the subjects included in Part One of the study for whom WRAT-R data were not available.

Determination of the Number of Clusters

Presented in Table 34 are the values of the fusion coefficients corresponding to the last 10 (out of 226) steps of the hierarchical agglomerative algorithm used to identify the clusters within this sample which display differences between profile elevation and/or shape. Inspection of the results indicate that between the 3 and the 2 cluster solution, there was an increase in the percentage of change in the fusion coefficient, followed by a further extreme increase as the last two groups or clusters are forced together. This pattern strongly points to the presence of a three cluster solution as being the optimal solution.

Table 34

HAP Fusion Coefficients: SB:FE Data Only Subsample Age 7-11-11

Number of Clusters	Clustering Coefficient	Percentage Change From Previous Step
10	74790.23	5.32
9	79207.87	5.90
8	84461.37	6.63
7	91374.43	8.18
6	98309.87	7.59
5	108515.75	10.38
4	120810.01	11.32
3	138588.75	14.71
2	188213.43	35.81
1	288718.62	53.39

To further confirm that the three cluster solutions optimally represented this data set, a k-means iterative procedure was performed for 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 11- and 12-clusters. The results of this analysis, along with the number of subjects and means for each group yielded by this procedure are reported in Table 35. For parsimony, only the results of the analysis 2 through 7 solutions are reported. Examination of the triangular matrices of sample mean Squared Euclidean distances within clusters, across the clustering solutions support a three cluster interpretation. The three cluster solution again maintains homogeneity within groups as well as the separation between groups. It thus appears to be the most optimal. The four cluster solution does not appear to increase cluster homogeneity, with relatively poor separation between the first and second clusters. The NNR for the two cluster solution was .59, which is lower than the three cluster solution, however the NNR for the three cluster solution was .66 which is acceptable. The NNR for the four cluster solution increased to .98. The NNR's for the solutions, 5, 6, and 7 are 1.00, 1.14, and .1.07 respectively. Not shown in Table 35 are the NNR's for the cluster solutions 8 through to 12. These were all above 1.00, and are thus unacceptable. As can be seen from the means reported in column 3-6, Table 35, and the corresponding graph in Figure 5, the three cluster solution groups the subjects into a below average group, an average group, and an above average group.

Table 35

IPP Data for 2-7 Cluster Solutions: SB:FE Only Subsample 7-11-11

		Reasoning Area Means				Mean D^2 Between and Within Clusters								
k	n	Ver. SAS	A/V SAS	Qnt. SAS	STM SAS	1	2	3	4	5	6	7	Range of Distances From Centre	NNR
1	130	92.8	91.2	89.8	86.4	23.2							7.5-100.2	
2	96	114.8	113.9	104.8	108.9	40.6	24.0						9.0-60.3	.59
1	140	100.4	98.8	93.5	94.4	18.8							5.9-45.7	
2	59	118.0	118.7	111.8	113.5	37.2	24.5						7.8-45.3	
3	27	76.8	72.1	76.2	65.8	48.6	85.5	22.7					10.6-64.3	.66
1	101	96.0	95.5	93.7	90.8	18.0							6.2-46.8	
2	72	111.9	109.4	96.1	104.1	26.3	19.1						5.7-39.3	
3	30	120.2	122.1	120.9	119.1	56.9	33.5	25.9					12.8-42.4	
4	23	75.0	69.7	75.0	63.1	45.3	71.2	101.8	22.5				9.7-63.5	.98
1	51	89.9	94.5	89.3	84.0	17.3							1.6-28.7	
2	31	121.1	120.8	120.1	119.2	48.8	20.6						9.7-42.1	
3	71	104.5	95.0	97.9	97.0	24.8	25.4	16.7					4.6-44.3	
4	55	109.5	114.7	94.9	103.9	82.0	34.5	59.4	24.8				14.5-39.3	
5	18	72.8	65.1	72.0	61.6	47.1	95.1	69.8	128.7	18.2			4.0-47.4	1.00
1	44	87.4	87.4	89.9	82.5	17.1							3.3-28.3	
2	19	120.3	119.9	130.4	117.7	50.6	19.6						5.8-35.4	
3	57	104.8	95.0	96.7	101.6	24.0	28.7	15.5					6.4-40.8	
4	53	101.4	107.5	93.7	90.1	34.7	24.6	21.4	17.0				4.1-36.6	
5	40	117.5	119.0	99.3	112.3	79.4	30.1	58.6	48.8	24.5			12.2-36.6	
6	13	69.9	61.1	65.9	58.4	47.8	97.1	68.7	81.0	126.7	18.2		4.0-47.4	1.14
1	41	86.9	86.7	90.4	81.8	17.2							3.7-28.4	
2	28	114.8	110.8	115.4	106.2	53.7	16.2						7.2-36.4	
3	51	106.4	94.1	96.7	100.7	23.7	31.0	15.3					5.4-40.7	
4	58	116.8	123.2	95.5	114.7	40.5	25.9	25.9	18.5				3.7-37.2	
5	26	99.5	106.6	91.7	92.0	41.5	22.4	23.6	23.5	18.6			8.3-35.6	
6	9	130.0	128.0	134.9	127.2	82.4	31.4	61.4	48.5	45.6	24.1		13.3-38.6	
7	13	69.9	61.1	65.7	58.4	48.1	99.8	68.8	87.2	87.6	129.8	18.2	4.0-47.4	1.07

Note: Within each triangular cluster matrix mean distances within clusters are read along the diagonal. Distances between clusters are read at the intersection of the vertical and horizontal of respective clusters.

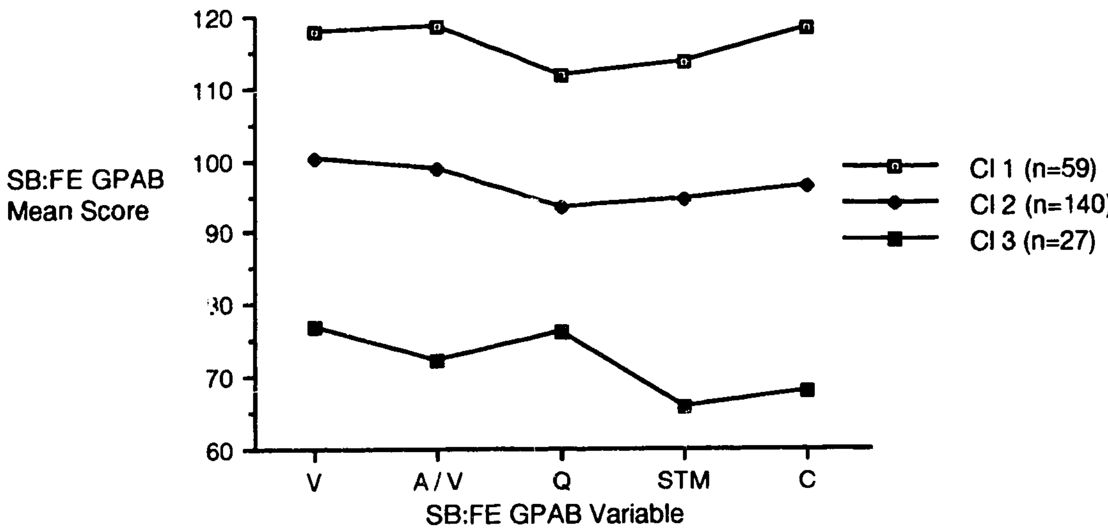

Figure 5. Graph displaying cluster scores on SB:FE GPAB Reasoning Area and Composite SAS's: SB:FE GPAB only subsample age 7-11-11.

Agreement Between Hierarchical Agglomerative and Iterative Procedures Assignment of Subjects to Clusters

The level of agreement between the hierarchical agglomerative solutions and the iterative solutions was examined next. The classification matrix for this comparison appears in Table 36. From Table 36, the percentage agreement between the two sets identified, 92%, was significant [X^2 (4, 226) = 301.5: $p < .001$].

Table 36

HAP vs. IPP Assignment of Subjects: SB:FE GPAB Only Subsample Age 7-11-11

		Hierarchical Agglomerative Procedures			
		Group 1	Group 2	Group 3	n
	Group 1	59	3	0	62
Iterative	Group 2	0	[illegible]	15	152
Procedures	Group 3	0		12	12
	n	59		27	226

Multivariate Profile Analysis

The groups derived through the iterative partitioning procedures solution were then subjected to multivariate profile analysis. There was a significant multivariate main effect for profile elevation [$F(8, 440) = 62.90$, $p < .001$]. The results, presented in Table 37, indicate that each of the group means for the SB:FE GPAB variables in the three cluster solution are significantly different from each other. There was also a significant multivariate main effect for the three group solution when the profile line segments were tested for parallelism [$F(6, 442) = 2.66$, $p < .05$]. Univariate comparisons indicated that all three groups were similar to each other in the shape (segment) of the profile defined by their Verbal - Abstract/Visual SAS scores. Groups 1 and 3, and Groups 2 and 3 were different to each other in the shape of their profiles defined by the Abstract/Visual - Quantitative and Quantitative - Short Term Memory SAS line segments. The results are presented in Table 38.

Table 37

Differences Between SB:FE Reasoning Area Group Means for IPP Solution: SB:FE GPAB Only Subsample Age 7-11-11

	Means of Iterative Clusters			Univariate F tests (2, 223)			
SB:FE Variable	1 (n=59)	2 (n=140)	3 (n=27)	F ratio	1 vs 2	1 vs 3	2 vs 3
Verbal SAS	116.8	100.4	76.8	137.7	*	*	*
A/V SAS	118.7	98.8	72.1	148.3	*	*	*
Quantitative SAS	111.8	93.5	76.2	97.2	*	*	*
STM SAS	113.5	94.4	65.8	176.8	*	*	*

Multivariate Main Effect for Elevation:
$F(8, 440) = 62.90, p < .001$

*denotes pairs of means significantly different at the $p < .05$ level.

Table 38

Tests for Profile Parallelism on IPP Clusters: SB:FE GPAB Only Subsample Age 7-11-11

	Iterative Clusters		
SB:FE Segment	1 vs 2	1 vs 3	2 vs 3
Verbal - A/V SAS	NS	NS	NS
A/V - Quantitative SAS	NS	*	*
Quantitative - STM SAS	NS	*	*

Multivariate Main effect for Parallelism:
$F(6, 442)\ 2.66, p < .05$

*denotes pairs of line segments significantly different at the $p < .05$ level; NS denotes pairs of line segments not significantly different.

Comparison between the solutions derived from the application of hierarchical agglomerative and iterative partitioning procedures to the two subsamples. Overall the results suggest some similarity between the cluster solutions applied to the two subsamples with both solutions yielding high average, average, and low average clusters, as can be seen in Table 39. A three cluster solution appeared to optimally represent both data sets.

Table 39

SB:FE GPAB Reasoning Area and Composite SAS Cluster Means for the Two Samples Age 7-11-11

Subsample	n	Ver. SAS	A/V SAS	Qnt. SAS	STM SAS	Com. SAS
SB:FE with WRAT-R data						
Cluster 1	97	116.7	118.3	109.3	116.0	117.8
Cluster 2	155	101.5	100.9	96.4	92.8	97.7
Cluster 3	49	79.9	82.2	82.1	72.8	76.1
Total	301	102.9	103.5	98.2	97.0	100.6
SB:FE Data only						
Cluster 1	59	117.9	118.7	111.8	113.5	118.2
Cluster 2	140	100.4	98.8	93.5	94.4	96.3
Cluster 3	27	76.81	72.1	76.2	65.8	67.9
Total	226	102.2	100.8	96.2	95.96	98.6

Cluster Analytic Procedures Applied to the Subsample Age 12-23-11

The subsample of subjects for whom both SB:FE GPAB and WRAT-R data were available in the age group 12-23-11 comprised 186 subjects.

Determination of the Number of Clusters

Presented in Table 40 are the values of the fusion coefficients corresponding to the last 10 (out of 186) steps of the hierarchical agglomerative algorithm used to identify the clusters within this sample which display differences between profile elevation and/or shape. Inspection of the results reveals a slight increase in fusion values when going from 4 to 3 clusters, followed by an even larger increase in fusion values when going from 3 to 2 clusters. Thus, it would appear that there are 3 subgroups within the 12-23-11 age group, although a 4 cluster solution may also warrant consideration.

Table 40

HAP Fusion Coefficients: SB:FE GPAB and WRAT-R Data Subsample 12-23-11

Number of Clusters	Clustering Coefficient	Percentage Change From Previous Step
10	52794.6	6.23
9	56239.8	6.77
8	61120.9	6.53
7	66264.0	8.68
6	71987.8	8.41
5	79972.1	8.64
4	89236.6	11.09
3	115280.8	29.19
2	173367.9	50.39
1	273310.1	57.65

To further determine whether a 4 or 3 cluster solution optimally represented this data set, a *k*-means iterative procedure was performed for 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 11- and 12-clusters. The results of this analysis, along with the number of subjects and means for each group yielded by this procedure are reported in Table 41. Only the results of the analysis 2 through 7 solutions are reported. Examination of the triangular matrices of sample mean D^2 distances between and within clusters presented in Table 41, across clustering solutions generally suggests that the 3 cluster solution is optimal. The three cluster solution maintains the homogeneity within groups and separation between groups. The four cluster solution does not appear to significantly increase cluster homogeneity, with relatively poor separation between the first and second groups. The NNR for the two cluster solution was .47, which is lower than the three cluster solution (.55), which was also acceptable. The NNR for the four cluster solution increased to .69 which is unacceptable. The NNR's for the solutions, 5, 6, and 7 are .81, .88, and .83 respectively. Not shown in Table 41 are the NNR's for the cluster solutions 8 through to 12. These were all above .85, and are thus unacceptable. As can be seen from the means reported in columns 3-6, Table 41, and the corresponding graph of this solution displayed in Figure 6, the three cluster solution groups the subjects into a below average group, an average group, and an above average group.

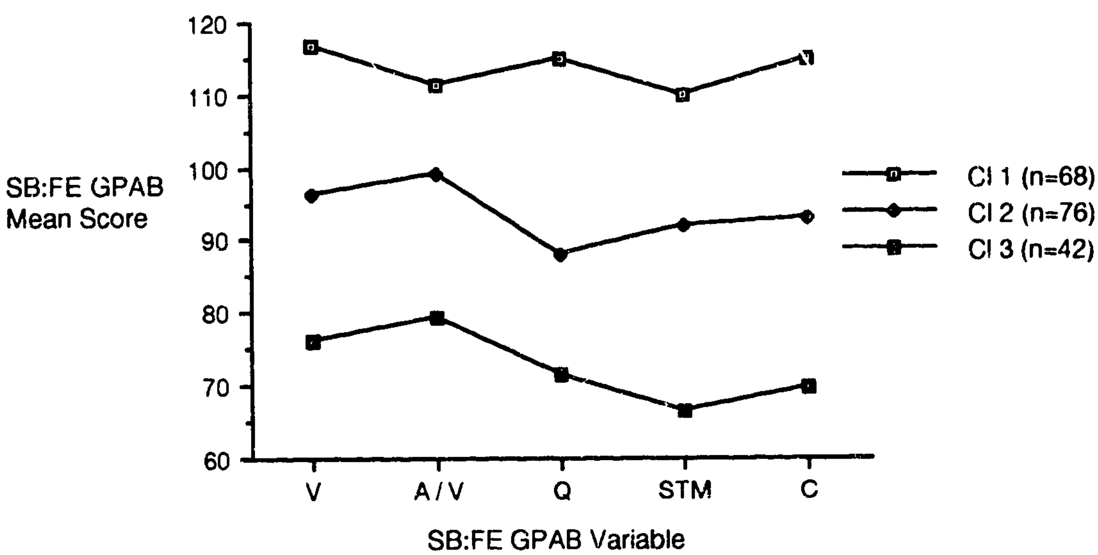

Figure 6. Graph displaying clusters scores on SB:FE GPAB Reasoning Area and Composite SAS's: SB:FE GPAB and WRAT-R data subsample age 12-23-11.

Table 41

IPP Data for 2-7 Cluster Solutions: SB:FE GPAB and WRAT-R Data Subsample 12-23-11

		Reasoning Area Means				Mean D^2 Between and Within Clusters								
k	n	Ver. SAS	A/V SAS	Qnt. SAS	STM SAS	1	2	3	4	5	6	7	Range of Distances From Centre	NNR
1	103	111.4	109.2	107.6	105.6	24.6							6.0-48.5	
2	83	84.4	86.7	77.5	76.9	54.4	25.9						7.5-68.6	.47
1	76	96.4	99.1	88.2	92.1	21.5							6.2-37.2	
2	68	116.7	111.3	114.8	109.8	39.8	21.8						3.1-41.9	
3	42	76.2	79.5	71.6	66.6	41.4	79.9	22.2					7.4-45.4	.55
1	56	91.4	90.4	83.2	83.0	21.1							10.6-36.9	
2	58	103.6	109.2	95.2	100.3	30.6	18.5						8.2-33.6	
3	46	119.7	110.7	120.8	113.1	59.4	32.9	20.3					6.5-40.9	
4	26	70.6	74.8	68.4	61.2	37.1	67.2	95.6	20.8				5.7-43.9	.69
1	47	119.5	110.6	120.9	112.4	20.6							6.7-40.8	
2	23	70.6	71.8	67.6	60.1	97.3	20.4						6.0-41.1	
3	38	95.9	83.4	85.3	86.2	56.8	42.1	18.9					9.7-35.7	
4	54	104.1	[illegible]	94.6	100.9	32.4	70.4	32.3	17.9				7.3-31.0	
5	24	81.9	[illegible]	80.5	77.7	65.6	39.6	25.5	35.8	17.1			11.1-28.0	.81
1	42	119.9	111.1	122.0	113.9	19.7							6.9-41.6	
2	22	69.8	72.2	67.3	59.4	100.0	20.2						5.6-40.6	
3	34	96.4	81.9	82.4	86.1	61.2	41.8	17.8					10.6-31.6	
4	35	107.9	110.7	90.5	107.2	34.4	75.8	38.2	16.2				4.8-29.8	
5	24	80.8	101.9	79.9	78.6	68.1	39.1	[illegible]	41.6	17.0			11.0-27.5	
6	29	100.1	106.0	104.2	89.9	36.2	65.9	32.8	23.7	33.2	15.5		6.9-27.0	.88
1	42	119.9	111.1	122.0	113.9	19.7							6.9-41.6	
2	5	54.0	56.4	52.0	51.8	126.9	19.5						12.8-24.0	
3	31	97.7	82.8	83.1	86.8	59.5	69.2	17.5					10.7-28.1	
4	35	107.9	110.7	90.5	89.9	34.4	101.9	36.7	16.2				4.8-29.8	
5	22	81.0	103.4	80.5	79.1	67.2	67.0	27.6	40.7	16.8			10.6-27.1	
6	29	100.1	106.0	104.2	107.2	36.2	93.5	31.5	23.7	32.3	15.5		4.9-27.0	
7	22	76.1	77.1	72.4	65.1	89.0	38.8	32.9	65.0	31.2	55.1	15.1	8.0-28.6	.83

Note: Within each triangular cluster matrix mean distances within clusters are read along the diagonal. Distances between clusters are read at the intersection of the vertical and horizontal of respective clusters.

Agreement Between Hierarchical Agglomerative and Iterative Procedures Assignment of Subjects to Clusters

Comparison of the hierarchical agglomerative and iterative clustering procedures, together with the rules for determining the number of clusters, generally pointed to a three cluster solution as being the most optimal representation of the data set. Therefore the level of agreement between the assignment of individuals to cluster groups by the hierarchical agglomerative solution and the assignment of individuals to the cluster groups by the iterative partitioning procedure was examined next. The classification matrix for this comparison appears in Table 42. Again, groups were matched according to SB:FE GPAB Area score. From Table 42, the percentage agreement between the two sets identified, 78%, was significant [X^2 (4, 186) = 193.2: $p < .001$]. This finding tends to support the internal validity of the obtained three cluster solutions.

Table 42

HAP vs. IPP Assignment of Subjects: SB:FE GPAB and WRAT-R Data Subsample Age 12-23-11

		Hierarchical Agglomerative Procedures			
		Group 1	Group 2	Group 3	n
	Group 1	45	0	0	45
Iterative	Group 2	23	76	17	116
Procedures	Group 3	0	0	25	25
	n	68	76	49	186

Multivariate Profile Analysis

The groups derived through the iterative partitioning procedure were subjected to multivariate profile analysis. There was a significant multivariate main effect for profile elevation [$F(8, 360) = 67.51$, $p < .001$] when the three profiles were tested simultaneously. Therefore univariate F tests for each SB:FE GPAB Reasoning Area SAS were conducted next. The results are presented in Tables 43. This analysis cannot be considered a validation procedure but highlights the significant differences between the groups on the SB:FE variables used in the clustering procedure. There was also a significant multivariate main effect for the three group solution when the profile line segments were tested for parallelism [$F(6, 362) = 6.55, p < .001$]. All three groups were different from each other in the shape of the portion of the profile defined by their Verbal - Abstract/Visual SAS scores. Groups 1 and 3 were also different from each other in the shape of the portion of the profile defined by their Quantitative - Short Term Memory SAS line segments. Finally, Groups 2 and 3 also differed with respect to the shape of their Abstract/Visual - Quantitative SAS line segments. Results are presented in Table 44.

Table 43

Differences Between SB:FE Reasoning Area Group Means for IPP Cluster Solution: SB:FE GPAB and WRAT-R Data Subsample Age 12-23-11

	Means of Iterative Clusters			Univariate F tests (2, 183)			
SB:FE Variable	1 (n=68)	2 (n=76)	3 (n=42)	F ratio	1 vs 2	1 vs 3	2 vs 3
Verbal SAS	116.8	96.4	76.2	166.4	*	*	*
A/V SAS	111.3	99.1	79.5	88.6	*	*	*
Quantitative SAS	114.8	88.2	71.6	184.6	*	*	*
STM SAS	109.8	92.1	66.6	183.8	*	*	*

Multivariate Main Effect for Elevation:
$F(8, 360) = 67.51, p < .001$

*denotes pairs of means significantly different at the $p < .01$ level.

Table 44

Tests for Profile Parallelism on IPP Clusters: SB:FE GPAB and WRAT-R Data Subsample Age 12-23-11

	Iterative Clusters		
SB:FE Segment	1 vs 2	1 vs 3	2 vs 3
Verbal - A/V SAS	*	NS	*
A/V - Quantitative SAS	*	NS	*
Quantitative - STM SAS	*	*	NS

Multivariate Main effect for Parallelism:
$F(6, 362)$ 6.55, $p < .001$

*denotes pairs of line segments significantly different at the $p < .01$ level; NS denotes pairs of line segments not significantly different.

Cross-validation of Cluster Solutions for the Subsample Age 12-23-11

Identical clustering techniques (hierarchical agglomerative and iterative partitioning procedures) were applied to a different subsample of subjects with SB:FE GPAB data of the same age (n=113). This sample comprised the subjects included in Part One of the study for whom WRAT-R data were not available.

Determination of the Number of Clusters

Presented in Table 45 are the values of the fusion coefficients corresponding to the last 10 (out of 113) steps of the hierarchical agglomerative algorithm used to identify the clusters within this sample which display differences between profile elevation and/or shape. Inspection of the results indicate that between the 4 and the 3 cluster solution, there was an increase in the percentage of change in the fusion coefficient. This is followed by a slightly lower jump between the 3 and the 2 cluster solution, and as the last two groups are forced together, a large jump, pointing to either a four or a three cluster solution as being the optimal.

Table 45

HAP Fusion Coefficients: SB:FE Only Subsample Age 12-23-11

Number of Clusters	Clustering Coefficient	Percentage Change From Previous Step
10	26656.26	9.84
9	29386.17	10.24
8	32428.14	10.35
7	36265.16	11.83
6	40774.33	12.43
5	46398.82	13.79
4	52497.72	13.14
3	69522.06	32.43
2	103487.12	48.86
1	185777.18	79.52

To assist in determining the optimal solution for this particular data set, a k-means iterative procedure was performed for 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 11- and 12-clusters. The results of this analysis, along with the number of subjects and means for each group yielded by this procedure are reported in Table 46. Results of the analysis for 2 through 7 solutions are reported. Examination of the triangular matrices of sample mean Squared Euclidean distances within clusters, across the clustering solutions support a three cluster interpretation, and suggests that a four cluster solution is not acceptable. The two cluster solution is also acceptable, although the three cluster solution maintains homogeneity within groups as well as separation between groups. The four cluster solution does not appear to increase cluster homogeneity, with relatively poor separation between the first and second clusters, and the second and third clusters. The NNR for the two cluster solution was .51, which is lower than the three cluster solution, however the NNR for the three cluster solution was .61 which is acceptable. The NNR for the four cluster solution increased to .75 which is unacceptable. The NNR's for the solutions, 5, 6, and 7 are 1.00, .98, and .75 respectively. Not shown in Table 46 are the NNR's for the cluster solutions 8 through to 12. These were all above 90, and are thus unacceptable. As can be seen from the means reported in columns 3-6, Table 46, and the corresponding graph in Figure 7, the three cluster solution groups the subjects into a below average group, an average group, and an above average group.

Table 46

IPP Data for 2-7 Cluster Solutions: SB:FE GPAB Only Subsample 12-23-11

		Reasoning Area Means				Mean D^2 Between and Within Clusters								
k	n	Ver. SAS	A/V SAS	Qnt. SAS	STM SAS	1	2	3	4	5	6	7	Range of Distances From Centre	NNR
1	43	78.7	80.8	73.4	72.8	29.4							6.1-74.4	
2	70	107.4	108.9	102.4	102.5	57.7	24.6						9.0-60.3	.51
1	69	97.6	101.3	87.7	90.9	22.7							7.7-39.0	
2	27	115.3	110.4	118.0	113.4	44.6	21.7						5.6-43.3	
3	17	62.1	66.2	64.0	57.4	64.5	104.1	27.3					10.1-46.5	.61
1	31	87.1	89.0	79.5	81.3	19.7							9.8-34.7	
2	50	102.4	107.2	95.0	96.9	32.4	18.8						6.4-38.7	
3	20	119.7	113.0	120.8	116.5	67.6	37.1	18.6					2.1-37.0	
4	12	57.3	59.5	57.7	50.8	56.3	88.5	122.7	24.4				9.6-34.2	.75
1	20	119.7	113.0	120.8	116.5	18.6							2.1-37.6	
2	12	57.3	59.5	57.7	50.8	122.7	24.4						9.6-34.2	
3	25	93.6	84.5	83.1	86.6	61.6	62.2	18.6					10.9-31.8	
4	43	103.2	108.8	95.6	97.9	35.5	90.6	31.1	18.2				7.6-38.6	
5	13	80.1	102.2	78.9	76.5	70.9	58.8	24.9	36.2	15.7			10.3-19.8	1.00
1	16	120.8	113.0	121.4	121.9	16.9							3.8-33.0	
2	12	57.3	59.5	57.7	50.8	126.5	24.4						9.6-30.8	
3	22	92.7	83.2	82.0	85.7	67.3	60.2	18.0					11.1-30.7	
4	25	103.3	109.4	107.9	91.4	43.0	90.1	32.4	16.3				4.4-31.2	
5	13	80.1	102.2	78.9	76.5	75.0	58.8	24.8	37.5	15.7			11.5-19.8	
6	25	104.8	107.1	86.5	103.3	37.7	93.7	38.8	24.6	40.6	14.7		3.5-30.1	.98
1	16	120.8	113.0	121.4	121.9	16.9							3.8-33.0	
2	6	46.5	48.3	46.7	42.5	146.2	14.4						9.3-25.8	
3	20	94.5	84.2	81.8	86.6	65.7	82.2	18.0					10.4-25.2	
4	24	105.7	106.8	86.3	103.6	42.7	110.5	30.6	16.0				6.6-31.5	
5	12	81.3	106.2	80.5	79.0	71.4	83.8	26.7	35.0	15.5			9.7-21.9	
6	25	103.3	109.4	107.9	91.4	37.7	114.5	37.6	24.0	37.4	14.7		8.1-30.1	
7	10	70.7	74.2	73.0	64.8	98.0	49.4	34.9	62.9	37.3	65.0	16.4	9.3-23.7	.75

Note: Within each triangular cluster matrix mean distances within clusters are read along the diagonal. Distances between clusters are read at the intersection of the vertical and horizontal of respective clusters.

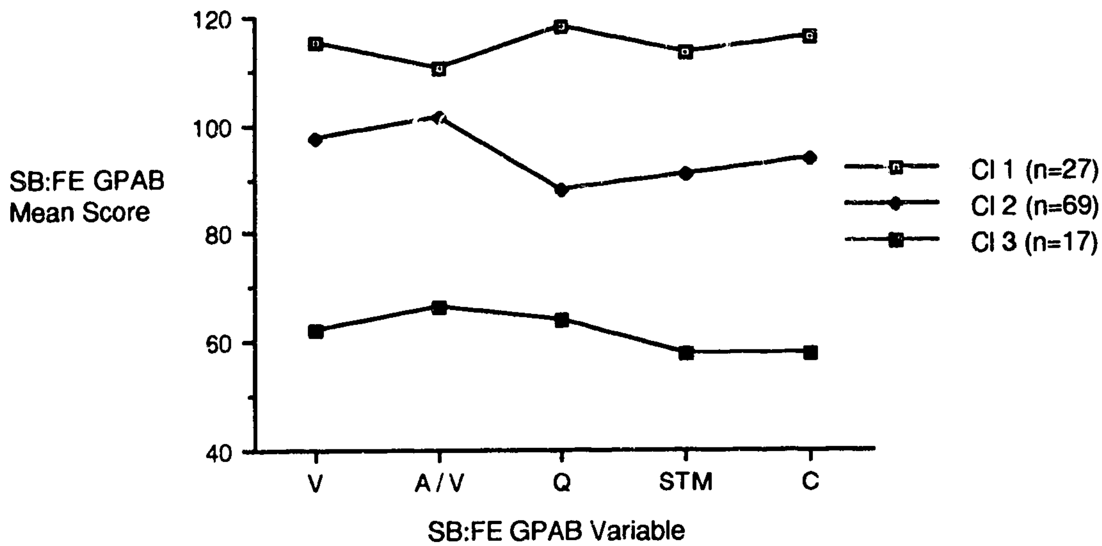

Figure 7. Graph displaying cluster scores on SB:FE GPAB Reasoning Area and Composite SAS's: SB:FE GPAB only subsample age 12-23-11.

Agreement Between Hierarchical Agglomerative and Iterative Procedures Assignment of Subjects to Clusters

The level of agreement between the hierarchical agglomerative solutions and the iterative partitioning solutions was examined next. The classification matrix for this comparison appears in Table 47. From Table 47, the percentage agreement between the two sets identified, 63%, was significant [X^2 (4, 113) = 96.3: $p < .001$]. This finding tends to support the internal validity of the obtained solutions.

Table 47

HAP vs. IPP Assignment of Subjects: SB:FE GPAB Data Only Subsample Age 12-23-11

		Hierarchical Agglomerative Procedures			
		Group 1	Group 2	Group 3	n
	Group 1	26	36	0	62
Iterative	Group 2	1	33	5	39
Procedures	Group 3	0	0	12	12
	n	27	69	17	113

Multivariate Profile Analysis

The groups derived through the iterative partitioning procedures were then subjected to multivariate profile analysis. Significant differences [$F(8, 214) = 40.66$, $p < .001$] were noted between all group means in all Reasoning Area SAS's. The results, presented in Table 48, indicate that each of the group means for the SB:FE GPAB variables in the three cluster solution are significantly different from each other. There was also a significant multivariate main effect for the three group solution when the profile line segments were tested for parallelism [$F(6, 216) = 8.62$, $p < .001$]. Groups 1 and 2 were different from each other in the segments of the profile defined by their Verbal - Abstract/Visual and Abstract/Visual - Quantitative SAS line segments. Groups 1 and 3 differed with respect to their Abstract/Visual - Quantitative SAS line segments. No differences were noted between Groups 2 and 3. The results are presented in Table 49.

Table 48

Differences Between SB:FE Reasoning Area Group Means for IPP Cluster Solution: SB:FE GPAB Data Subsample Age 12-23-11

	Means of iterative Clusters			Univariate F tests (2, 110)			
SB:FE Variable	1 (n=27)	2 (n=69)	3 (n=17)	F ratio	1 vs 2	1 vs 3	2 vs 3
Verbal SAS	115.3	97.6	62.1	110.4	*	*	*
A/V SAS	110.4	101.3	66.2	66.7	*	*	*
Quantitative SAS	118.0	87.7	64.0	106.2	*	*	*
STM SAS	113.4	90.9	57.4	97.1	*	*	*

Multivariate Main Effect for Elevation:
$F(8, 214) = 40.66, p < .001$

*denotes pairs of means significantly different at the $p < .05$ level.

Table 49

Tests for Profile Parallelism on IPP Clusters: SB:FE GPAB Data Only Subsample 12-23-11

	Iterative Clusters		
SB:FE Segment	1 vs 2	1 vs 3	2 vs 3
Verbal - A/V SAS	*	NS	NS
A/V - Quantitative SAS	*	*	NS
Quantitative - STM SAS	NS	NS	NS

Multivariate Main effect for Parallelism:
$F(6, 216)$ 8.62, $p < .05$

*denotes pairs of line segments significantly different at the $p < .05$ level; NS denotes pairs of line segments that are not significantly different.

Comparison betwe… t… solutions derived from the application of hierarchical agglomerative and iterative partitioning procedures to the two subsamples. Overall, the results suggest some similarity between the cluster solutions applied to the two samples. The second sample has slightly lower means across all Reasoning Area and Composite SAS scores (see Table 50). This likely reflects differences in the sample composition. However, the main trend to emerge, that cognitive profiles appe…red to be based primarily on overall cognitive ability, and that a three cluster solutions appeared optimal, tends to support the findings of the first sample, particularly when data obtained from the iterative partitioning procedures were examined.

Table 50

SB:FE GPAB Reasoning Area and Composite SAS Cluster Means for the Two Samples Age 12-23-11

Subsample	n	Ver. SAS	A/V SAS	Qnt. SAS	STM SAS	Com. SAS
SB:FE with WRAT-R data						
Cluster 1	68	116.8	111.3	114.8	109.8	114.9
Cluster 2	76	96.4	99.1	88.2	92.1	93.2
Cluster 3	42	76.2	79.5	71.6	66.6	69.8
Total	186	99.3	99.1	94.2	92.8	95.8
SB:FE Data only						
Cluster 1	27	115.3	110.4	118.0	113.4	116.1
Cluster 2	69	97.6	101.4	87.7	90.9	93.6
Cluster 3	17	62.1	66.2	64.0	57.4	57.7
Total	113	96.5	98.2	91.4	91.2	93.6

Cluster Analytic Procedures Applied to the Subsample Age 2-4-11

As pointed out, the WRAT-R is administered to subjects age 5-00 and older. No subjects in the age group 2-4-11 were therefore administered the WRAT-R. However, SB:FE GPAB scores were available for 110 subjects within this age group. Clustering procedures were therefore applied to this subsample.

Determination of the Number of Clusters

Presented in Table 51 are the values of the fusion coefficients corresponding to the last 10 (out of 110) steps of the hierarchical agglomerative algorithm used to identify the clusters or within this sample which display differences between profile elevation and/or shape. Inspection of the results reveals a sudden increase in fusion values when going from 2 to 1 cluster. It would appear that there are two subgroups within the 2-4-11 age group.

Table 51

HAP Fusion Coefficients: SB:FE Subsample Age 2-4-11

Number of Clusters	Clustering Coefficient	Percentage Change From Previous Step
10	27720.35	6.93
9	29518.43	6.49
8	31757.08	7.58
7	34085.72	7.33
6	36545.82	7.22
5	40025.48	9.52
4	45863.78	14.59
3	51878.58	13.11
2	59736.07	15.15
1	98736.06	65.29

To further confirm that the two cluster solution optimally represented this data set, a k-means iterative procedure was performed for 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 11- and 12-clusters. The results of this analysis, along with the number of subjects and means for each group yielded by this procedure are reported in Table 52. Results of the analysis 2 through 7 solutions are reported. Examination of the triangular matrices of sample mean D^2 distances within and between clusters, presented in Table 52, across clustering solutions generally suggests that the two cluster solution is optimal. There is relatively poor separation between the second and third groups in the three cluster solution. The NNR for the two cluster solution was .57. The NNR's for the three cluster solution, .82, and the four cluster solution, .75, are both unacceptable. The NNR's for the solutions, 5, 6, and 7 are .89, .79, and .76 respectively. Not shown in Table 52, are the NNR's for the cluster solutions 8 through to 12. These were all above .8, and are thus unacceptable. As can be seen from the means in column 3-6, Table 52, and the corresponding graph of this solution displayed in Figure 8, the two cluster solution groups the subjects into a high average group and an above average group.

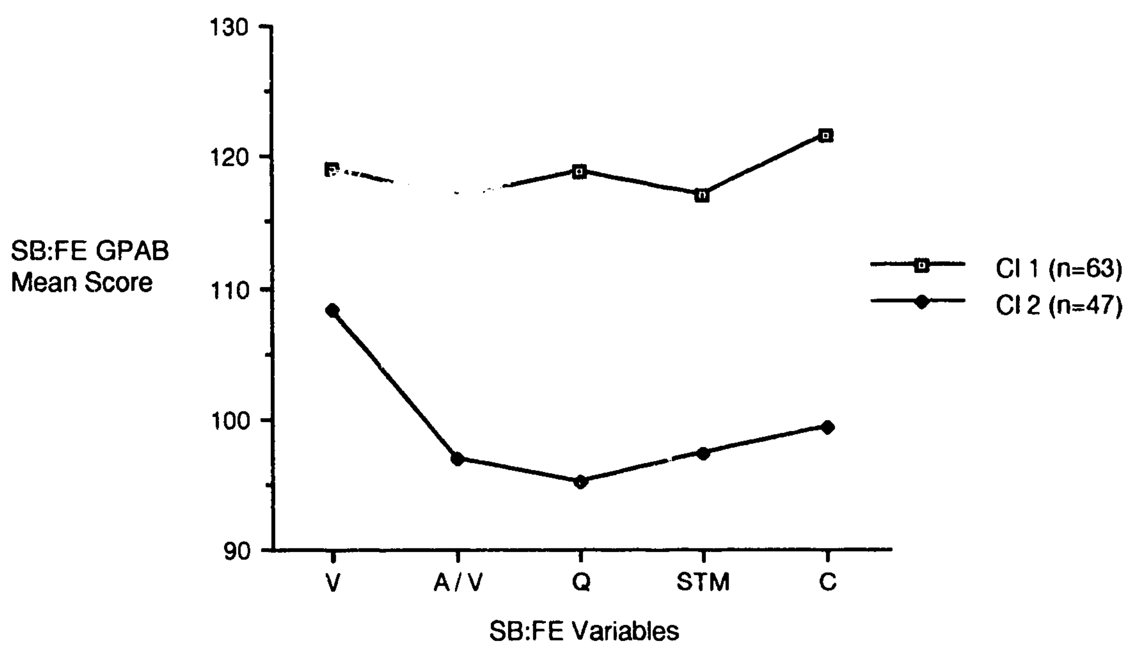

Figure 8. Graph displaying cluster scores on SB:FE GPAB Reasoning Area and Composite SAS: SB:FE only subsample age 2-4-11.

Table 52

IPP Data for 2-7 Cluster Solutions: SB:FE Subsample Age 2-4-11

		Reasoning Area Means				Mean D^2 Between and Within Clusters								
k	n	Ver. SAS	A/V SAS	Qnt. SAS	STM SAS	1	2	3	4	5	6	7	Range of Distances From Centre	NNR
1	63	119.0	116.9	118.9	117.0	22.2							8.1-40.4	
2	47	108.4	96.9	95.1	97.4	38.3	20.9						6.9-41.0	.57
1	39	107.7	95.9	92.5	95.7	19.1							4.2-40.6	
2	49	130.8	125.2	124.9	120.7	32.4	19.2						7.0-38.5	
3	22	112.6	110.8	114.5	113.6	55.3	26.3	21.7					4.6-32.3	.82
1	31	130.3	127.5	125.5	120.9	16.7							6.2-35.4	
2	39	103.1	95.5	93.5	93.0	36.5	17.7						7.4-35.7	
3	20	123.9	104.4	95.2	112.8	59.6	27.7	20.8					6.6-32.5	
4	20	110.7	110.9	119.2	112.6	30.1	28.2	39.4	19.1				11.6-29.6	.75
1	27	109.7	106.7	127.2	108.7	16.7							5.4-33.8	
2	27	101.0	97.6	93.5	91.8	37.5	15.7						8.0-33.1	
3	19	115.8	115.1	108.2	117.6	39.7	23.3	15.0					5.5-28.0	
4	20	130.3	127.5	125.5	120.9	60.1	25.9	31.7	20.8				6.6-32.9	
5	17	120.6	94.2	93.6	106.6	24.8	28.1	37.4	49.1	15.5			6.8-26.2	.89
1	25	100.9	97.8	94.2	90.6	16.4							5.3-33.1	
2	24	116.3	114.3	107.9	117.5	37.6	14.3						7.9-24.7	
3	18	108.3	106.4	126.0	109.2	38.5	22.8	14.2					5.6-28.4	
4	14	136.2	119.0	124.3	125.4	61.5	27.3	34.5	19.5				4.8-33.5	
5	10	117.0	136.0	126.0	112.4	56.0	28.0	31.0	28.6	17.9			10.4-26.4	
6	19	118.7	94.3	92.7	106.7	24.2	27.4	36.9	47.4	53.0	18.0		6.6-27.[illegible]	.79
1	27	101.0	97.6	93.5	91.2	16.7							5.4-33.8	
2	25	115.5	114.3	108.2	117.5	36.9	14.6						8.6-24.7	
3	17	109.0	106.0	126.6	108.6	38.8	22.9	13.8					5.1-28.3	
4	11	134.1	121.5	130.7	116.9	60.6	30.0	30.8	17.3				8.8-27.0	
5	8	114.0	138.0	124.0	114.1	56.5	28.6	32.9	27.0	17.0			11.8-24.6	
6	17	120.6	94.2	93.6	106.6	24.8	27.5	36.9	49.0	54.1	17.5		6.8-26.2	
7	5	138.0	117.2	112.0	135.4	65.2	29.1	43.5	26.9	40.0	44.6	15.8	14.5-19.3	.76

Note: Within each triangular cluster matrix mean distances within clusters are read along the diagonal. Distances between clusters are read at the intersection of the vertical and horizontal of respective clusters.

Agreement Between Hierarchical Agglomerative and Iterative Procedures Assignment of Subjects to Clusters

Comparison of the hierarchical agglomerative and iterative clustering procedures, together with the rules for determining the number of clusters, pointed to a two cluster solution as being the most optimal representation of this data set. The level of agreement between the assignment of individuals to cluster groups by the hierarchical agglomerative solution and the assignment of individuals to the cluster groups by the iterative partitioning procedure was examined next. The classification matrix for this comparison appears in Table 53. From Table 53, the percentage agreement between the two sets identified, 96%, was significant [χ^2 (3, 110) = 268.9; $p < .001$]. This finding tends to support the internal validity of the obtained solutions.

Table 53

HAP vs. IPP Assignment of Subjects: SB:FE Subsample age 2-4-11

		Hierarchical Agglomerative Procedures		
		Group 1	Group 2	n
Iterative Procedures	Group 1	43	0	43
	Group 2	4	63	67
	n	47	63	110

Multivariate Profile Analysis

The groups derived through the iterative partitioning procedure were subjected to multivariate profile analysis. There was a significant multivariate main effect for profile elevation [$F(6, 105) = 76.16$, $p < .001$] when the two profiles were tested simultaneously. Univariate F tests for each SB:FE GPAB Reasoning Area SAS were conducted next. The results are presented in Table 54. There was also a significant multivariate main effect for the two group solution when the profile line segments were tested for parallelism [$F(3, 106) = 6.38$, $p < .001$]. The two groups were similar to each other in the shape of the portion of the profile defined by their Quantitative -

Short Term Memory and Abstract/Visual - Quantitative SAS segments, but different to each other in the shape of their profiles defined by the Verbal - Abstract/Visual SAS segment. The results are presented in Table 55.

Table 54

Differences Between SB:FE Reasoning Area Group Means for IPP Cluster Solution: SB:FE Subsample Age 2-4-11

	Iterative Cluster Means		
SB:FE Variable	1 (n=63)	2 (n=47)	Univariate F Ratio (2, 108)
Verbal SAS	119.0	108.4	18.6*
A/V SAS	116.9	96.9	81.7*
Quantitative SAS	118.9	95.1	116.4*
STM SAS	117.0	97.4	85.4*

Multivariate Main Effect for Elevation:
$F(4, 105) = 76.16, p < .001$

*denotes means significantly different at the $p < .01$ level.

Table 55

Tests for Profile Parallelism on IPP Clusters: Subsample Age 2-4-11

SB:FE Segment	F ratio	Significance
Verbal - A/V SAS	9.179	*
A/V - Quantitative SAS	1.606	NS
Quantitative - STM SAS	1.737	NS

Multivariate Main effect for Parallelism:
$F(3, 106) = 6.38, p < .001$

*$p < .05$ level; NS denotes non significant differences.

Summary of the Results of Chapter Six

In this part of the study it was hypothesized that application of hierarchical agglomerative and iterative partitioning clustering procedures to the SB:FE GPAB data would result in the identification of subgroups within the various age samples that displayed significant differences with respect to profile elevation and/or shape. The results obtained generally supported this hypothesis. For the age sample 2-4-11, a two cluster solution appeared optimal, with a very high scoring group and an average group being formed. For the remaining three age samples, 5-6-11, 7-11-11, and 12-23-11, a three cluster solution appeared optimal. The three cluster solutions generally comprised a high scoring group, an average group, and a low scoring group. The various rules used to determine the number of clusters in the data set and the two clustering procedures adopted generally yielded consistent findings. There was considerable agreement between hierarchical agglomerative and iterative procedures assignment of subjects to clusters across the four age groups considered and in all the subsamples analyzed. In addition, there was considerable similarity between results obtained from the application [illegible] clustering procedures to different samples of the same age level drawn from the overall data set.

Thus, the third objective of the study, as outlined in the first chapter, was achieved. Internally valid, reliable, and replicable groups that displayed differences in profile elevation and/or shape were obtained through application of multivariate clustering procedures to different samples drawn from the same population, providing evidence for the stability and generality of the solutions obtained. In the next part of the study, the external validity of these solutions will be explored. Particular emphasis, in the next part of the study, is placed on an examination of the clinical validity and utility of the solutions derived here, in addition to exploring, from an empirical perspective, the external validity of the cluster solutions obtained.

CHAPTER SEVEN

Part Four

Presented in this chapter are the results of the examination of the external validity of the three groups identified through application of the empirical iterative partitioning procedure to the SB:FE GPAB scores of the age groups considered. Evidence pertaining to the external validity of the cluster solutions was collected in two separate steps. The first step involved use of Multivariate Analysis of Variance (MANOVA) to test for mean differences among the three cluster groups in each age group on WRAT-R Reading, Spelling, and Arithmetic achievement subtests. The second step involved examining the agreement between the three subtypes derived using the iterative procedure and groups identified using clinical categorizations procedures based on intellectual classification and academic achievement criteria. Four a priori clinical categorization or inferential models, which were described in Chapter Three, were considered.

As pointed out in Chapter Three, the WRAT-R is not administered to children less than 5 years of age. Consequently, examination of the external validity of the three empirically derived subgroups was restricted to the subsamples at ages 5-6-11 (n=120), 7-11-11 (n=301), and 12-23-11 (n=186) for whom both SB:FE GPAB data and WRAT-R data were available.

External Validation Procedures With Subsamples Age 5-6-11, 7-11-11, and 12-23-11

Hypothesis 3:1

Subgroups within the age samples 5-6-11, 7-11, and 12-23-11 identified on the basis of iterative partitioning cluster analytic procedures applied to SB:FE data will demonstrate significant mean differences on the external criteria of WRAT-R achievement scores.

Presented in Table 56 are the WRAT-R means for the three groups identified using the iterative clustering procedure within each of the age samples, as well as the results of the statistical tests to identify significant differences between group means. As shown, within each age sample, the three cluster groups performed significantly differently on each of the three WRAT-R subtests. In each case, the above average group outperformed the average group who, in turn, outperformed the below average group. These results suggest that the groups determined on the basis of cognitive ability variables also differ significantly on variables of academic achievement. Hypothesis 3.1 was supported across the three age groups considered.

Table 56

WRAT-R Differences Between Iterative Partitioning Clusters for the Three Age Groups

Subsample Age 5-6-11	Means of Iterative Clusters			Univariate F tests (2, 117)			
WRAT-R Subtest	1 (n=12)	2 (n=80)	3 (n=28)	F ratio	1 vs 2	1 vs 3	2 vs 3
Reading	124.4	105.4	88.3	14.92	*	*	*
Spelling	123.3	102.4	85.6	21.61	*	*	*
Arithmetic	122.9	109.9	85.7	28.35	*	*	*

Multivariate Main Effect:
$F(6, 230) = 10.28$, $p < .001$

Subsample Age 7-11-11	Means of Iterative [illegible]			Univariate F tests (2, 298)			
WRAT-R Subtest	1 (n=97)	2 (n=155)	3 (n=49)	F ratio	1 vs 2	1 vs 3	2 vs 3
Reading	106.9	90.8	73.4	7.73	*	*	*
Spelling	101.6	86.9	72.6	60.59	*	*	*
Arithmetic	101.9	89.8	73.2	88.07	*	*	*

Multivariate Main Effect:
$F(6, 592) = 31.47$, $p < .001$

Subsample Age 12-23-11	Means of Iterative Clusters			Univariate F tests (2, 183)			
WRAT-R Subtest	1 (n=68)	2 (n=[illegible])	3 (n=42)	F ratio	1 vs 2	1 vs 3	2 vs 3
Reading	106.8	91.8	73.3	78.56	*	*	*
Spelling	101.4	87.4	73.3	49.68	*	*	*
Arithmetic	108.3	84.2	66.8	95.49	*	*	*

Multivariate Main Effect:
$F(6, 362) = 31.46$, $p < .001$

*denotes pairs of means significantly different at the $p < .01$ level.

Comparison Between Empirically and Clinically Derived Classifications

This part of the study examined the agreement between the empirical classification procedures and clinical classification models. In particular, what was examined was whether the categories within each of the a priori clinical classifications systems could be collapsed to produce three unique subgroupings which, when ordered by ability level, corresponded to the empirically ordered cluster groups. The agreement in classification between each of four a priori clinical inferential models identified in Chapter Three with the iterative partitioning procedure cluster solutions was examined within each of the age samples. The hypothesis designed to explore this agreement is presented below and the results are reported separately for each of the age groups following the hypothesis.

Hypothesis 3.2

There will be congruence between a priori clinically determined subgroups and the three subgroups empirically derived through application of iterative partitioning cluster analytic procedures within the 5-6-11, 7-11-11, and 12-23-11 subsamples, for whom both SB:FE data and WRAT-R data were available.

Cross-classifications for the Subsample Age 5-6-11

Model One: Intellectual classification groups. Under Model One, each subject within this age group was assigned to one of seven groups: Very Superior (n=7); Superior (n=19); High Average (n=39); Average (n=42); Low Average (n=11); Slow Learner (n=2); and Mentally Retarded (n=0) on the basis of SB:FE GPAB Composite SAS scores. Membership in these cluster groups was then crossed with membership in the empirically derived groups of above average (n=12), average (n=80), and below average (n=28). The results are displayed in Table 57. As shown, the 12 students placed in the above average group using the iterative clustering procedure were identified as "very superior" (n=7) or "superior" (n=5) using the clinical intellectual classification system. This would seem to indicate that by collapsing these two intellectually determined groups, the above average clinical group would be formed. However, this is clearly not the case. Examination of those students classified as "superior", shows that while five were in the first empirical group, 14 were placed in the average empirical group. Further, of the 42 students classified as "average" in the clinical Model One, 27 were placed in empirical group 2 and 15 in empirical group 3. Of the 11 students classified as Low Average, all fell in the empirical low average group, as did all the subjects classified as "slow learners" by the clinical classification procedure. Consequently, it was concluded that the categories within the intellectual classification model could not be collapsed to form groups which uniquely corresponded to one of the three groups derived using the empirical iterative partitioning procedure.

Table 57

3 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 5-6-11: Model One

		Three Cluster Solution		
Group	n	1	2	3
VS	7	7	0	0
S	19	5	14	0
HA	39	0	39	0
A	42	0	27	15
LA	11	0	0	11
SL	2	0	0	2
MR	0	0	0	0
	120	12	80	28

Note: VS=Very Superior; S=Superior; HA=High Average; A=Average; LA=Low Average; SL=Slow Learner; MR=Mentally Retarded.

Given these results, it was speculated that perhaps greater agreement between the intellectual categorizations and empirical categorizations could be realized by increasing the number of empirical groups. That is, could an empirical solution be fitted to the seven group intellectual classification system? Inspection of the cross-tabulations of the clinical intellectual groups with the k= 4, 5, 6, and 7 empirical groups presented in Table 58 reveals that there is no gain. Indeed in all the cases, a similar distribution of empirical group membership with intellectual levels are found suggesting very little differences between empirical groups (e.g., empirical groups 2 and 3 for k=4 group solution; 4 and 5 for the k=5 solution and 4 and 5 for the k=7 solution). Clearly, compared to the k=3 empirical solution, no improvement is to be gained by increasing the number of empirical groups to effect greater agreement with the clinical classification system. And while the three group empirical solution appeared to fit "best", there was overlap between "superior" and "average" levels in the intellectual classification system and one of the three empirical groups, preventing a unique collapsing as postulated. Hypothesis 3.2 was not supported by the findings derived from using this clinical classification system. Further, this finding foreshadowed the likelihood that there would be no better fits observed when using the remaining clinical systems.

Table 58

4-7 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 5-6-11: Model One

		Four Cluster Solution				Six Cluster Solution					
Group	n	1	2	3	4	1	2	3	4	5	6
VS	7	7	0	0	0	5	0	2	0	0	0
S	19	1	7	11	0	0	11	8	0	0	0
HA	39	0	22	17	0	0	18	9	9	3	0
A	42	0	17	11	14	0	4	0	15	16	7
LA	11	0	0	0	11	0	0	0	0	0	11
SL	2	0	0	0	2	0	0	0	0	0	2
MR	0	0	0	0	0	0	0	0	0	0	0
	120	8	46	39	27	5	33	19	24	19	20

		Five Cluster Solution					Seven Cluster Solution						
Group	n	1	2	3	4	5	1	2	3	4	5	6	7
VS	7	5	2	0	0	0	5	0	2	0	0	0	0
S	19	0	13	6	0	0	0	11	8	0	0	0	0
HA	39	0	12	25	2	0	0	18	9	9	3	0	0
A	42	0	0	13	20	9	0	4	0	13	16	3	6
LA	11	0	0	0	0	11	0	0	0	0	0	4	7
SL	2	0	0	0	0	2	0	0	0	0	0	0	2
MR	0	0	0	0	0	0	0	0	0	0	0	0	0
	120	5	27	44	22	22	5	33	19	22	19	7	15

Note: VS=Very Superior; S=Superior; HA=High Average; A=Average; LA=Low Average; SL=Slow Learner; MR=Mentally Retarded.

Model Two: Normal learners compared with homogeneous Learning Disabled groups. With this clinical model, subjects were assigned to one of seven a priori groups: Gifted (Very Superior and Superior) (n=26); Normal Learner (n=76); Slow Learner (n=2); Mentally Retarded (n=0); and one homogeneous LD group (n=16). As shown in Table 59, no unique collapsing of the five groups to form three groups which coincided closely with the empirical three group solution can be found. The subjects classified as "gifted" are fairly evenly divided between the first two empirical groups; likewise, although not evenly, the students classified as either "normal learner" or as "learning disabled" are present in empirical groups 2 and 3.

Table 59

Cluster Iterative Partitioning and a priori Groups for the Subsample Age 5-6-11: Model Two

		Three Cluster Solution		
Group	n	1	2	3
VS/S	26	12	14	0
NL	76	0	60	16
SL	2	0	0	2
MR	0	0	0	0
LD	16	0	6	10
	120	12	80	28

Note: VS/S=Very Superior/Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; LD=Learning Disabled.

Again, the question arises whether the agreement between the empirically derived solution and the homogeneous learning disability classification model would improve if other than an empirical three cluster solution was used. Generally, like the clinical intellectual model, the cross-tabulations revealed the alternate solutions (k=4, 5, 6, and 7) were less satisfactory. The solutions are displayed in Appendix A. The findings with the homogeneous learning disabled model fail to support hypothesis 3.2.

Model Three: Normal learners compared with heterogeneous Learning Disabled groups. In this model, the 16 Learning Disabled subjects in the previous model were divided into seven homogeneous subgroups according to patterns of academic performance. In addition to Gifted (Very Superior and Superior) (n=26), Normal Learner (n=76), Slow Learner (n=2), and Mentally Retarded (n=0), the following groups of LD subjects were created: Spelling Disabled (n=5); Arithmetic Disabled (n=3); Reading Disabled (n=4); Reading and Spelling Disabled (n=1); Reading and Arithmetic Disabled (n=1); Spelling and Arithmetic Disabled (n=1), and Reading, Spelling, and Arithmetic Disabled (n=1). As might be expected, the results of the cross-classification of the groups identified with this model with the three empirically derived groups and shown in Table 60 are similar to the findings obtained with Model Two. Again, the 26 "very superior/superior" subjects were essentially evenly divided between empirical groups 1 and 2, the "normal learner groups" were split in a ratio of approximately 4 to 1 between the empirical groups 2 and 3, and the learning disabled subjects were found in groups 2 and 3. Attempts to improve the fit between the clinical model and a greater number of empirical groups led to results similar to those found for the previous two models (see Appendix A). Hypothesis 3.2 did not receive support.

Table 60

3 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 5-6-11: Model Three

		Three Cluster Solution		
Group	n	1	2	3
VS/S	26	12	14	0
NL	76	0	60	16
SL	2	0	0	2
MR	0	0	0	0
SLD	5	0	2	3
ALD	3	0	0	3
RLD	4	0	3	1
R/SLD	1	0	1	0
R/ALD	1	0	0	1
S/ALD	1	0	0	1
R/S/ALD	1	0	0	1
	120	12	80	28

Note: VS/S=Very Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; SLD=Spelling Disabled; ALD=Arithmetic Disabled; RLD=Reading Disabled; R/SLD=Reading/Spelling Disabled; R/ALD=Reading/Arithmetic Disabled; S/ALD=Spelling/Arithmetic Disabled; R/S/ALD=Reading/Spelling/Arithmetic Disabled.

Model Four: Normal Learners compared with Learning Disabled subjects based on Rourke's classification system. Based on Rourke and colleagues classification schema, Gifted (Very Superior/Superior) (n=26); Normal Learner (n=82); Slow Learner (n=2); and Mentally Retarded (n=0); and three Learning Disabled subgroups were identified: Reading LD with Arithmetic normal (n=5); Reading and Arithmetic LD (n=2); and Arithmetic LD with Reading normal (n=3). As shown in Table 61, when crossed with the three empirical groups, the results are somewhat similar to the findings obtained with the homogeneous LD group in Model Two and the heterogeneous LD group in Model Three. "Very superior/superior" and "normal learner" subjects were found, respectively, in empirical groups 1 and 2 and 2 and 3. However, with one exception, RLD/AN, the subjects in the Rourke LD groups were found in the third group.

Table 61

3 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 5-6-11: Model Four

		Three Cluster Solution		
Group	n	1	2	3
VS/S	26	12	14	0
NL	82	0	62	20
SL	2	0	0	2
MR	0	0	0	0
RLD	0	0	0	0
RLD/AN	5	0	4	1
R/ALD	2	0	0	2
ALD/RN	3	0	0	3
	120	12	80	28

Note: VS/S=Very Superior/Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; RLD=Reading Disabled; RLD/AN=Reading LD with Math Normal: R/ALD=Reading/Arithmetic Disabled; ALD/RN=Arithmetic Disabled/Reading Normal.

When alternate numbers of empirically driven solutions were considered (k=4, 5, 6, and 7) the cross-tabulations again revealed no improvement overall (see Appendix A). While there was some improvement in the fit between Rourke's classification and the three group empirical solution, together the results suggest that Hypothesis 3.2 is again rejected.

Thus, it would appear that the speculation that it would be possible to form unique groups within each of the 4 clinical models by collapsing categories which would fit uniquely with the three group empirical solution was not supported. There seemed to be limited agreement between the clinical models and the empirical solutions even when alternate values for k were considered. Group membership within the empirically derived clusters, although determined by profile similarity, appeared to be related more to overall SB:FE GPAB profile elevation (i.e., Model One, intellectual classification) than to diagnostic composition (i.e., Models Two, Three, and Four). LD subjects tended to fall within the lower average clusters; however, no empirically derived groups existed for which there was unique agreement with any of the clinically derived learning disabled groups. What remains to be seen is whether the findings for subjects in the 5-6-11 age sample hold for the two older age groups.

Cross-classifications for the Subsample Age 7-11-11

Model One: Intellectual classification groups. Under Model One, each subject within this age group was assigned to one of seven groups: Very Superior (*n*=9); Superior (*n*=23); High Average (*n*=46); Average (*n*=168); Low Average (*n*=33); Slow Learner (*n*=15); and Mentally Retarded (*n*=7) based on SB:FE GPAB Composite SAS scores. Membership in these cluster groups was crossed with membership in the empirically derived groups of above average (*n*=97), average (*n*=155), and low average (*n*=49). As shown in Table 62, the 78 students placed in the "very superior", "superior", and "high average" groups were all found in the above average empirical group. Similarly, there was a clear cross-classification for the "slow learner" and "mentally retarded" subjects; all were located in the below average empirical group. The same clear cross-classifications just mentioned did not occur for the subjects identified as "average" and "low average". Nineteen of the "average" st[illegible] found in the above average empirical group, while 149 were placed in empirical [illegible] group. Likewise, 6 "low average" subjects were placed in the empirical [illegible] in comparison to the results observed earlier for the sample ag[illegible] over" is proportionately smaller. If one accepts that the 19 "aver[illegible] ts are "misclassified", then seemingly the intellectual categories [illegible] groups which coincide with the three empirical groups.

Again, it was speculated that perhaps greater agreement between the intellectual and empirical classifications could be realized by increasing the number of empirical groups. That is, by increasing the number of empirical groups, the number of overlapping classifications would be reduced. The cross-tabulations, however, which appear in Appendix A, revealed the alternate solutions were clearly less satisfactory. Thus, it appears that within this age group, the three empirically derived groups, above average, average, and below average, although not corresponding exactly to the combination of unique categories within the intellectual classification system, provided a reasonable fit to the clinical classification system, and certainly provided the best fit of the empirical models considered. Hypothesis 3.2 received modest support from the clinical classification model based on overall intellectual ability.

Table 62

3 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 7-11-11: Model One

		Three Cluster Solution		
Group	n	1	2	3
VS	9	9	0	0
S	23	23	0	0
HA	46	46	0	0
A	168	19	149	0
LA	33	0	6	27
SL	15	0	0	15
MR	7	0	0	7
	301	97	155	49

Note: VS=Very Superior; S=Superior; HA=High Average; A=Average; LA=Low Average; SL=Slow Learner; MR=Mentally Retarded.

Model Two: Normal learners compared with homogeneous Learning Disabled groups. Within this clinical model, subjects were assigned to one of seven a priori determined groups: Gifted (Very Superior and Superior) (n=28); Normal Learner (n=168); Slow Learner (n=15); Mentally Retarded (n=7); and one homogeneous LD group (n=83). As shown in Table 63, no unique collapsing of the five groups to form three groups which coincided closely with the empirical group solution can be found. When alternate solutions for the empirical solution were considered, without exception, the cross-tabulations revealed the alternate solutions were less satisfactory: The degree of multiple overlap increased as k increased, particularly within the "normal learner" group. The solutions are displayed in Appendix A. The findings with Model Two fail to support Hypothesis 3.2.

Table 63

3 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 7-11-11: Model Two

		Three Cluster Solution		
Group	n	1	2	3
VS/S	28	28	0	0
NL	168	48	102	18
SL	15	0	0	15
MR	7	0	0	7
LD	83	21	53	9
	301	97	155	49

Note: VS/S=Very Superior/Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; LD=Learning Disabled.

Model Three: Normal learners compared with heterogeneous Learning Disabled groups. In this model, the 83 Learning Disabled subjects were divided into seven homogeneous LD subgroups according to patterns of academic performance. In addition to Gifted (Very Superior and Superior) (n=28), Normal Learner (n=168), Slow Learner (n=15), and Mentally Retarded groups (n=7), the following groups of LD subjects are created: Spelling Disabled (n=15); Arithmetic Disabled (n=14); Reading Disabled (n=7); Reading and Spelling Disabled (n=22); Reading and Arithmetic Disabled (n=4); Spelling and Arithmetic Disabled (n=9); and Reading, Spelling, and Arithmetic Disabled (n=12). As might be expected, the results of the cross-classification of the groups identified with this model with the three empirically derived three cluster solution and shown in Table 64 provided a less than ideal fit to the data set, not only for the aforementioned "normal learner" group, but also for the specific learning disability subgroups where subjects so classified were found in all three empirical groups. Results of the cross-classification of the groups identified with this model are similar to those obtained with the previous model. Next, alternate values for k were considered (4, 5, 6, and 7); however, the cross-tabulations, presented in Appendix A, revealed the alternate solutions were less satisfactory. Hypothesis 3.2 received no support.

Table 64

3 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 7-11-11: Model Three

		Three Cluster Solution		
Group	n	1	2	3
VS/S	28	28	0	0
NL	168	48	102	18
SL	15	0	0	15
MR	7	0	0	7
SLD	15	8	6	1
ALD	14	3	10	1
RLD	7	2	5	0
R/SLD	22	3	17	2
R/ALD	4	0	2	2
S/ALD	9	3	5	1
R/S/ALD	12	2	8	2
	301	97	55	49

Note: VS/S=Very Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; SLD=Spelling Disabled; ALD=Arithmetic Disabled; RLD=Reading Disabled; R/SLD=Reading/Spelling Disabled; R/ALD=Reading/Arithmetic Disabled; S/ALD=Spelling/Arithmetic Disabled; R/S/ALD=Reading/Spelling/Arithmetic Disabled.

Model Four: Normal Learners compared with Learning Disabled subjects based on Rourke's classification system. Gifted (Very Superior and Superior) (n=30), Normal Learner (n=187), Slow Learner (n=15), and Mentally Retarded groups (n=7), and four LD subgroups were identified: Reading LD with Arithmetic normal (n=24); Reading and Arithmetic LD (n=16); Arithmetic LD with Reading normal (n=14); and a Reading LD group (n=8) that did not met criteria for membership in the first three groups. As shown in Table 65, results are similar to the findings obtained with the homogeneous LD group in Model Two and the heterogeneous LD subgroups in Model Three. The three cluster solution provided a fairly poor fit to the clinical model. However, when alternate numbers of empirically driven solutions were considered (4, 5, 6, and 7) the cross-tabulations revealed the alternate solutions were even less satisfactory (see Appendix A).

Table 65

3 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 7-11-11: Model Four

		Three Cluster Solution		
Group	n	1	2	3
VS/S	30	30	0	0
NL	187	57	108	22
SL	15	0	0	15
MR	7	0	0	7
RLD	8	2	6	0
RLD/AN	24	3	19	2
R/ALD	16	2	11	3
ALD/RN	14	3	11	0
	301	97	155	49

Note: VS/S=Very Superior/Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; RLD=Reading Disabled; RLD/AN=Reading LD with Math Normal; R/ALD=Reading/Arithmetic Disabled; ALD/RN=Arithmetic Disabled/Reading Normal.

Cross-classifications for the Subsample Age 12-23-11

Results obtained in the sample age 12-23-11 are generally consistent with those obtained with the two younger samples. The empirical three cluster solution comprised an above average group of 68 subjects, an average group of 76 subjects, and a lower average group of 42 subjects and generally provided less than perfect fits to the various clinical classification models. As shown in Table 66, the 68 subjects placed in the above average group using the empirical procedure were identified as "very superior" (n=2), "superior" (n=14), "high average" (n=28) and "average" (n=24). Examination of those students classified as "average" by the clinical classification system, shows that while 24 were in the first empirical group, 54 were in the second group. Further, of the students classified as "low average" by the clinical model, 22 were placed in the empirical group 2, an average group, and seven in empirical group 3, a low average group. The results of the remaining cross-tabulations between the three cluster empirical solutions and the clinical Models Two through Four are presented together in Table 66.

Table 66

3 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 12-23-11; Models Oneto Four

Model One Group	n	Three Cluster Solution 1	2	3
VS	2	2	0	0
S	14	14	0	0
HA	28	28	0	0
A	78	24	54	0
LA	29	0	22	7
SL	19	0	0	19
MR	16	0	0	16
	186	68	76	42

Model Three Group	n	Three Cluster Solution 1	2	3
VS/S	16	16	0	0
NL	99	42	50	7
SL	19	0	0	19
MR	16	0	0	16
SLD	8	4	4	0
ALD	10	2	8	0
RLD	2	1	1	0
R/SLD	4	0	4	0
R/ALD	0	0	0	0
S/ALD	4	2	2	0
R/S/ALD	8	1	7	0
	186	68	76	42

Model Two Group	n	Three Cluster Solution 1	2	3
VS/S	16	16	0	0
NL	99	42	50	7
SL	19	0	0	19
MR	16	0	0	16
LD	36	10	26	0
	186	68	76	42

Model Four Group	n	Three Cluster Solution 1	2	3
VS/S	16	16	0	0
NL	113	48	58	7
SL	19	0	0	19
MR	16	0	0	16
RLD	4	0	4	0
RLD/AN	2	1	1	0
R/ALD	6	1	5	0
ALD/RN	10	2	8	0
	186	68	76	42

Note Model One: VS=Very Superior; S=Superior; HA=High Average; A=Average; LA=Low Average; SL=Slow Learner; MR=Mentally Retarded.
Note Model Two: VS/S=Very Superior/Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; LD=Learning Disabled.
Note Model Three: VS/S=Very Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; SLD=Spelling Disabled; ALD=Arithmetic Disabled; RLD=Reading Disabled; R/SLD=Reading/Spelling Disabled; R/ALD=Reading/Arithmetic Disabled; S/ALD=Spelling/Arithmetic Disabled; R/S/ALD=Reading/Spelling/Arithmetic Disabled.
Note Model Four: VS/S=Very Superior/Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; RLD=Reading Disabled; RLD/AN=Reading LD with Math Normal; R/ALD=Reading/Arithmetic Disabled; ALD/RN=Arithmetic Disabled/Reading Normal.

Of the four clinical classification models examined, the three cluster empirical solution fit best with the clinical classification model based on the intellectual categories, Model One. When alternate values for k (4, 5, 6, and 7) were considered for each of the four clinical classification models, the results of the cross-tabulations revealed essentially no gain whatsoever. These results are presented in Appendix A. In sum, Hypothesis 3.2 was not supported by the findings derived from using the four clinical classification systems within this age group.

As a final means of exploring the differences in profile between "normal learners" and "learning disabled" subjects, two groups from the homogeneous learning disability classification Model Two were compared across each of the three age groups. These are two groups with essentially the same SB:FE score ranges, but separated on the basis of WRAT-R scores. The results of tests of the differences in means is presented in Table 67. As shown, of the 15 comparisons made on SB:FE variables, there was only one statistically significant difference ($p < .05$) in the SB:FE Verbal Reasoning Area scores within the sample age 5-6-11. The learning disabled group within this age range appeared to have consistently lower SB:FE Reasoning Area and Composite SAS scores and the absence of statistical significance across the remaining SB:FE variables likely reflects the small and unequal sample sizes in the two groups being compared. Examination of the results generally suggest that the LD subjects in the younger age group typically evidenced lower overall SB:FE scores than the "normal learners". This may be an artifact of the sample composition within this age group. However, as the ages within the sample increase, for example in the subsamples at age 7-11-11 and 12-23-11, the differences between the LD groups and the "normal learner" groups are clearly not statistically significant, and in addition, do not appear to be clinically different, suggesting that in the older two age groups the SB:FE Reasoning Area and Composite SAS's may contribute little to the differential diagnosis of homogeneous learning disabled subjects compared to normal achievers. However, as expected the differences within each age sample on each of the three WRAT-R subtests are all clinically and statistically different, suggesting that there were meaningful differences between the normal and learning disabled subjects with respect to academic functioning. In each case, the normal learne group outperformed the learning disabled group. However, these differences in academic ability, by and large, did not translate into differential performance on cognitive ability variables.

Table 67

<u>Comparison of Mean SB:FE GPAB and WRAT-R Scores of the Normal Learner and LD Groups for the Three Age Samples</u>

	n		Ver. SAS	A/V SAS	Qnt. SAS	STM SAS	Com. SAS	WRAT RD	WRAT SP	WRAT ARITH
Age 5-6-11										
Normal	76	M	109.3	106.7	107.7	101.2	107.3	100.3	102.0	105.8
		SD	12.1	14.9	12.7	12.7	10.5	14.62	17.8	15.8
LD	16	M	98.2*	94.8	104.9	92.8	97.2	77.4**	80.9**	85.5**
		SD	12.2	18.7	15.4	15.9	12.5	14.5	13.3	26.6
Age 7-11-11										
Normal	168	M	104.1	103.3	98.8	96.5	100.8	100.4	96.0	95.7
		SD	13.2	13.6	11.5	13.3	10.4	14.32	14.2	11.6
LD	83	M	100.9	104.1	96.7	96.8	99.6	78.5**	75.3**	82.5**
		SD	13.6	14.2	9.6	15.1	10.9	14.6	10.2	12.3
Age 12-23-11										
Normal	99	M	103.8	101.9	100.2	91.5	96.6	100.4	95.7	95.8
		SD	13.5	12.9	11.8	15.9	13.8	13.2	13.1	14.9
LD	36	M	98.8	105.7	97.9	91.9	95.9	83.6**	75.4**	78.4**
		SD	13.1	12.4	10.8	16.2	13.9	12.3	10.3	15.8

*denotes pairs of means significantly different at $p < .05$ level; **denotes pairs of means significantly different at $p < .01$ level.

Given the small and uneven sample sizes within the various heterogeneous learning disability models, Models Three and Four, it was not possible to compare each learning disabled subgroups performance on the SB:FE and WRAT-R variables.

Summary of the Results of Chapter Seven

In the first part of this chapter evidence supporting the external validity of the three cluster solutions derived through application of multivariate iterative partitioning procedures to SB:FE GPAB data in the previous chapter was found. Within each age sample, the three cluster groups performed significantly differently on each of the three WRAT-R subtests, Reading, Spelling, and Arithmetic. These results suggest that the groups determined on the basis of cognitive ability variables also differ significantly on variables of academic achievement.

In the second step of the external validation procedure, the empirically derived cluster solutions were compared with clinically derived a priori groups based on intellectual classifications using SB:FE GPAB Composite SAS scores and learning disability models based on patterns of cognitive ability and academic achievement. Overall, the results revealed that collapsing categories identified using the intellectual model, Model One, yielded a somewhat closer fit with the three empirically determined groups than was observed for each of Models Two, Three, and Four, which were based on cognitive ability and patterns of academic achievement. However, the degree of agreement was less than ideal. Further, when alternate numbers of groups were considered (k=4, 5, 6, and 7) less satisfactory solutions were observed. The influence of overall ability levels or SB:FE GPAB Composite SAS appeared to be significant and pervasive, even when various homogeneous and heterogeneous learning disability subtyping models were considered. To reiterate, when the various models were considered, there appeared to be no empirical clusters that were more clearly related to psychoeducational diagnosis than to SB:FE overall or Composite SAS scores, i.e., profile elevation. The ability of SB:FE Reasoning Area scores to distinguish normal subjects from learning disabled subjects is thus questionable. LD subjects did not demonstrate profiles that were significantly different from profiles of "normal learners", either when empirical groups were compared with a priori clinically derived classifications or when the two groups were directly compared. This latter point was clearly illustrated when "normal learners" and "learning disabled" groups with essentially the same Composite SAS ranges were compared on SB:FE GPAB and WRAT-R variables in the three age samples considered. Generally, no differences in group means were found in SB:FE Reasoning Area and Composite scores between the two groups selected for comparison. One exception in the age sample 5-6-11 is noted in the SB:FE GPAB Verbal Reasoning Area and within this age group, caution may be needed in interpreting findings as sample sizes are small. Most noticeably, within this age group, practical differences in mean scores were present, with the learning disabled group typically scoring below the normal achieving comparison group. Differences in performance on the SB:FE Reasoning Area and Composite SAS's in the two older samples were nonsignificant, from both a statistical and practical perspective. However, as expected, all academic achievement scores were significantly lower in the learning disabled group within each age level.

Consideration of alternate values for k in comparisons involving the empirical solutions in this part of the study was not intended to detract from the statistical validity of the findings documented in Part Three of this study. However, users of test instruments such as the SB:FE for clinical and research purposes are not only interested in distinguishing between subjects who may score within the high or low ranges on intellectual measures. As noted in the review of the literature presented in Chapter Two, a great deal of interest and research has focused on the identification of particular cognitive profiles that distinguish normal from learning disabled subjects who may typically score within the same overall cognitive ability ranges. Thus, the question of interest shifts from comparisons based on overall ability levels, to somewhat finer and more subtle comparisons between subjects with overall ability levels in the same range. To fully address this question, and to explore the clinical utility of the SB:FE within this context, alternate values for k were therefore considered. The results obtained in this analysis cast doubt about the validity of profiles derived from SB:FE Reasoning Area scores in distinguishing normal learners from learning disabled subjects. None of the empirical clusters for the alternate values for k corresponded uniquely with either heterogeneous or homogeneous LD subgroups and the empirically derived clusters appeared to be more similar with respect to overall cognitive ability than with respect to the psychoeducational categories considered.

CHAPTER EIGHT: DISCUSSION

Overview

This chapter contains three sections. First, a summary of the study including its purpose, procedure, and findings is presented. This is followed by a discussion of the conclusions of the study. Finally, the chapter ends with implications for school psychology practice and research.

Summary

Purpose and Problem

The SB:FE was published in 1986 (Thorndike et [illegible] a) with a number of important changes in development and organization from its pre[illegible]essors. The theoretical model of intelligence underlying the test was considerably expanded. The traditional mental age format and IQ scores were replaced with Standard Age Scores (SAS's) reported for the four Reasoning Areas (Verbal, Abstract/Visual, Quantitative, and Short Term Memory) and the Composite index. As well, standard scores were used for the 15 subtests which comprise the total test. Further, an adaptive format by which the test administrator can vary, within limits, the subtests administered was adopted to reduce administration time required. Flexibility in interpretation and administration is supported by the option to calculate Reasoning Area and Composite SAS's based on numerous combinations of subtests (Glutting, 1989). Each of the four Reasoning Area scores and therefore Composite Standard Age Scores (SAS's) may be composed of various numbers and different subtests.

The adaptive format of the SB:FE poses particular problems in clinical use (Sattler, 1988, 1992) and in multivariate research (Fritzke, 1988; Glutting, 1989; Spelliscy, 1991). To help deal with difficulties in empirically studying validity issues brought about by the variations presented in administration, Thorndike et al. (1986b) used the following age groupings: 2 through 6 years; 7 through 11 years; and 12 through 18-23 years. Abbreviated batteries, in addition, are available that reduce test administration time and overcome the difficulties posed by the adaptive format of the SB:FE full version. One such abbreviation is the six test General Purpose Abbreviated Battery (GPAB) which includes six subtests that are to be administered to all ages: Vocabulary;

Bead Memory; Quantitative; Memory For Sentences; Pattern Analysis; and Comprehension. This battery may, according to the test developers, be used for placement decisions (Thorndike et al., 1986a). The shortened version requires less testing time than the complete battery and possesses acceptable internal consistency reliabilities (Glutting, 1989; Thorndike et al., 1986b).

The SB:FE was intended to assist clinicians and educators in formulating educational diagnoses and interventions. One of the most important areas within education today is that of learning disabilities (Hooper & Willis, 1989; Winzer, 1993). Individuals with learning disabilities experience extreme difficulties in making academic progress (Childs & Finucci, 1983; Hooper & Willis, 1989; Hynd, 1988; Rourke, 1991; Wilson, 1985, Winzer, 1993). Deficits typically persist into adulthood (Kaste, 1971; Mendelson, Johnson, & Stewart, 1971; Silver & Hagin, 1964; Spreen, 1982). The use of intelligence tests is central to work in the area of learning disabilities (Francis et al., 1991) and currently much effort is being spent determining subtypes according to patterns of underlying cognitive dysfunction (Forness, 1990; Hooper & Willis, 1989).

The role of the SB:FE GPAB in the diagnosis of learning disabilities, and its potential contribution to learning disability subtyping research, was therefore investigated in order to attempt to improve current diagnostic procedures and ultimately facilitate better remedial programming. In the current literature relating to guidelines and standards for educational testing (APA, 1985; CPA, 1987; Principles for Fair Student Assessment Practices for Education in Canada, 1993) the need to demonstrate the validity of test instruments for the purposes for which they are intended is emphasized. The intent of this study was to add to the growing body of empirical data providing information about the validity of the SB:FE and SB:FE GPAB, particularly with respect to educational diagnosis, profile analysis, and placement decisions.

In particular, the research was designed to meet four main objectives:

1. To provide comprehensive descriptive data about the SB:FE used with a Canadian clinic sample;

2. To explore the relationship between the SB:FE full battery and the SB:FE GPAB within this sample;

3. Given close agreement between the two, to then investigate the applicability of multivariate cluster analytic procedures to SB:FE GPAB data in order to derive reliable and replicable (internally valid) groups of individuals with distinct cognitive profiles; and

4. To explore the external validity of groups derived through application of multivariate clustering procedures to SB:FE GPAB data through investigation of subgroup differences on the basis of external achievement criteria and to explore the agreement between empirically derived subtypes and clinical inferential models.

The sample comprised 1220 individuals tested at the University of Alberta, Faculty of Education Clinic between 1986 to 1993. The sample of psychoeducational assessments (N=1220) included: SB:FE data; comprehensive demographic data for subjects age 2 years, 0 months through 23 years, 11 months and WRAT-R achievement data; for 607 of the subjects age 5-00 and older.

Findings

For convenience, the findings are separated into four main parts corresponding to the research objectives. In the first part, comprehensive descriptive data about the SB:FE used with a Canadian clinic sample was reviewed. The ratio of males to female in the predominantly Caucasian sample was approximately 3 to 2. Subtest and Reasoning Area and Composite SAS means generally decreased as the sample age increased and for the sample age 2-6-11 were higher than the standardization values (Thorndike et al., 1986b). However, for the age sample 12-23-11, the respective scores were slightly below standardization means (Thorndike et al., 1986b). Intercorrelations among subtest and Reasoning Area and Composite SAS's were consistent across the age group sampled and appear similar to those reported by Thorndike et al. (1986b).

Results were supportive of the four cognitive areas underlying the theoretical construct of the test posited by Thorndike et al. (1986b). Evidence for the SB:FE as a measure of *g* was also found. The significant intercorrelations among subtests and Reasoning Area and Composite SAS's suggest that the SB:FE is a valid measure of *g* and that the Composite SAS's can be interpreted with relative confidence.

In the second part the relationship between the SB:FE full battery and the SB:FE GPAB was explored. Within each age sample, differences between means and variances computed using the full battery and the reduced battery were generally significant ($p < .05$). A similar result was obtained for the pairs of variances. However, this needs to be tempered as the statistical differences found may have been more a function of large samples (Glass & Hopkins, 1984). The differences are of less consequence from a practical standpoint, as well as from a research perspective: e.g., differences in mean scores of approximately half a score point were found to be statistically significant.

With one exception, correlations between SB:FE full battery and GPAB Composite and Reasoning Area scores were close to unity. It is likely that the discrepancies reflect attenuation due to unreliability; lack of available internal consistency estimates precluded calculation of disattenuated correlation coefficients. The one exception to the strong relationships was the Abstract/Visual Reasoning Area where the correlations varied between .85 to .87 within the three age levels examined. This lower correlation between the full battery and the abbreviated version may reflect the lower reliability of the Copying subtest (Mason, 1992), as well as very stringent scoring criteria that may be difficult to operationalize uniformly. In addition, the Copying subtest comprises two essentially different sections. In the first section (items 1 to 12), the subject reproduces a block design built by the examiner. In the second section, the subject copies printed line drawings directly in the SB:FE record booklet and these drawings are then scored by the examiner according to explicit criteria. Difficulties in scoring may arise in particular in this latter section. These factors help explain the lower correlations observed within this area. Finally, significant and similar correlations were found for both the SB:FE full battery and the GPAB with an external criteria, the WRAT-R. Thus, from a practical view, the results suggested that the SB:FE GPAB represents the SB:FE full battery, particularly with respect to the four Area and Composite SAS's. Consequently, this version of the test was used in subsequent analyses.

In the third part of the study, multivariate cluster analytic techniques were applied to SB:FE GPAB data in order to derive reliable and replicable clusters at each age group. Two subsamples were considered: The first subsample comprised subjects for whom both SB:FE and WRAT-R data were available at the following age levels, 5-6-11, 7-11-11, and 12-23-11; the second subsample comprised subjects for whom only SB:FE data were available at the following age levels, 2-4-11, 5-6-11, 7-11-11, and 12-23-11. Hierarchical agglomerative (SPSSx, 1990) and k-means iterative partitioning (Dixon, 1981) clustering procedures were used to identify three groups - - low average, average, and above average, within all of the subsamples except the age group 2-4-11, where two groups were identified. The cluster solutions were then subjected to multivariate profile analysis in order to highlight the differences between the groups. Similar patterns were represented in the two older age groups 7-11-11 and 12-23-11, although the overall group means of each cluster in these age groups tended to be lower than those obtained in the age sample 5-6-11. It is likely that the emergence of a two cluster solution at the youngest age range reflects the sample composition and the greater frequency of referrals related to giftedness and parental interest within this age group, as well as the limited floor of the SB:FE at the lower age ranges.

In the fourth part of the study, the external validity of the cluster solutions was examined. First, significant differences were found between all three cluster groups on Reading, Spelling, and Arithmetic scores ($p < .01$); given the positive correlation between the SB:FE GPAB and WRAT-R, this is perhaps not surprising. While this provided some evidence of the external validity of the three cluster solutions, these findings were contradicted by the observation that there was poor agreement between the classification of subjects using empirical clustering procedures and the classification of subjects using one of four clinically derived a priori models based on cognitive ability and patterns of academic achievement. Cross classification tables were used to examine whether clinically derived a priori identified procedures could be uniquely combined to reproduce the empirically derived groups. The intellectual classification model which emphasized profile elevation appeared to provide the best fit for each of the three age samples with the empirical clustering solutions. Both homogeneous and heterogeneous learning disability models were essentially unrelated to the cluster solutions across the three age groups, 5-6-11, 7-11, and 2-23-11, even when alternate numbers of groups within the empirical cluster solutions were considered in an attempt to increase the clarity of the joint classifications. The number of clusters derived through the empirical procedures was varied from k=3 to 7. However, the alternate solutions provided essentially no gain in fit over the 3 cluster model. Results of the external validation procedures suggested that overall profile elevation contributed most to between group differences.

Limitations

Interpretations of the findings of the study must be viewed in terms of the following limitations. First, the idiosyncratic nature of the sample needs to be considered. The study involved post hoc analyses of data abstracted from psychoeducational assessment protocols and the sample comprised subjects referred to the University of Alberta Education Clinic which, as a group, represented the full range of cognitive ability and academic achievement levels. However, the sample cannot be considered representative of the general population. As such, direct comparisons with the norm sample results (Thorndike et al., 1986) should not be made. Further, the demographic data for the subjects in the clinic were based on information abstracted from psychoeducational assessment data. There may be some overlap among the classifications used in the original sources. For example, distinctions between "parental interest" and "giftedness" and "learning problems" and "school placement" may be poorly differentiated. As well, reasons for referral differed significantly across the age ranges sampled. This must be considered when making comparisons among the age groups considered. Second, lack of available reliability data prevented correcting the correlations for attenuation due to unreliability. Consequently, uncorrected correlations were used when examining convergence among tests scores which were hypothesized to converge, and these likely provide lower bound estimates. Third, given the number of subjects with missing data, and the absence of reliability data, particularly at the subtest level, clustering procedures were restricted to using Reasoning Area scores based on the SB:FE GPAB. Inferences, therefore, cannot be drawn about the subtests. Finally, in the part of the study in which the external validity of the empirically derived cluster solutions is addressed, the latter two clinically derived classification models based on heterogeneous learning disabled subjects comprised groups with various samples sizes. Larger and more even sized samples across these subtypes would have enabled statistical comparisons to be made between the groups. The discussion that follows and the interpretation of the research findings needs to be considered with these limitations in mind.

Discussion

The relative lack of success in discriminating between distinct clusters of learning disabled subjects and normal achievers in this study on the basis of SB:FE GPAB cognitive ability profiles suggests caution may be needed in utilizing SB:FE cognitive area scores and profiles in educational diagnosis, classification, and planning. The three cluster solution derived for all groups appeared to be related more to overall profile elevation and less to distinct patterns of performance characterized by intra-group strengths and weaknesses. While notions of differential diagnosis of learning difficulties and other educational problems using cognitive ability tests obtained from the SB:FE may be intuitively appealing, such practice may not be warranted. Normal learners and learning disabled groups with essentially the same composite cognitive ability score ranges were compared in the age groups, 5-6-11, 7-11-11, and 12-23-11. With the exception of the youngest age group, 5-6-11, no statistically significant or clinically significant differences were found between SB:FE scores. In contrast, WRAT-R Reading, Spelling, and Arithmetic scores were all significantly different ($p < .01$). Further, while the findings apply to the SB:FE GPAB, it is unlikely that markedly different results would be obtained with expanded versions of the SB:FE, given the close relationship between the two measures documented in this study.

As noted, profile differences that were found following the application of multivariate cluster analytic procedures to the various age groups, were largely based on differences in profile elevation across the four reasoning areas and not based on differences in shape among these areas. In the first part of the study, the significant correlations among the four cognitive areas which comprise the SB:FE were documented. Thus, the profiles that were obtained through application of the cluster analytic procedures may therefore be a function of the intercorrelations among Area SAS's analyzed to form groups. With this in mind, the SB:FE may most validly be viewed as a measure of g or overall ability. Thus, the SB.FE may remain a valuable tool for the assessment of global intellect, but may have limited utility in educational planning and classification, and in generating distinctly different patterns that are not based primarily on overall ability.

The findings tend to support the growing body of research in which arguments have been mounted against the use of cognitive measures for the classification of children on any other basis than overall intellect or global intelligence (Berk, 1983; Galvin, 1981; Kavale & Forness, 1984; Kramer, Henning-Stout, Ullman, & Schellenberg, 1987; Miller, 1980; Miller & Walker, 1981; Mueller et al., 1983, 1984, 1986).

Two recent studies conducted by Kline, Snyder, Guilmette, and Castellanos (1992, 1993) with the SB:FE are relevant to this dissertation. Kline et al. (1992) explored the relative usefulness of profile scatter information from the WISC-R, K-ABC, and the SB:FE. All three test batteries were administered to 146 children with an average age of 9.3 years referred for learning problems. Kline et al. (1992) predicted that because of the nature of information the SB:FE and K-ABC were designed to provide, the latter two tests' scatter information would provide greater incremental validity than the information provided by the WISC-R. Using hierarchical cluster analysis (SPSSx, 1990) of SB:FE subtest data to represent profile shape information three groups were obtained: the first group had relatively low verbal scores; the second group had a generally flat profile except for low scores on the Copying subtest; and the third group also had a flat profile, except for marginally lower scores on Bead Memory. Some differences were noted on achievement measures. However, when indices of profile elevation, variability, and shape were evaluated with hierarchical multiple regression, profile elevation was the most important predictor of achievement. Profile variability and shape had essentially no incremental validity. WISC-R profile shape, in contrast, had moderate unique predictive power. Kline et al. (1992) concluded that "profile shape information from the SB:FE may be even less useful than that from the WISC-R" (p. 431). Results of the analyses based on profile elevation, variability, and shape are convincing. However, the findings based on the application of cluster analysis form only a small part of their overall study, and tend to be difficult to interpret. Kline et al. (1992) fail to provide information on how the number of clusters was determined in their study, suffice to state "results of cluster analyses for all three IQ batteries indicated three-group solutions" (p. 428). Moreover, no information is provided about the clustering algorithm used, although, cosines, which are typically less commonly used, were selected as the similarity measure.

In a subsequent study, Kline et al. (1993) found that profile variability on the SB:FE increased with overall Composite SAS scores, which resulted in a very modest correlation between variability and achievement, but in the opposite direction expected. SB:FE variability, however, was unrelated to predicted achievement-aptitude discrepancies, i.e., to learning disabilities. Kline et al. (1993) suggested that profile variability had very limited validity, but they conceded that although unlikely, it is possible that profile variability may have external validity for certain types of profiles.

Results of this current study suggest that this is not the case. In this dissertation an attempt was made to address the issue whether specific types of profiles existed that were predictive of patterns of academic performance and that could differentiate normal from abnormal learners. The results clearly suggest that cognitive ability profiles of learning disabled subjects were not readily distinguishable from patterns of normal academic performers.

In examining the results of this study, two factors need to be considered in explaining the lack of success of the SB:FE in generating subtypes based on cognitive ability profiles that differentiated normal from abnormal learners, i.e., learning disabled subjects. First, as has been discussed, there were significant correlations among the four Reasoning Area SAS's of the SB:FE which suggest that each Reasoning Area SAS may have contributed in similar ways to the profile elevation differences that determined the group membership within the cluster solutions. Second, the significant correlation between cognitive ability and achievement scores may have a bearing on results. This may mean that little residual variance may be left over to be uniquely explained by other predictors, leaving elevation the predominant factor in determination of profiles or subgroup.

However, learning disabled subjects, simply described, are those individuals for whom the positive correlation between achievement and cognitive ability is not necessarily present. Thus, a significant discrepancy exists between performance in these two areas, albeit within clearly restricted or defined parameters, for example in certain academic achievement areas, e.g., in reading and spelling, but not in others, e.g., arithmetic, or vice-versa. However, this exceptionality or discrepancy did not appear to translate into consistent or meaningful profile differences on the SB:FE GPAB that could be used to differentiate subjects.

Kline et al. (1993) maintain that the clinical lore about profile variability is "old and well-established and makes conceptual sense" and that this belief may even be "resistant to change" despite empirical evidence to the contrary (p. 565). The notion of profile variability continues to generate much clinical interest. A number of parallels can be drawn between notions of profile variability and the clinical application of profile analysis in educational settings. Despite a number of negative findings relating to profile analysis, there remains considerable interest in the topic and ongoing research. However, it is important that educational decisions, diagnosis, and treatment planning be based firmly on empirical data and not on clinical lore that may have limited validity. On the basis of information derived from empirical data in this study and in others (Kline et al., 1992, 1993), the SB:FE appears to have limited validity in making fine distinctions among subjects and in identifying cognitive ability profiles associated with learning problems. Other than as a measure of overall intellectual ability, the SB:FE may be of limited utility in an educational setting. Uses of the SB:FE in educational diagnosis, placement, and treatment planning must therefore be restricted to these parameters for which validity has been demonstrated.

Conclusion

When the SB:FE was first published in 1986, it represented an ambitious departure from the traditional Binet scales and held much promise as an instrument with the potential to yield potentially rich and useful information about cognitive profiles and to contribute to educational diagnosis and planning. Educational diagnosis addresses both the issue of classification and the issue of determination of the nature of problems that individuals may experience so as to guide appropriate interventions and remediation. However the SB:FE appears valid primarily as an instrument to classify individuals according to overall ability level, and lacks validity as a diagnostic instrument when working with the four reasoning area scores that Thorndike et al. (1986b) suggested are important in educational decision making. It is possible that the SB:FE has failed to live up to the potentially rich uses it seemed to offer. Canter (1990) argued that while the SB:FE represented a number of technical advances over the earlier editions, it offered an unreasonably complex means of addressing typical referral concerns, but most importantly lacks treatment utility and failed to address the desired outcome of psychoeducational assessments, that of providing information to help professionals when formulating potential intervention strategies. Canter (1990) concluded that the Binet may contribute more to research than daily assessment practices of school psychologists. However, it is only through research that the validity and utility of the instrument can be demonstrated. This study does not provide much support for the use of the SB:FE other than as an index of global cognitive ability within an educational setting. In addition, on the basis of the findings of this study caution may be warranted in interpreting the results of the reasoning areas scores. There are, however, a number of implications for both practice and research that arise from the findings of this study.

Implications for School Psychology Practice

1. In the absence of unequivocal evidence to suggest that the SB:FE and the SB:FE GPAB contribute to the differential diagnosis of learning problems, caution is needed in using information from the SB:FE and abbreviated versions. This refers particularly to the use of cognitive ability profiles derived from the four Reasoning Areas scores to type individuals and plan remedial programs and interventions. The SB:FE and the SB:FE GPAB may most be useful as a measure of composite or global intelligence. Decisions about use of these tests must therefore be based on factors such as the need for a global estimate of cognitive ability, ease of use, and time considerations, but must not be guided primarily by the perhaps ill-founded belief that the four areas of the SB:FE may provide intuitively appealing information concerning diagnostic planning and remediation. To reiterate, in the current standards and guidelines for test use (APA, 1985; CPA, 1987; Principles for Fair Assessment Practice for Education in Canada, 1993) it is emphasized that validity must be demonstrated for particular test uses before such use is justified. The validity of the SB:FE and GPAB Reasoning Areas scores for such purposes has clearly not yet been demonstrated, despite its intuitive appeal and suggestions by Thorndike et al. (1986a, 1986b) that such practice is warranted.

2. Caution may also be needed in using the SB:FE with younger subjects, particularly where referral concerns are related to mental retardation or developmental delay. The instrument may lack adequate floors to make such assessments viable. Results of this study are consistent with reservations expressed by Glutting (1989) and Sattler (1992). Concomitantly, with respect to referral concerns related to giftedness in subjects within older age ranges (12-23-11), ceiling effects may be a factor. In trying to cover too wide an age range in one test, the designers of the SB:FE may have limited its utility and applicability.

3. The SB:FE GPAB may be a viable short form of the full battery that provides reasonably accurate representations of the Reasoning Area and Composite SAS's yielded by the more comprehensive battery. However, caution may be needed in interpreting the results in the

Abstract/Visual Reasoning Area; this caveat applies when dealing with either the SB:FE GPAB, or the SB:FE full battery when the Copying subtest is used.

<u>Implications for Further Research</u>

1. The results of this study have cast some doubt on the validity of profile analysis and subtyping applications with the SB:FE GPAB, and in particular suggest that the SB:FE GPAB may not contribute to differential diagnosis of normal versus learning disabled subjects. However, further research examining treatment by aptitude interventions based on profiles derived from samples of learning disabled subjects may shed further light on the validity of the SB:FE in educational settings. It may also be possible to empirically cluster samples of learning disabled children to determine whether subtypes consistent with findings in the literature with the WISC-R are supported. This does not, however, address the issue of whether profiles contribute to differential diagnosis or not.

2. Further, research with clinical samples consisting of larger and more balanced numbers and subtypes of learning disabled subjects, based on academic achievement criteria, may also be useful. For example, Williams (1991) criticized Rourke's (1990) research with learning disabled subjects because of limited coverage and limited clinical relevance, noting the infrequency with which subjects met the criteria for the non-verbal learning disabled subtype. Future research with the SB:FE should take this consideration into account. In addition, further research with clinical samples using not only Area SAS's, but also subtest SAS's, should be undertaken to assess the issue of the generalizability of the solutions obtained in this study, either supporting or disconfirming the findings presented here.

3. Research using different statistical clustering procedures as well as additional research with different samples is needed. An extremely large number of cluster procedures are possible. At each step in the clustering process - - determining the number of variables to be used, selecting distance measures, choosing clustering algorithms, and identifying the number of clusters - - the possible permutations for data analysis increase exponentially. Research is needed to explore the generalizability of the findings obtained in this study across different cluster methodologies.

4. Definitive interpretation of the trend, in which subtest and Reasoning Area and Composite SAS means decreased with age, was not possible in this present study given the fact that the referral concerns within each age group sampled differed significantly. Further research in this area may be warranted to fully address the limitations floor and ceiling effects may have on score interpretation in young and older subjects within the age range covered by the SB:FE.

5. Further research may be needed to address concerns relating to the lack of reliability and possible questionable validity of the Copying subtest.

6. A major difficulty in this study related to controversy over the diagnosis and definition of learning disabilities. Strict criteria, emphasizing academic performance, were adopted to operationalize the definition of learning disabilities for the purpose of this study. However, given the lack of consensus in the field, efforts need to be directed toward reaching agreement concerning diagnosis and definition. The need for a comprehensive taxonomy and organization of learning disabilities is also paramount, especially if effective aptitude by treatment interventions are to be developed. However, the notion that learning disabled subjects are qualitatively different from normal learners on the basis of cognitive ability profiles may not be viable. Ceci (1990) maintains that processing efficiency may be an important factor in learning disability research. Current measures of cognitive ability may not be sensitive enough to reflect subtle differences in profile ability that may also underlie or contribute to learning difficulties. Of the 15 subtests which comprise the four Area scores within the SB:FE, only one, Pattern Analysis, requires mandatory time limits (Glutting, 1989). Therefore the SB:FE may be further disadvantaged as a measure of processing efficiency. These are issues that need to be considered in future research with such populations.

REFERENCES

Adelman, H. S., & Taylor, L. (1985). The future of the LD field: A survey of fundamental concerns. Journal of Learning Disabilities, 18, 423-427.

Adelman, H. S., & Taylor, L. (1986). The problems of definition and differentiation and the need of classification schema. Journal of Learning Disabilities, 19, 154-520.

Agresti, A., & Finlay, B. (1986). Statistical methods for the social sciences. San Francisco, CA: Dellen Publishing Company.

Alberta Education (1986). Learning Disabilities: A resource manual for Teachers. Edmonton, AB, Canada: Author.

Alberta Education (1988). A teachers guide to learning disabilities in the classroom. Edmonton, AB, Canada: Author.

Alberta Education (1992). Students with challenging needs: Updated facts and figures. Edmonton, AB, Canada: Author.

Aldenderfer, M. S., & Blashfield, R. K. (1984). Cluster analysis. Beverley Hills, CA: Sage Publications.

Algozzine, B., & Ysseldyke, J. E. (1986). Tr future of the LD field: Screening and diagnosis. Journal of Learning Disabilities, 19, 394-398.

American Psychological Association (1985). Standards for educational and psychological testing. Washington, DC: Author.

Atkinson, L. (1991). Short forms of the Stanford-Binet Intelligence Scale, Fourth Edition, for children with low intelligence. Journal of School Psychology, 29(2), 177-181.

Atkinson, L., Bevc, I., Dickens, S. Blackwell, J. (1992). Concurrent validities of the Stanford-Binet (Fourth Edition), Leiter and Vineland with developmentally disabled children. Journal of School Psychology, 30(2), 165-173.

Anastasi, A. (1982). Psychological testing. New York: MacMillan.

Baddeley, A. D. (1976). The psychology of memory. New York: Basic Books.

Badian, N. A. (1983). Dyscalculia and nonverbal disorders of learning. In H. R. Myklebust (Ed.), Progress in learning disabilities (pp. 235-264). New York: Grune & Stratton.

Bakker, D. J. (1973). Hemispheric specialization and stages in the learning to read process. Bulletin of the Orton Society, 23, 15-27.

Bakker, D. J. (1979a). Perceptual asymmetries and reading strategies: Two dyslexias. Bulletin of the Orton Society, 29, 84-100.

Bakker, D. J. (1979b). Perceptual asymmetries and reading proficiency. In M. Bortner (Ed.), Cognitive growth and development (pp. 134-152). New York: Brunner/Mazzell.

Bannatyne, A. (1966). The colorphonics system. In J. Money & G. Schiffman (Eds.), The disabled reader: Education of the dyslexic child (pp. 193-214). Baltimore, MA: John Hopkins Press.

Bateman, B. (1968). Interpretation of the 1961 Illinois Test of Psycholinguistic Abilities. Seattle, WA: Special Child Publications

Bender, L. A. (1956). Psychology of children with organic brain disorders. Springfield, IL: Charles C. Thomas.

Bender, L. A. (1957). Specific reading disabilities as a maturational lag. Bulletin of the Orton Society, 7, 9-18.

Bennett, D. E., & Clarizio, H. F. (1988). A comparison of methods for calculating a severe discrepancy. Journal of School Psychology, 26, 359-369.

Berk, R. (1983). The value of Wisc-R profile analysis for differential diagnosis of learning disabled children. Journal of Clinical Psychology, 39(1), 133-136.

Biemiller, A., & Siegel, L. S. (1988). Developmental changes in relationships between reading comprehension, reading speed, reading accuracy, and working memory. Unpublished manuscript.

Binet, A., & Simon, T. (1905). Upon the necessity of establishing a scientific diagnosis of inferior states of intelligence. In E. S. Kit, 1916, (Trans.), The development of intelligence in children (pp. 9-181). Baltimore, MA: Williams & Wilkins.

Binet, A., & Simon, T. (1908). New methods for the diagnosis of the intellectual level of subnormals. In E. S. Kit, 1916, (Trans.), The development of intelligence in children (pp. 182-337). Baltimore, MA: Williams & Wilkins.

Birch, H. G., & Belmont, L. (1964). Auditory-visual integration in normal and retarded readers. American Journal of Orthopsychiatry, 34, 852-861.

Birch, H. G., & Belmont, L. (1965). Auditory-visual integration, intelligence, and reading ability in school children. Perceptual and Motor Skills, 20, 295-305.

Blakeslee, D. E. (1987). Psychometric properties, subtest patterning and concurrent validity of the Stanford-Binet Intelligence Scale, Fourth Edition with emotionally disturbed children (public school setting). Unpublished doctoral dissertation, Biola University, La Mirada, CA.

Blashfield, R. K. (1976). Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. Psychological Bulletin, 83, 377-388.

Blashfield, R. K. (1980). Propositions regarding the use of cluster analysis in clinical research. Journal of Consulting and Clinical Psychology, 48(4), 456-459.

Boder, E. (1970). Developmental dyslexia: A new diagnostic approach based on the identification of three subtypes. The Journal of School Health, 40, 289-290.

Boder, E. (1971). Developmental dyslexia: Prevailing diagnostic concepts and a new diagnostic approach. In H. R. Mycklebust (Ed.), Progress in learning disabilities (pp. 293-321). New York: Grune & Stratton.

Boder, E. (1973). Developmental dyslexia: A diagnostic approach based on three atypical reading-spelling patterns. Developmental Medicine & Child Neurology, 15, 663-687.

Boder, E., & Jarrico, S. (1982) The Boder Test of Reading-Spelling Patterns. New York: Grune & Stratton.

Boyle, G. J. (1989). Confirmation of the structured dimensionality of the Stanford-Binet Intelligence Scale: Fourth Edition. Personality & Individual Differences, 10(7), 709-715.

Boyle, G. J. (1990). Stanford-Binet IV Intelligence Scale: Is its structure supported by LISREL congeneric factor analysis? Personality & Individual Differences, 11(11), 1175-1181.

Breen, M. J. (1986). Cognitive patterns of learning disability subtypes as measured by the Woodcock-Johnson Psychoeducational Battery. Journal of Learning Disabilities, 19, 131-141.

Brodie, N. E. (1988). A validational investigation of the Luria-Nebraska Neuropsychological Battery for use with adolescents. Unpublished doctoral dissertation, University of Alberta, Edmonton, AB, Canada.

Brown, T. L., & Morgan, S. B. (1991). Concurrent validity of the Stanford-Binet. 4th Edition: Agreement with the WISC-R in classifying learning disabled children. Psychological Assessment, 3(2), 247-253.

Burgemeister, B. B., Blum, L. H., & Lorge, I. (1972). Columbia Mental Maturity Scale (3rd. ed.). San Antonio, TX: The Psychological Corporation.

Burt, C. L., & John, E. (1942a). A factor analysis of the Terman Binet tests: Part I. British Journal of Educational Psychology, 12, 117-127.

Burt, C. L., & John, E. (1942b). A factor analysis of the Terman Binet tests: Part II. British Journal of Educational Psychology, 12, 156-161.

Burt, C. L., & Stephenson, W. (1939). Alternative views on correlations between persons. Psychometrika, 4, 269-281.

Bush, W., & Waugh, K. (1976). Diagnosing learning disabilities. Columbus, OH: Charles Merrill.

Canadian Psychological Association (1987). Guidelines for educational and psychological testing. Old Chelsea, Quebec: Author.

Canter, A. (1990). A new Binet, and old premise: A mismatch between technology and evolving practice. Journal of Psychoeducational Assessment, 8(3), 443-450.

Carlson, L., & Reynolds, C. R. (1981). Factor structure and specific variance of the WPPSI subtests at six age levels. Psychology in the Schools, 18, 49-54.

...nter, R. L., & Miller, W. (1981). Spelling ability of reading disabled LD students and able ...aders. Learning Disability Quarterly, 5, 65-70.

...ajal, H. H., & Gerber. (1987). 1986 Stanford-Binet abbreviated forms. Psychological Reports, 61, 285-286.

Carvajal, H. H., Gerber, J., & Smith, P. D. (1987). Relationship between scores of young adults on Stanford-Binet IV and Peabody Picture Vocabulary Test-Revised. Psychological Reports, 63(3), 721-722.

Carvajal, H. H., Gerber, J., Hewes, P. & Weaver, K. A. (1987). Correlations between scores on Stanford-Binet IV and Wechsler Adult Intelligence Scale-Revised. Psychological Reports, 61, 83-86.

Carvajal, H. H., Hardy, K., Harmon, K, Sellers, T. A., & Holmes, C. B. (1987). Relationships among scores on the Stanford-Binet IV, Peabody Picture Vocabulary Test, and the Columbia Mental Maturity Scale. Bulletin of the Psychonomic Society, 25(4), 275-276.

Carvajal, H. H., Hardy, K., Smith, K. L., & Weaver, K. A. (1988). Relationships between scores on the Stanford-Binet IV and the Wechsler Preschool and Primary Scale of Intelligence. Psychology in the Schools, 25, 129-131.

Carvajal, H. H., Hayes, J. E., Lackey, K. L., Rathke, M. L., Wiebe, D. A., & Weaver, K. A. (1993). Correlations on scores on the Wechsler Intelligence Scale for Children-III and the General Purpose Abbreviated Battery of the Stanford-Binet IV. Psychological Reports, 72(3), 1167-1170.

Carvajal, H. H., & McNab, P. (1990). Relationships between scores of gifted students on Stanford-Binet IV and the SRA Educational Ability Series. Gifted Children Quarterly, 34(2), 80-82.

Carvajal, H. H., McVey, S., Sellers, T., Weyand, K., & McKnab, P. (1987). Relationships between scores on the General Purpose Abbreviated Battery, Columbia Mental Maturity Scale, and Goodenough Harris Drawing Scale. Psychological Record, 37(1), 127-130.

Carvajal, H. H., Parks, J. P., Bays, K. J., Logan, R. A., Lujano, C. I., Page, G. L., & Weaver, K. A. (1991). Relationships between scores on the Wechsler Preschool and Primary Scale of Intelligence-Revised and the Stanford-Binet IV. Psychological Reports, 69(1), 23-26.

Carvajal, H. H., & Weyand, K. (1986). Relationship between scores on the Stanford-Binet and Wechsler Intelligence Scale for Children-Revised. Psychological Reports, 59, 963-966.

Ceci, S. J. (1990). A sideways glance at this thing called LD: A context by process by person framework. In H. L. Swanson & B. K. Keogh (Eds.). Learning disabilities: Theoretical and research issues (pp. 73). Hillsdale, NJ: Erlbaum & Associates.

Childs, B., & Finucci, J. M. (1983). Genetics, epidemiology, and specific reading disability. In M. Rutter (Ed.), Developmental Neuropsychiatry (pp. 507-519). New York: Guilford Press.

Clarke, L. (1988). The Stanford-Binet Intelligence Scale: Fourth Edition: A validity study. Unpublished masters thesis, University of Alberta, Edmonton, Alberta.

Clarke, M. M. (1970). Reading difficulties in school. Harmondsworth, UK: Penguin Press.

Cohen, R. L., & Netley, C. (1978). Cognitive deficits, learning disabilities and WISC-R performance consistency. Developmental Psychiatry, 14(6), 624-634.

Cohen, R. L., & Netley, C. (1981). Short term memory deficits in reading disabled, children in the absence of opportunity for rehearsal strategies. Intelligence, 5(1), 69-76.

Conte, R., Samuels, M., & Zirk, H. (1983). Cognitive Correlates of reading and spelling patterns: An analysis of Boder's method of subtyping disabled readers. Calgary, AB, Canada: The Learning Centre.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change" - or should we? Psychological Bulletin, 74, 68-80.

Cronbach, L. J., & Gleser, G. C. (1953). Assessing the similarity between profiles. Psychological Bulletin, 50, 456-473.

CTB/McGraw Hill (1979). California Achievement Tests. Monterey, CA: Author.

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

Del Dotto, J. E., & B. P. Rourke. (1985). Subtypes of left-handed learning disabled children. In B. P. Rourke (Ed.), Neuropsychology of learning disabilities: Essentials of subtype analysis (pp. 89-132). New York: Guilford Press.

DeLamatre, J., & Hollinger, C. (1990). Utility of the Stanford Binet IV abbreviated form for placing exceptional children. Psychological Reports, 67(3), 973-974.

Delaney, E. A., & Hopkins, T. F. (1987). The Stanford-Binet Intelligence Scale: Fourth Edition Examiners Handbook: An expanded guide. Chicago, IL: The Riverside Publishing Company.

DeLuca, J. D., Adams, K. M., & Rourke, B. P. (1991). Methodological and statistical issues in cluster analysis. In B. P. Rourke (Ed.), Neuropsychological validation of learning disability subtypes (pp. 45-56). New York: Guilford Press.

DeLuca, J., Del Dotto, J., & Rourke, B. P. (1987). Subtypes of arithmetic disabled children: A neuropsychological taxonomic approach [Abstract]. Journal of Clinical and Experimental Neuropsychology, 9, 26.

Denckla, M. B. (1979). Childhood learning disabilities. In K. Heilman & E Valenstein (Eds.), Clinical Neuropsychology (pp. 535-573). New York: Oxford University Press.

Denckla, M. B., & Rudel, R. (1976a). Naming of pictured objects by dyslexic and other learning disabled children. Brain and Language, 3, 1-15.

Denckla, M. B., & Rudel, R. (1976b). Rapid "automatized" naming (RAN): Dyslexia differentiated from other learning disabilities. Neuropsychologia, 14, 471-479.

Dixon, W. (1981). BMDP statistical software. Los Angeles, CA: University of California Press.

Doehring, D. G., & Hoshko, I. M. (1977). Classification of reading problems by Q-technique of factor analysis. Cortex, 13, 281-294.

Doehring, D. G., Hoshko, I. M., & Bryans, B. N. (1979). Statistical classification of children with reading problem. Journal of Clinical Neuropsychology, 1, 289-310.

Duane, D. D. (1979). Toward a definition of dyslexia: A summary of views. Bulletin of the Orton Society, 29, 56-64.

Dunn, L. M., & Dunn, L. M. (1959). Peabody Picture Vocabulary Test. Circles Pines, MN: American Guidance Service.

Dunn, L. M., & Dunn, L. M. (1981). Peabody Picture Vocabulary Test-Revised: Manual for Forms L and M. Circles Pines, MN: American Guidance Service.

Dunn, L. M., & Markwardt, F. C. (1970). Peabody Individual Achievement Test. Circle Pines, MN: American GuidanceService.

Dykman, R. A., Ackerman, P. T., Clements, S. D., & Peters, J. E. (1971). Specific learning disabilities: An attentional deficit syndrome. In H. R. Myklebust (Ed.), Progress in learning disabilities (pp. 56-93). New York: Grune & Stratton.

Edelbrock, C. (1979). Comparing the accuracy of hierarchical clustering algotrithms: The problem of classifying everybody. Multivariate behavioural Research, 14, 367-384.

Epps, S., Ysseldyke, J. E., & Algozzine, B. (1983). Impact of different definitions of learning disabilities on the number of students identified. Journal of Psychoeducational Assessment, 1, 341-352.

Epps, S., Ysseldyke, J. E., & Algozzine, B. (1985). An analysis of the conceptual framework underlying definitions of learning disabilities. Journal of School Psychology, 23, 133-144.

Everitt, B. S. (1974). Cluster analysis. London, UK: Halsted Press.

Everitt, B. S. (1979). Unresolved problems in cluster analysis. Biometrics, 35, 169-181.

Everitt, B. S. (1980). Cluster analysis. New York: Halsted Press.

Fancher, R. E. (1985). The intelligence men: Makers of the IQ controversy. New York: Norton.

Fiedorowcz, C. A. M., & Trites, R. L. (1991). From theory to practice with subtypes of reading disabilities. In B. P. Rourke (Ed.), Neuropsychological validation of learning disability subtypes (pp. 243-270). New York: Guilford Press.

Fisk, J. L., & Rourke, B. P. (1979). Identification of subtypes of learning disabled children at three age levels: A neuropsychological multivariate approach. Journal of Clinical Neuropsychology, 1, 289-310.

Fleiss, J. L., & Zubin, J. (1969). On the methods and theory of clustering. Multivariate Behavior Research, 4, 235-250.

Fletcher, J. M. (1983). Verbal and spatial selective reminding in learning disability subtypes. Paper presented at the Society for Research in Child Development, Detroit, MI.

Fletcher, J. M. (1985). External validation of learning disability subtypologies. In B. P. Rourke (Ed.), Neuropsychology of learning disabilities: Essentials of subtype analysis (pp. 187-211). New York: Guilford Press.

Fletcher, J. M., & Morris R. (1986). Classification of disabled learners: Beyond exclusionary definitions. In S. J. Ceci (Ed.), Handbook of cognitive, social, and neuropsychological aspects of learning disabilities (pp. 55-80). Hillsdale, NJ: Lawrence Erlbaum.

Fletcher, J. M., & Satz, P. (1985). Cluster analysis and the search for learning disability subtypes. In B. P. Rourke (Ed.), Neuropsychology of learning disability: Essentials of subtype analysis (pp. 40-64). New York: Guilford Press.

Forness, S. R. (1990). Subtyping in learning disabilities: Introduction to the issues. In H. L. Swanson & B. K. Keogh (Eds.), Learning disabilities: Theoretical and research issues (pp. 195-200). Hillsdale, NJ: Erlbaum & Associates.

Francis, D. J., Espy, K. A., Rourke, B. P., & Fletcher, J. M. (1991). Validity of intelligence test scores in the definition of learning Disability: A critical analysis. In B. P. Rourke (Ed.), Neuropsychological validation of learning disability subtypes (pp. 15-44). New York: Guilford Press.

Frank, G. (1983). The Wechsler enterprise: An assessment of the development, structure, and use of the Wechsler tests of intelligence. New York: Pergamon.

Frauentreim, J. G. (1978). Academic achievement characteristics of male adults who were diagnosed as dyslexic in childhood. Journal of Learning Disabilities, 11, 476-483.

Freeman, F. S. (1955). Theory and practice of psychological testing. New York: Holt.

Fritzke, B. E. (1988). Validation of the Stanford-Binet Intelligence Scale: Factor analytic study of the Binet and WISC-R. Unpublished masters thesis, University of Alberta, Edmonton, AB, Canada.

Frostig, M. (1964). Frostig Developmental Test of Visual Perception. Palo Alto, CA: Consulting Psychologists Press.

Gaddes, W. H. (1981). An examination of the validity of neuropsychological knowledge in educational diagnosis and remediation. In G. W. Hynd & J. E. Obrzut (Eds.), Neuropsychological assessment and the school-aged child (pp. 27-84). New York: Grune & Stratton.

Gaddes, W. H. (1985). Learning disabilities and brain dysfunction: A neuropsychological approach. New York: Springer-Verlag.

Galaburda, A. M., & Eidelberg, D. (1982). Symmetry and asymmetry in the human posterior thalamus: Thalamic lesions in a case of developmental dyslexia. Archives of Neurology, 39, 333-336.

Galaburda, A. M., Sherman, G. F., Rosen, G. D., Aboitz, F., & Geshwind, N. (1985). Developmental dyslexia: Four consecutive cases with cortical anomalies. Annals of Neurology, 18, 222-233.

Galvin, G. (1981). Uses and abuses of the WISC-R with the learning disabled. Journal of Learning Disabilities, 14(6), 326-329.

Garrett, H. E. (1965). A developmental theory of intelligence. In A. Anastasi (Ed.), Individual Differences (pp. 73-83). New York: Wiley.

Gardner, E. F., Rudman, H. C., Karlson, B., & Merwin, J. C. (1982). Stanford Achievement Test. San Antonio, TX: The Psychological Corporation.

Glass, G. V (1976). Primary, secondary, and meta-analysis of research. Educational Researcher, 5, (10), 3-8.

Glass, G. V (1977). Integrating findings: the meta-analysis of research. In L. Shulman (Ed.), Review of Research in Education (pp. 351-379). Itasca, IL: Peacock.

Glass, G. V, McGaw, B., & Smith, M. L. (1981). Meta-analysis on class size and achievement. Educational Evaluation and Policy Analysis, 1, 2-16.

Glass, G. V, & Hopkins, K. D. (1984). Statistical methods in education and psychology. Englewood Cliffs, NJ: Prentice Hall.

Glutting, J. J. (1989). Introduction to the structure and application of the Stanford-Binet Intelligence Scale-Fourth Edition. The Journal of School Psychology, 27, 69-80.

Goldberg, H., & Schiffman, G. (1972). Dyslexia: Problems of reading disabilities. New York: Grune & Stratton.

Greene, A. C., Sapp, G. L., & Chissom, B. (1990). Validation of the Stanford-Binet Intelligence Scale: Fourth Edition with exceptional black male students. Psychology in the Schools, 27(1), 35-41.

Gross, A. L. (1972). A Monte Carlo study of the accuracy of a hierarchical grouping procedure. Multivariate Behavioural Research, 7, 379-389.

Grossman, F. M. (1983). Classification in mental retardation. Washington, DC: American Association on Mental Deficiency.

Gridley, B. E., & McIntosh, D. E. (1991). Confirmatory factor analysis of the Stanford-Binet: Fourth Edition for a normal sample. Journal of School Psychology, 29(3). 237-248.

Hale, R. L., & Sax, J. E. (1983). Profile analysis of the Wechsler Intelligence Scale for Children-Revised. Journal of Psychoeducational Assessment, 1, 155-162.

Hallahan, D. P., Ball, D. W., & Payne, J. S. (1973). Factorial composition of the short form of the Stanford-Binet with culturally disadvantaged Head Start Children. Psychological Reports, 32, 1048-1050.

Hammill, D. D. (1990). On defining learning disabilities: An emerging consensus. Journal of Learning Disabilities, 23(2), 74-84.

Hammill, D. D., Leigh, J., McNutt, G., & Larsen, S. (1981). A new definition of learning disabilities. Learning Disability Quarterly, 4, 336-342.

Hartigan, J. (1975). Clustering algorithms. New York: Wiley

Hartwig, S. S., Sapp, G. L., & Clayton, G. A. (1987). Comparison of the Stanford-Binet Intelligence Scale: Form L-M and the Stanford-Binet Intelligence Scale: Fourth Edition. Psychological Reports, 60(3), 1215-1218.

Hayden, D. C., Furlong, M. J., & Linnemeyer, S. (1988). A comparison of the Kaufman Assessment Battery for Children and the Stanford-Binet IV for the assessment of gifted children. Psychology in the Schools, 25(3), 239-243.

Halstead, W. C. (1947). Brain and intelligence: A quantitative study of the frontal lobes. Chicago, IL: University of Chicago Press.

Hendershott, J. L., Searight, H. R., Hatfield, J. L., & Rogers, B. J. (1990). Correlations between the Stanford-Binet, Fourth Edition and the Kaufman Assessment Battery for Children for a preschool sample. Perceptual and Motor Skills, 71(3), 819-825.

Henry, A. (1975). Specific difficulties in reading. Remedial Education, 10, 81-85.

Hermann, K. (1959). Reading disability: A medical study of word-blindness and related handicaps. Copenhagen: Munksgaard.

Hodapp, A. E. (1993). Correlation between Stanford-Binet IV and PPVT-R scores for young children. Psychological Reports, 73, 1152-1154.

Hoffman, R. (1988). Comparability of Binet (4th) full scale and abbreviated IQ's. NASP Communique, May, 27.

Hoffman, R., Carleton, R., Bishop-Marbury, A., & Goodwin, G. (1988). Exploratory and extended confirmatory analyses of the common factor of the new Binet intelligence test: A re-evaluation hypothesized structure. Paper presented at the annual meeting of the National Association of School Psychologists, Chicago, IL.

Hollinger, C. L., & Baldwin, C. (1990). The Stanford-Binet Fourth Edition: A small study of concurrent validity. Psychological Reports, 66(3), 1331-1336.

Holmes, J. M. (1986). Natural histories in learning disabilities: Neuropsychological Difference/Environmental Demand. In S. J. Ceci (Ed.), Handbook of cognitive, social, and neuropsychological aspects of learning disabilities (pp. 303-321). Hillsdale, NJ: Lawrence Erlbaum.

Hopkins, T. F. (1988). The fourth edition of the Stanford-Binet: Alfred Binet would have been proud... Measurement and Evaluation in Counseling and Development, 21, 40-41.

Hooper, S. R., & Willis, W. G. (1989). Learning disability subtyping: Neuropsychological foundations, conceptual models, and issues in clinical differentiation. New York: Springer-Verlag.

Horn, J. L. (1976). Human abilities: A review of research and theory in the early 1970s. Annual Review of Psychology, 27, 437-485.

Horn, J. L. (1982). The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood. In F. I. M. Craik & S. Trehub (Eds.), Aging and cognitive process (pp. 237-278). New York: Plenum.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. Journal of Educational Psychology, 57, 253-270.

Hynd, G. W. (1988). Neuropsychological assessment in clinical child psychology. Newbury Park, CA: Sage Publications.

Hynd, G. W. (1989). Foreword. In S.R. Hooper, S. R., & W. G. Willis (Eds.), Learning disability subtyping: Neuropsychological foundations, conceptual models, and issues in clinical differentiation (pp, vii-viii). New York: Springer-Verlag.

Hynd, G. W., Marshall, R., & Goncalez, J. (1991). Learning disabilities and presumed central nervous system dysfunction. Learning Disability Quarterly, 14, 283-296.

Hynd, G. W., & Obrzut, J. E. (1981). Neuropsychological assessment and the school-age child: Issues and procedures. New York: Grune & Stratton.

Ingram, T. S., Mason, A. W., & Blackburn, I. (1970). A retrospective study of 82 children with reading disability. Developmental Medicine & Child Neurology, 10, 19.

Interagency Committee on Learning Disabilities. (1987). Learning disabilities: A report to the U. S. Congress. Bethesda, MD: National Institute of Health.

Jastak, J. F., & Jastak, S. R. (1936). The Wide Range Achievement Test. Wilmington, DE: C. L. Story Co.

Jastak, J. F., & Jastak, S. R. (1946). The Wide Range Achievement Test. Wilmington, DE: C. L. Story Co.

Jastak, J. F., & Jastak, S. R. (1965). The Wide Range Achievement Test. Wilmington, DE: Guidance Associates.

Jastak, J. F., & Jastak, S. R. (1976). The Wide Range Achievement Test. Wilmington, DE: Guidance Associates.

Jastak, J. F., & Jastak, S. R. (1978). The Wide Range Achievement Test. Wilmington, DE: Guidance Associates.

Jastak, S. R., & Wilkinson, G. S. (1984). The Wide Range Achievement Test-Revised: Administration Manual. Wilmington, DE: Jastak & Associates Inc.

Jones, L. V. (1949). A factor analysis of the Stanford-Binet at four age levels. Psychometrika, 14, 299-331.

Johnson, D. L., Howie, V. M., Owen, M., & Baldwin, C. D. (1993). Assessment of three year olds with the Stanford-Binet Fourth Edition. Psychological Reports, 73(1), 51-57.

Johnson, D. J., & Myklebust, H. R. (1967). Learning disabilities, Educational Principles and practices. New York: Grune & Stratton.

Johnston, C. S., Fennell, E. B., & Satz, P. (1987). Learning disability subtypes: A cross-validation. Journal of Clinical & Experimental Neuropsychology, 9, 28.

Joreskog, K. G., & Sorbom, D. (1986). Lisrel IV user's guide, Fourth Edition. Chicago, IL: National Educational Resources.

Joschko, M., & Rourke, B. P. (1985). Neuropsychological subtypes of learning-disabled children who exhibit the ACID pattern on the WISC. In B. P. Rourke (Ed.), Neuropsychology of learning disabilities: Essentials of subtype analysis (pp. 65-88). New York: Guilford Press.

Kail, R. V., & Pelligrino, J. W. (1985). Human intelligence perspectives and prospects. New York: Freeman.

Kaplan, E., Fein, D., Morris, R., & Delis, D. C. (1991) WAIS-R NI. San Anotnio, TX: Psychological Corporation.

Karnes, F. A., & D'Ilio, V. R. (1987). Correlations of scores on verbal and nonverbal measures of intelligence for intellectually gifted students. Perceptual and Motor Skills, 64(1), 101-102.

Kaste, C. M. (1971). A ten year follow-up study of children diagnosed in a child guidance clinic as having cerebral dysfunction. Unpublished doctoral dissertation, University of Minnesota, MN.

Kaufman, A. S. (1975). Factor analysis of the WISC-R at 11 age levels between 61/2 and 16 1/2 years. Journal of Consulting & Clinical Psychology, 43, 135-147.

Kaufman, A. S. (1979). Intelligent testing with the WISC-R. New York: Wiley.

Kaufman, A. S. (1980). Issues in psychological assessment: Interpreting the WISC-R intelligently. In B. Lahey & A. Kazdin (Eds.), Advances in clinical child psychology (pp. 177-214). New York: Plenum.

Kaufman, A. S., & Kaufman, N. L. (1983). K-ABC: Kaufman Assessment Battery for Children of Educational Achievement. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (1985). Kaufman Test of Educational Achievement. Circle Pines, MN: American Guidance Service.

Kavale, K. A. (1990). A critical appraisal of empirical subtyping research in learning disabilities. In H. L. Swanson & B. K. Keogh (Eds.), Learning disabilities: Theoretical and research issues (pp. 215-227). Hillsdale, NJ: Erlbaum & Associates.

Kavale, K. A., & Forness, S. (1984). A meta-analysis assessing the validity of Wechsler Scale profiles and recategorizations: Patterns or parodies? Learning Disability Quarterly, 7(2), 136-156.

Keith, T. Z. (1987). Assessment research: An assessment and recommended interventions. School Psychology Review, 16, 276-289.

Keith, T. Z., Cool, V. A., Novak, C. G., White, L. J., & Pottebaum, S. M. (1988a). Confirmatory factor analysis of the Stanford-Binet Fourth Edition: Testing the theory-test match. Journal of School Psychology, 26, 253-274.

Keith, T. Z., Cool, V. A., Novak, C. G., White, L. J., & Pottebaum, S. M. (1988b). Hierarchical confirmatory factor analysis of the Stanford-Binet Fourth Edition: Testing the theory-test match. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Keogh, B. K. (1990). Definitional assumptions and research issues. In H. L. Swanson & B. K. Keogh (Eds.), Learning disabilities: Theoretical and research issues (pp. 13-19). Hillsdale, NJ: Erlbaum & Associates.

Kephart, N. C. (1971). The slow learner in the classroom. Columbus, OH: Charles E. Merrill.

Kinsbourne, M., & Warrington, E. K. (1963). Developmental factors in reading and writing backwardness. British Journal of Psychology, 54, 145-146.

Kirk, S. A. (1962). Educating exceptional children. Boston, MA: Houghton Mifflin.

Kirk, S. A. (1963). Behavioral Diagnosis and remediation of learning disabilities. Proceedings of the conference on exploration into the problems of the Perceptually Handicapped Child. Chicago, IL.

Kitano, M. K., & de Leon, J. (1988). Use of the Stanford-Binet Fourth Edition in identifyir : young gifted children. Roeper Review, 10(3), 156-159.

Kline, R. B. (1989). Is the Fourth Edition Stanford-Binet a four-factor test? Confirmatory factor analysis of alternative models for ages 2 through 23. Journal of Psychoeducational Assessment, 7(1), 4-13.

Kline, R. B., Snyder, J., Guilmette, S., & Castellanos, M. (1992). Relative usefulness of elevation, variability, and shape information from WISC-R, K-ABC, and Fourth Edition Stanford-Binet profiles in predicting achievement. Psychological Assessment, 4(4), 426-432.

Kline, R. B., Snyder, J., Guilmette, S., & Castellanos, M. (1993). External validity of the Profile Variability index for the K-ABC, Stanford-Binet, and WISC-R: Another a cul-de-sac. Journal of Learning Disabilities, 26(8), 557-567.

Kluever, R. C., & Green, K. E. (1990). Identification of gifted children: A comparison of the scores on Stanford-Binet 4th Edition and Form LM. Roeper Review, 13(1), 16-20.

Knight, C. B., Baker, E. H., & Minder, C. C. (1990). Concurrent validity of the Stanford-Binet: Fourth Edition and Kaufman Assessment Battery for Children with learning disabled children. Psychology in the Schools, 27(2), 116-125.

Knights, R. M., & Bakker, D. J. (1976). The neuropsychology of learning disorders. Baltimore, MA: University Park Press.

Kraemer, H. C., & Andrews, G. (1982). A nonparametric technique for meta-analysis effect size calculation. Psychological Bulletin, 94, 190-192.

Kramer, J. J., Henning-Stout, M., Ullman, D. P., & Schellenberg, R. P. (1987). The viability of scatter analysis on the WISC-R and the SBIS: Examining a vestige. Journal of Psychoeducational assessment, 5, 37-47.

Krohn, E. J., & Lamp, R. E. (1989). Concurrent validity of the K-ABC, Binet L-M and Binet Four with preschoolers. Paper presented at the 98th Annual convention of the American Psychological Association, Boston, MA.

Krohn, E. J., & Lamp, R. E. (1990). Predictive validity of the Stanford-Binet Fourth Edition and K-ABC for Head Start Children. Journal of School Psychology, 27(1), 59-67.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. Psychometrika, 2, 151-160.

Kuhlmann, F. (1912). A revision of the Binet-Simon system for measuring the intelligence of children. Journal of Psycho-asthenics, Monograph Supplement, 1, 1-41.

Kuiper, F. K., & Fisher, L. A. (1975). A Monte Carlo comparison of six clustering procedures. Biometrics, 31, 777-783.

Kussmaul, A. (1877). Disturbance of speech. Cyclopedia of Practical Medicine, 14, 581.

Laurent, J., Swerdlik, M., & Ryburn, M. (1992). Review of validity research on the Stanford-Binet Intelligence Scale: Fourth Edition. Psychological Assessment, 4(1), 102-112.

Learning Disabilities Association of America. (1986). ACLD Description: Specific Learning Disabilities. ACLD Newsbriefs, 15-16.

Learning Disabiliies Association of Canada (1981). Annual Report. Toronto: Author.

Lee, H. B., & MacQueen, J. B. (1980). A k-means cluster analysis computer program with cross-tabulations and Next-Nearest-Neighbour Analysis. Educational and Psychological Measurement, 40, 133-139.

Leiter, R. G. (1948). Leiter International Performance Scale. Chicago, IL: Stoelting Co.

Lennox, C., & Siegel, L. S. (1988). Visual and phonetic spelling errors in subtypes of children with learning disabilities. Paper presented at the 16th annual International Neuropsychological Society meeting, New Orleans, LA.

Lennox, C., & Siegel, L. S. (1993). Visual and phonological spelling errors in subtypes of children with learning disabilities. Applied Psycholinguistics, 14(4), 473-488.

Lewis-O'Donnell, M. (1986). A comparison study of the performance of inpatient emotionally disturbed non-psychotic children and the standardization sample using the Stanford-Binet (Fourth Edition). Unpublished doctoral dissertation, Biola University, La Mirada, CA.

Livesay, K. K. (1986, October). Comparisons of the Stanford Binet: Fourth Edition to the S-B-L-M and WISC-R with gifted referrals. Paper presented at the annual conference of the Florida Association of School Psychologists, Jacksonville, FL.

Lorr, M. (1983). Cluster analysis for social scientists. San Francisco, CA: Jossey-Bass.

Lubin, B., Larsen, R. M., & Matarazzo, J. D. (1984). Patterns of psychological test usage in the United States: 1935-1982. American Psychologist, 39, 451-454.

Lubin, B., Larsen, R. M., Matarazzo, J. D., & Seever, M. (1985). Psychological test usage patterns in five professional settings. American Psychologist, 40, 857-861.

Lukens, J. (1987). Comparison of the Fourth Edition and the L-M Edition of the Stanford-Binet used with mentally retarded persons. Journal of School Psychology, 26(1), 87-89.

Lukens, J. (1990). Stanford-Binet: Fourth Edition and the WISC-R in the lower range of intelligence. Perceptual and Motor Skills, 70(3), 819-822.

Lyle, J. G. (1969). Reading retardation and reversal tendency: A factorial study. Child Development, 40, 833-843.

Lyle, J. G., & Goyen, J. (1968). Visual recognition, development lag and strephosymbolia in reading retardation. Journal of Abnormal Psychology, 73, 25-29.

Lyle, J. G., & Goyen, J. (1975). Effects of speed of exposure and difficulty of discrimination on visual recognition of retarded readers. Journal of Abnormal Psychology, 84, 673-676.

Lyon, G. R. (1983). Learning disabled readers: Identification of subgroups. In H. R. Myklebust (Ed.), Progress in learning disabilities (pp. 103-133). New York: Grune & Stratton.

Lyon, G. R. (1985). Identification and remediation of learning disability subtypes: Preliminary findings. Learning disabilities Focus, 1, 21-35.

Lyon, G. R., & Flynn, J. M. (1991). Educational validation studies with subtypes of learning-disabled readers. In B. P. Rourke (Ed.), Neuropsychological validation of learning disability subtypes (pp. 233-242). New York: Guilford Press.

Lyon, G. R., & Risucci, D. (1988). Classification of learning disabilities. In K. Kavale, (Ed.), Learning disabilities: State of the art and practice (pp. 44-70). Boston, MA: Little Brown/College Hill.

Lyon, G. R., Stewart, N., & Freedman, D. (1982). Neuropsychological characteristics of empirically derived subgroups of learning disabled readers. Journal of Clinical Neuropsychology, 4, 343-365.

Lyon, G. R., & Watson, B. (1981). Empirically derived subgroups of learning disabled readers: Diagnostic characteristics. Journal of Learning Disabilities, 14, 256-261.

MacGregor, K., Rosenbaum, S., & Skoutajan, K. (1982). Putting the pieces together: A parent's guide to special education in Ontario. Toronto: Association for Children with Learning Disabilities.

Mason, E. M. (1992). Percentage of agreement among raters and rater reliability of the Copying subtest of the Stanford-Binet Intelligence: Fourth Edition. Perceptual and Motor Skills, 74(2), 347-353.

McCall, V. W., Yates, B., Hendricks, S., Turner, K., & McNabb, B. (1989). Comparison between the Stanford-Binet: L-M and the Stanford-Binet: Fourth Edition with a group of gifted children. Contemporary Educational Psychology, 14(2), 93-96.

McCallum, R. S. (1990). Determining the factor structure of the Stanford-Binet --Fourth Edition: The right choice. Journal of Psychoeducational Assessment, 8(3), 436-442.

McCallum, R. S., & Karnes, F. A. (1987). Comparison of intelligence tests: Responses of gifted Pupils to the Stanford-Binet Intelligence Scale (4th ed.), the British Ability Scales, and the Wechsler Intelligence Scale for Children-Revised. School Psychology International, 8(2-3) 133-139.

McCallum, R. S., & Karnes, F. A. (1990). Use of a brief form of the Stanford-Binet Intelligence Scale (Fourth) for gifted children. Journal of School Psychology, 28(3) 279-283.

McCallum, R. S., Karnes, F. A., & Crowell, M. (1988). Factor structure of the Stanford-Binet Intelligence Scale (Fourth Edition) for gifted children. Contemporary Educational Psychology, 13(4) 331-338.

McNemar, Q. (1942). The revision of the Stanford-Binet Scale: An analysis of the standardization data. Boston, MA: Houghton Mifflin Company

Meloff, R. (1987). The Stanford Binet Intelligence Scale: Fourth Edition: A descriptive study. Unpublished masters thesis, University of Alberta, Edmonton, AB, Canada.

Mendelson, W., Johnson, N., & Stewart, M. (1971). Hyperactive children as teenagers: A follow-up study. Journal of Nervous and Mental Diseases, 153, 272-279.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational Measurement (pp. 11-104). New York: MacMillan.

Miller, M. (1980). On the attempt to find WISC-R profiles for learning and reading disabilities. Journal of Learning Disabilities, 13(6), 52-54.

Miller, M., Stoneburger, R., & Brecht, R. (1978). WISC-R subtest patterns as discriminators of perceptual disability. Journal of Learning Disabilities, 11(6), 449-452.

Miller, M. & Walker, K. (1981). The myth of the LD Wisc-R profile. Exceptional Child, 28 (2), 83-88.

Milligan, G. W. (1981). A review of Monte Carlo tests of cluster analysis. Multivariate Behavioural Research, 16, 379-407.

Mitterer, J. O. (1982). There are at least two kinds of poor readers: Whole word readers and recoding poor readers. Canadian Journal of Psychology, 36, 445-461.

Mojena, R. (1977). Hierarchical grouping methods and stopping rules: An evaluation. The Computer Journal, 20, 359-363.

Molfese, V. J., Helwig, S., & Holcomb, L. (1993). Standardized assessments of verbal intelligence in 3-year-old children: A comparison of biomedical and psychoeducational data in a longitudinal sample. Journal of Psychoeducational Assessment, 11(10), 55-66.

Molfese, V. J., Yaple, K., Helwig, S., Harris, L., & Connell, S. (1992). Stanford-Binet Intelligence Scale (Fourth Edition): Factor structure and verbal scale scores for three year olds. Journal of Psychoeducational Assessment, 10, 47-58.

Morris, R. D., Blashfield, R. K., & Satz, P. (1986). Developmental classification of reading-disabled children. Journal of Clinical and Experimental Neuropsychology, 8, 371-392.

Mueller, H. H., Dash, U. N, Matheson, D. W., & Short, R. H. (1984). WISC-R subtest patterning of below average, average, and above average IQ children: A meta-analysis. The Alberta Journal of Educational Research, 30(1), 68-85.

Mueller, H. H., Dennis, S. S., & Short, R. H. (1986). A meta-exploration of WISC-R factor scores as a function of diagnosis and intellectual level. Canadian Journal of School Psychology, 2, 21-43.

Mueller, H. H., Matheson, D. W., & Short, R. H. (1983). Bannatyne-recategorized WISC-R patterns of mentally retarded, learning disabled, normal and intellectually superior children: A meta-analysis. The Mental Retardation and Learning Disability Bulletin, 11, 60-78.

Murphy, K. R., & Davidshofer, C. O. (1991). Psychological testing: Principles and applications. Englewood Cliffs, NJ: Prentice Hall.

Naidoo, S. (1972). Specific dyslexia. New York, Wiley.

Nagle, R. J., & Bell, N. L. (1993). Validation of the Stanford-Binet Intelligence Scale: Fourth Edition abbreviated batteries with college students. Psychology in the Schools, 30(3), 227-231.

Naglieri, J. A. (1988). Interpreting the subtest profile on the fourth edition of the Stanford-Binet Scale of Intelligence. Journal of Clinical Child Psychology, 17(1), 62-65.

Naglieri, J. A., & Kaufman, A. S. (1983). How many factors underlie the WAIS-R? Journal of Psychoeducational Assessment, 1, 113-119.

National Joint Committee on Learning Disabilities (1988). Letter to NJCLD members.

Nelson, H. E., & Warrington, E. K. (1974). Developmental spelling retardation and its relation to other cognitive abilities. British Journal of Psychology, 65, 265-274.

Nolan, D. R., Hammeke, T. A., & Barkley, R. A. (1983). A comparison of the patterns of the neuropsychological performance in two groups of learning disabled children. Journal of Clinical Child Psychology, 12, 13-21.

Nurss, J. R., & McGauvran, M. E. (1986). Metropolitan Readiness Test. San Antonio, TX: Psychological Corporation.

Obrzut, J. E., & Hynd, G. W. (1986). Child Neuropsychology. Orlando, FL: Academic Press.

Orton, S. T. (1928). Specific reading disability-strephosymbolia. Journal of the American Medical Association, 90, 1005-1009.

Orton, S. T. (1937). Reading, writing and speech problems in children. New York: Norton.

Osberg, D. (1986). New Tests: New Options. NASP Communique, 15(3), 3.

Ownby, R., &, Carmin, C. (1988) Confirmatory factor analysis of the Stanford-Binet Intelligence Scale, Fourth Edition. Journal of Psychoeducational Assessment, 6(4), 331-340.

Ozols, E. J., & Rourke, B. P. (1991). Classification of young learning disabled children according to patterns of academic achievement: Validity studies. In B. P. Rourke (Ed.), Neuropsychological validation of learning disability subtypes (pp. 97-121). New York: Guilford Press.

Petrauskas, R., & Rourke, B. P. (1979). Identification of subgroups of retarded readers: A neuropsychological multivariate approach. Journal of Clinical Neuropsychology, 1, 17-37.

Phelps, L. (1989). Comparison of scores for gifted students on the WISC-R and the Fourth Edition of the Stanford-Binet. Psychology in the Schools, 26(2), 125-129.

Phelps, L., Bell, C. M., & Scott, M. J. (1988). Correlations between the Stanford-Binet: Fourth Edition and the WISC-R with learning disabled populations. Psychology in the Schools, 25, 380-382.

Phillips, S. E., & Clarizio, H. F. (1988). Limitations of standard scores in individual achievement testing. Educational Measurement: Issues and practice, 7, 8-15.

Powers, A., D., Church, J. K., & Treolar, J. (1989). The fourth edition of the Stanford-Binet Intelligence Scale and the Woodcock-Johnson Tests of Achievement: A criterion validity study. Paper presented at the Annual meeting of the National Association of School Psychologists, Boston, MA.

Pirrozzolo, F. J. (1979). The neuropsychology of developmental reading disorders. New York: Prager Press.

Prewett, P. N. (1992). Short forms of the Stanford-Binet Intelligence Scale: Fourth Edition. Journal of Psychoeducational Assessment, 10(3), 257-264.

Prewett, P. N., & Giannuli, M. N. (1991). Correlations of the WISC-R, Stanford-Binet Intelligence Scale: Fourth Edition, and the reading subtests of three popular achievement tests. Psychological Reports, 69(3), 1232-1234.

Prewett, P. N., & McCaffery, L. K. (1993). A comparison of the Kaufman Brief Intelligence Test (K-BIT) with the Stanford-Binet Intelligence Scale (4th ed.) and the Kaufman Test of Educational Achievement (K-TEA) brief form. Psychology in the Schools, 30(4), 299-304.

Principles of Fair Student Assessment Practices for Education in Canada. (1993). Edmonton Alberta, Joint Advisory Committee.

Quiros, J. B. (1964). Dysphasia and dyslexia in school children. Folia Phoniatrica, 16, 201.

Ramsey, P. H., & Vane, J. R. (1970). A factor analytic study of the Stanford-Binet with young children. Journal of School Psychology, 8, 278-283.

Rasch, G. (1980). Probabilistic Models for some intelligence and attainment tests. Chicago, IL: The University of Chicago Press,.

Reitan, R. M., & Davidson, L. A. (1974). Clinical neuropsychology: Current status and applications. New York: Wiley and Sons.

Reynolds, C. R. (1981). The fallacy of "two years below grade level for age" as a diagnostic criteria for reading disorders. Journal of School Psychology, 19, 350-358.

Reynolds, C. R. (1984). Critical measurement issues in learning disabilities. Journal of Special Education, 18, 451-476.

Reynolds, C. R. (1988). Putting the individual into aptitude treatment interaction. Exceptional Children, 54(4), 324-331.

Reynolds, C. R., & Clarke, J. R. (1985). Profile analysis of standardized intelligence test performance of very low functioning individuals. Journal of School Psychology, 23, 277-283.

Reynolds, C. R., Kamphaus, R., & Rosenthal, B. (1988). Factor analysis of the Stanford-Binet Fourth Edition for ages 2 through 23. Measurement and Evaluation in Counseling and Development, 21, 52-63.

Robinson, N. M., Dale, P. S., & Landesman, S. (1990). Validity of the Stanford-Binet IV with precocious toddlers. Intelligence, 14(2), 173-176.

Robinson, E. L., & Nagle, R. J. (1992). The comparability of the Test of Cognitive Skills with the Wechsler Intelligence Scale for Children-Revised and the Stanford-Binet: Fourth Edition with gifted children. Psychology in the Schools, 29(2), 107-112.

Rosenthal, B. L., & Kamphaus, R. W. (1988). Interpretive tables for test scatter on the Stanford-Binet Intelligence Scale: Fourth Edition. Journal of Psychoeducational Assessment, 6(4), 359-370.

Ross, A. O. (1976). Psychological aspects of learning disabilities and reading disorders. New York: McGraw Hill.

Rothlisberg, B. A. (1987). Comparing the Stanford-Binet, Fourth Edition to the WISC-R: A concurrent validity study. Journal of School Psychology, 25(2), 193-196.

Rothlisberg, B. A., & McIntosh, D. E. (1991). Performance of a referred sample on the Stanford-Binet IV and the K-ABC. Journal of School Psychology, 29(4), 367-370.

Rourke, B. P. (1975). Brain behavior relationships in children with learning disabilities: A research program. American Psychologist, 30, 911-920.

Rourke, B. P. (1978). Reading, spelling, arithmetic disabilities: A neuropsychological perspective. In H. R. Myklebust (Ed.), Progress in learning disabilities (pp. 97-120). New York: Grune & Stratton.

Rourke, B. P. (1982). Central processing deficiencies in children: Toward a developmental neuropsychological model. Journal of Clinical Neuropsychology, 4, 1-18.

Rourke, B. P. (1983). Outstanding issues in research on learning disabilities. In M. Rutter (Ed.), Developmental neuropsychiatry (pp. 564-574). New York: Guilford Press.

Rourke, B. P. (1985). Overview of learning disability subtypes. In B. P. Rourke (Ed.), Neuropsychology of learning disabilities: Essentials of subtype analysis (pp. 3-16). New York: Guilford Press.

Rourke, B. P. (1989). Nonverbal Learning Disabilities: The syndrome and the model. New York: Guilford Press.

Rourke, B. P. (1991). Validation of learning disability subtypes. In B. P. Rourke (Ed.), Neuropsychological validation of learning disability subtypes (pp. 3-14). New York: Guilford Press.

Rourke, B. P., & Finlayson, M. A. P. (1978). Neuropsychological significance of variations in patterns of academic performance: verbal and visual spatial abilities. Journal of Abnormal Child Psychology, 3, 62-66.

Rourke, B. P., Fisk, J. L. & Strang, J. D. (1986). Neuropsychological assessment of children: A treatment-oriented approach. New York, Guilford Press.

Rourke, B. P., & Fuerst, D. R. (1991). Learning disabilities and psychosocial functioning: A neuropsychological perspective. New York: The Guilford Press.

Rourke, B. P., & Strang, J. D. (1978). Neuropsychological significance of variations in patterns of academic performance: Motor, psychomotor, and tactile-perceptual abilities. Journal of Pediatric Psychology, 3, 62-66.

Rourke, B. P., Bakker, D. J., Fisk, J. L., & Strang, J. D. (1983). Child neuropsychology: An introduction to theory, research, and clinical practice. New York: Guilford Press.

Rourke, B. P., Fisk, J. L., & Strang, J. D. (1986). Neuropsychological assessment of children: A treatment-oriented approach. New York: Guilford Press.

Rourke, B. P., Young, G. C., & Flewelling, R. W. (1971). The relationship between WISC verbal-performance discrepancies and selected verbal, auditory-perceptual, visual perceptual, and problem-solving abilities in children with learning disabilities. Journal of Clinical Psychology, 27, 475-479.

Saccuzzo, D., & Lewandowski, D. (1976). The WISC-R as a diagnostic tool. Journal of Clinical Psychology, 32(1), 115-124.

Salvia, J., & Ysseldyke, J. E. (1985). Assessment in special and remedial education (3rd ed.). New York: Houghton Mifflin.

Salvia, J., & Ysseldyke, J. E. (1988). Assessment in special and remedial education (4th ed.). New York: Houghton Mifflin.

Sandoval, J., & Irwin, M. G. (1986). A review: Stanford-Binet Intelligence Scale: Fourth Edition, technical manual. CASP Today, 36(2), 3-5.

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

Sattler, J. M. (1988). Assessment of children (3rd ed.). San Diego, CA: Author.

Sattler, J. M. (1992). Assessment of children (3rd edition-revised). San Diego, CA: Author.

Satz, P., & Morris, R. (1981). Learning disability subtypes: A review. In F. J. Pirozzolo & M. C. Wittrock (Eds.), Neuropsychological and cognitive processes in reading (pp. 109-141). New York: Academic Press.

Satz, P., Rardin, D. & Ross, J. (1971). An evaluation of a theory of specific developmental dyslexia. Child Development, 42, 2009-2021.

Scheibler, D., & Schneider, W. (1985). Monte Carlo tests of the accuracy of cluster analysis algorithms: A comparison of hierarchical and nonhierarchical methods. Multivariate Behavioural Research, 20, 283-304.

Schaie, K., W., & Willis, S. L. (1991). Adult development and aging. New York: Harper Collins.

Senf, G. M. (1969). Development of immediate memory of bisensory stimuli in normal children and children with learning disabilities. Developmental Psychology Monograph, Part 2, 1(6), 1-28.

Shepard, L. A. (1980). An evaluation of the regression discrepancy method for identifying children with learning disabilities. Journal of Special Education, 14, 79-91.

Shepard, L. A. (1983). The role of measurement in educational policy: Lessons from the identification of learning disabilities. Educational Measurement: Issues and Practice, 2, 4-8.

Shepard, L. A. (1989). Identification of mild handicaps. In R. L. Linn (Ed.), Educational Measurement (pp. 545-572). New York: MacMillan.

Shepard, L. A., Smith, M. L., & Vojir, C. P. (1983). Characteristics of pupils identified as learning disabled. American Educational Research Journal, 20, 309-331.

Siegel, L. S. (1984). On the adequacy of the Wide Range Achievement Test (WRAT): A reply to Snart, Dennis, and Brailsford. Canadian Psychology, 25, 73-74.

Siegel, L. S. (1986). Phonological deficits in children with a reading disability. Canadian Journal of Special Education, 2, 45-54.

Siegel, L. S. (1990). IQ and learning disabilities: R.I.P. In H. L. Swanson & B. K. Keogh (Eds.), Learning disabilities: Theoretical and research issues (pp. 111-128). Hillsdale, NJ: Erlbaum & Associates.

Siegel, L. S., & Heaven, R. (1986). Defining and categorizing learning disabilities. In S. J. Ceci (Ed.), Handbook of cognitive, social, and neuropsychological aspects of learning disabilities (pp. 95-121). Hillsdale, NJ: Lawrence Erlbaum.

Siegel, L. S., & Linder, B. A. (1984). Short-term memory processes in children with reading and arithmetic learning disabilities. Developmental Psychology, 20, 200-207.

Siegel, L. S., & Ryan, E. B. (1984). Reading disability as a language disorder. Remedial and Special Education, 5, 28-33.

Silver, A. A. (1968). Diagnostic considerations in children with reading disability. In G. Natchez (Ed.), Children with reading problems: Classic and contemporary issues in reading disability (pp. 240-250). New York: Basic Books.

Silver, A. A., & Hagin, R. A. (1964). Specific reading disability: Follow-up studies. American Journal of Orthopsychiatry, 34, 95-102.

Sim, J. (1987). The Stanford-Binet IV: Recent criticisms and issues. The Alberta School Psychologist, 7(1), 17-21.

Skinner, H. A. (1978). Differentiating the contribution of elevation, scatter, and shape in profile similarity. Educational and Psychological Measurement, 38, 297-308.

Skinner, H. A. (1981). Toward the integration of classification theory and methods. Journal of Abnormal Psychology, 90, 68-87.

Skinner, H. A., & Lei, H. (1980). Manual for modal profile analysis. Toronto, ON: Addiction Research Foundation.

Slate, J. (1986). A reaction to the revised Stanford-Binet Intelligence Scale: New does not necessarily mean better. NASP Communique, 15(1), 3.

Slavin, R. (1983). Meta-nonsense: Misuse of meta-analysis in educational research. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Slavin, R. (1984). Meta-analysis in education: How has it been used? Educational Researcher, 3(18), 6-15.

Smith M. M. (1970). Patterns of intellectual abilities in educationally handicapped children. Unpublished doctoral dissertation, Claremont College, Claremont, CA.

Smith, D. K., St. Martin, M. E., & Lyon, M. A. (1989). A validity study of the Stanford-Binet: Fourth Edition with students with learning disabilities. Journal of Learning Disabilities, 22(4), 260-261.

Smith, M. D., Coleman, J. M., Dokecki, P. R., & Davis, E. E. (1977). Recategorized WISC-R scores of learning disabled children. Journal of Learning Disabilities, 10, 444-449.

Sneath, P. (1957). The application of computers to taxonomy. Journal of General Microbiology, 17, 201-226.

Snow, J. H., & Hynd, G. W. (1985a). Factor structure of the Luria-Nebraska Neuropsychological Battery - Children's Revision with learning-disabled children. Journal of School Psychology, 23, 271-275.

Snow, J. H., & Hynd, G. W. (1985b). A multivariate investigation of the Luria-Nebraska Neuropsychological Battery-Children's Revision with learning-disabled children. Journal of Psychoeducational Assessment, 3, 101-109.

Snow, J. H., Cohen, M., & Holliman, W. B. (1985). Learning disability subtypes using cluster analysis of the WISC-R. Journal of Psychoeducational Assessment, 4, 391-397.

Snow, J. H., Koller, J. R., & Roberts, C. D. (1987). Adolescent and adult learning disability subgroups based on WAIS-R performance. Journal of Psychoeducational assessment, 5, 7-14.

Sokal, R., & Sneath, P. (1963). Principles of numerical taxonomy. San Francisco, CA: W. H. Freeman.

Spearman, C. (1904). General Intelligence objectively determined and measured. American Journal of Psychology, 15, 201-203.

Speece, D. L. (1990). Methodological issues in cluster analysis: How clusters become real. In H. L. Swanson & B. K. Keogh (Eds.), Learning disabilities: Theoretical and research issues (pp. 201-213). Hillsdale, NJ: Erlbaum & Associates.

Spelliscy, R. (1991). Theoretical and factorial conceptualizations of human intelligence: An evaluation of the Stanford-Binet Intelligence Test Fourth Edition. Unpublished doctoral dissertation, University of Alberta, Edmonton, Alberta, Canada.

SPSSx (1990). SPSSx reference guide. Chicago, IL: SPSSx.

Spreen, O. (1976). Neuropsychology of learning disabilities: Post conference review. In R. M. Knights & D. J. Bakker (Eds.), The neuropsychology of learning disabilities (pp. 445-468). Baltimore, OH: University Park Press.

Spreen, O. (1982). Adult outcome of reading disorders. In R. N. Malatesha & P. G. Aaron (Eds.), Reading disorders: Varieties and treatments (pp. 405-432). New York: Academic Press.

Spreen, O. (1988). Learning disabled children growing up: A follow-up into adulthood. New York: Oxford University Press.

Spruill, J. (1991). A comparison of the Wechsler Adult Intelligence Scale-Revised with the Stanford-Binet Intelligence Scale (4th Edition) for mentally retarded adults. Psychological Assessment, 3(1), 133-135.

Stanovich, K. (1986). New beginnings, old problems. In S. J. Ceci (Ed.), Handbook of cognitive, social and neurological aspects of learning disabilities (pp. 3-54). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stephenson, W. (1935). Technique of factor analysis. Nature, 136, 297.

Stern, W. (1914). The psychological methods of testing intelligence. Baltimore, MA: Warwick & York.

Strang, J. D., & Rourke, B. P. (1983). Concept-formation/non-verbal reasoning abilities of children who exhibit specific academic problems with arithmetic. Journal of Clinical Child Psychology, 12, 33-39.

Strang, J. D., & Rourke, B. P. (1985). Adaptive behavior of children who exhibit specific arithmetic disabilities and associated neuropsychological abilities and deficits. In B. P. Rourke (Ed.), Neuropsychology of learning disabilities: Essentials of subtype analysis (pp. 302-328). New York: Guilford Press.

Sutter, E. G., Bishop, P. C., & Battin, R. (1986). Factor similarities between traditional psychoeducational and neuropsychological test batteries. Journal of Psychoeducational Assessment, 4, 73-82.

Swanson, H. L. (1988). Learning disabled children's problem solving: Identifying mental processes underlying intelligent performance. Intelligence, 12, 261-278.

Sweeney, J. E., & Rourke, B. P. (1978). Neuropsychological significance of phonetically accurate and phonetically inaccurate spelling errors in younger and older retarded spellers. Brain & Language, 6, 212-225.

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

Terman, L. M. (1916). The measurement of intelligence. Boston, MA: Houghton, Mifflin.

Terman, L. M., & Merrill, M. A. (1937). Measuring Intelligence: A guide to the administration of the new revised Stanford-Binet tests of intelligence. Boston, MA: Houghton, Mifflin.

Terman, L. M., & Merrill, M. A. (1960). Stanford-Binet Intelligence Scale: Manual for the Third Revision, Form L-M. Boston, MA: Houghton, Mifflin.

Terman, L. M., & Merrill, M. A. (1973). Stanford-Binet Intelligence Scale. Chicago, IL: Riverside Publishing Company.

Thomson, M. E. (1982). The assessment of children with specific reading difficulties (dyslexia) using the British Ability Scales. British Journal of Psychology, 73, 461-478.

Thomson, M. E., & Wilsher, C. (1978). Some aspects of memory in dyslexics and controls. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical aspects of memory (pp. 545-560). New York: Academic Press.

Thompson, J. E. (1984). An empirical versus conceptual model of the Stanford-Binet for interpretive purposes. Unpublished masters degree, University of Alberta, Edmonton, AB, Canada.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986a). The Stanford-Binet Intelligence Scale: Fourth Edition. Guide for administering and scoring. San Diego State University, CA: Riverside Publishing Company.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986b). The Stanford-Binet Intelligence Scale: Fourth Edition. Technical manual. San Diego State University, CA: Riverside Publishing Company.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986c). The Stanford-Binet Intelligence Scale: Fourth Edition. Guide for administering and scoring. San Diego State University, CA: Riverside Publishing Company.

Thorndike, R. M. (1990). Would the real factors of the Stanford-Binet Fourth Edition please come forward? Journal of Psychoeducational Assessment, 8(3), 412-435.

Thorndike, R. M., & Lohman, D. F. (1990). A century of ability testing. Chicago, IL: The Riverside Publishing Company.

United States Office of Education. (1977). Definition and criteria for defining students as learning disabled. Federal Register, 42:250, p. 65083. Washington, DC: U. S. Government Printing Office.

Vallar, G., & Baddeley, A. D. (1982). Short-term forgetting and the articulatory loop. Journal of Experimental Psychology, 34, 53-60.

Valtin, R. (1973). Report of research on dyslexia in children. Paper presented at the Annual Meeting of the International Reading Association, Denver, CO.

van der Vlugt, H. (1991). Neuropsychological validation studies of learning disability subtypes: Verbal, visual spatial, and psychomotor abilities. In B. P. Rourke (Ed.), Neuropsychological validation of learning disability subtypes (pp. 140-159). New York: Guilford Press.

van der Vlugt, H., & Satz, P. (1985). Subgroups and subtypes of learning-disabled and normal children: A cross-cultural replication. In B. P. Rourke (Ed.), Neuropsychology of learning disabilities: Essentials of subtype analysis (pp. 212-227). New York: Guilford Press.

Vance, H., Wallbrown, F. H., & Blaha, J. (1978). Determining WISC-R profiles for reading disabled children. Journal of Learning Disabilities, 11, 657-661.

Vellutino, F. R. (1979). Dyslexia: Theory and research. Cambridge, MA: MIT Press.

Vernon, P. E. (1961). The structure of human abilities. London, England: Methuen.

Vernon, P. E. (1987). The demise of the Stanford-Binet Scale. Canadian Psychology, 28,(3) 251-258.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58, 236-244.

Wechsler, D. (1939). The measurement of adult intelligence. Baltimore, MA: Williams & Wilkens.

Wechsler, D. (1949). Manual for the Wechsler Intelligence Scale for Children. New York: Psychological Corporation.

Wechsler, D. (1955). Manual for the Wechsler Adult Intelligence Scale. New York: Psychological Corporation.

Wechsler, D. (1967). Manual for the Wechsler Preschool and Primary Scale of Intelligence. New York: Psychological Corporation.

Wechsler, D. (1974). Manual for the Wechsler Intelligence Scale for Children-Revised. New York: Psychological Corporation.

Wechsler, D. (1981). Manual for the Wechsler Adult Intelligence Scale-Revised. New York: Psychological Corporation.

Wechsler, D. (1989). Manual for the Wechsler Preschool and Primary Scale of Intelligence-Revised. New York: Psychological Corporation.

Wechsler, D. (1991). Manual for the Wechsler Intelligence Scale for Children-Third Edition. San Antonio, TX: Psychological Corporation.

Wepman, J. M. (1960). Auditory discrimination, speech, and reading. Elementary School Journal, 9, 325-333.

Wilde, W. D., & Sillito, M. T. (1986). Evaluation report of the challenge program, County of Strathcona #20. Edmonton, AB, Canada: Alberta Education.

Wilkinson, G. S. (1993). Wide Range Achievement Test-Three, Administration Manual. Wilmington, DE: Wide Range Inc.

Williams, J. (1991). (Review of Neuropsychological validation of learning disability subtypes). Journal of Psychoeducational Assessment, 9, 369-371.

Willis, W. G. (1988). Neuropsychological diagnosis with children: Actuarial and clinical models. In M. G. Tramontana & S. R. Hooper (Eds.), Assessment issues in child neuropsychology (pp. 93-111). New York: Plenum Press.

Wilson, L. R. (1985). Large scale learning disability identification: The reprieve of a concept. Exceptional Children, 52, 44-51.

Winzer, M. (1993). Children with exceptionalities: A Canadian Persepctive. Scarborough, ON: Prentice Hall.

Wishart, D. (1975). Clustan user manual. London, UK: Computer Centre, University of London,

Wolf, F. M. (1986). Meta-analysis: Quantitative methods for research synthesis. Beverly Hills, CA: Sage University Press.

Woodcock, R. W. (1977). Woodcock-Johnson Psycho-educational Battery: Technical Report. Allen, TX: DLM Teaching Resources.

Wright, R. E. (1939). A factor analysis of the original Stanford-Binet Scale. Psychometrika, 4, 209-220.

Wright, B. D., & Mead, R. J. (1978). Bical: Calibrating items and scales with the Rasch model. Research Memorandum: No.23a. Chicago, IL: Statistical Laboratory, Department of Education, University of Chicago.

Yerkes, R. M. (1917). The Binet versus the Point Scale Method of measuring intelligence. Journal of Applied Psychology, 1, 111-122.

Zigler, E., & Farber, E. A. (1985). Commonalities between intellectual extremes: Giftedness and mental retardation. In F. D. Horowitz & M. O'Brien (Eds.), The gifted and talented: Developmental perspectives (pp. 387-408). Washington, DC: American Psychological Association.

APPENDIX A

Table 68

4-7 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 5-6-11: Model Two

		Four Cluster Solution				Six Cluster Solution					
Group	n	1	2	3	4	1	2	3	4	5	6
VS/S	26	8	11	7	0	5	10	11	0	0	0
NL	76	0	25	36	15	0	8	21	21	14	12
SL	2	0	0	0	2	0	0	0	0	0	2
MR	0	0	0	0	0	0	0	0	0	0	0
LD	16	0	3	3	10	0	1	1	3	5	6
	120	8	39	46	27	5	19	33	24	19	20

		Five Cluster Solution					Seven Cluster Solution						
Group	n	1	2	3	4	5	1	2	3	4	5	6	7
VS/S	26	5	15	6	0	0	5	10	11	0	0	0	0
NL	76	0	11	35	17	13	0	8	21	19	14	4	10
SL	2	0	0	0	0	2	0	0	0	0	0	0	2
MR	0	0	0	0	0	0	0	0	0	0	0	0	0
LD	16	0	1	3	5	7	0	1	1	3	5	3	3
	120	5	27	44	22	22	5	19	33	22	19	7	15

Note: VS/S=Very Superior/Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; LD=Learning Disabled.

Table 69

4-7 Cluster terative Partitioning and a priori Groups for the Subsample Age 5-6-11: Model Three

Group		Four Cluster Solution				Six Cluster Solution					
Group	n	1	2	3	4	1	2	3	4	5	6
VS/S	26	8	11	7	0	5	10	11	0	0	0
NL	76	0	25	36	15	0	8	21	21	14	12
SL	2	0	0	0	2	0	0	0	0	0	2
MR	0	0	0	0	0	0	0	0	0	0	0
SLD	5	0	1	1	3	0	0	0	1	3	1
ALD	3	0	0	0	3	0	0	0	0	1	2
RLD	4	0	2	1	1	0	1	1	0	1	1
R/SLD	1	0	0	1	0	0	0	0	1	0	0
R/ALD	1	0	0	0	1	0	0	0	1	0	0
S/ALD	1	0	0	0	1	0	0	0	0	0	1
R/S/ALD	1	0	0	0	1	0	0	0	0	0	1
	120	8	39	46	27	5	19	33	24	19	20

		Five Cluster Solution					Seven Cluster Solution						
Group	n	1	2	3	4	5	1	2	3	4	5	6	7
VS/S	26	5	15	6	0	0	5	10	11	0	0	0	0
NL	76	0	11	35	17	13	0	8	21	19	14	4	10
SL	2	0	0	0	0	2	0	0	0	0	0	0	2
MR	0	0	0	0	0	0	0	0	0	0	0	0	0
SLD	5	0	0	1	3	1	0	0	0	1	3	0	1
ALD	3	0	0	0	1	2	0	0	0	0	1	1	1
RLD	4	0	1	1	1	1	0	1	1	0	1	1	0
R/SLD	1	0	0	1	0	0	0	0	0	1	0	0	0
R/ALD	1	0	0	0	0	1	0	0	0	1	0	0	0
S/ALD	1	0	0	0	0	1	0	0	0	0	0	1	0
R/S/ALD	1	0	0	0	0	1	0	0	0	0	0	0	1
	120	5	27	44	22	22	5	19	33	22	19	7	15

Note: VS/S=Very Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; SLD=Spelling Disabled; ALD=Arithmetic Disabled; RLD=Reading Disabled; R/SLD=Reading/Spelling Disabled; R/ALD=Reading/Arithmetic Disabled; S/ALD=Spelling/Arithmetic Disabled; R/S/ALD=Reading/Spelling/Arithmetic Disabled.

Table 70

4-7 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 5-6-11: Model Four

		Four Cluster Solution				Six Cluster Solution					
Group	n	1	2	3	4	1	2	3	4	5	6
VS/S	26	8	11	7	0	5	10	11	0	0	0
NL	82	0	26	37	19	0	8	21	22	18	13
SL	2	0	0	0	2	0	0	0	0	0	2
MR	0	0	0	0	0	0	0	0	0	0	0
RLD	0	0	0	0	0	0	0	0	0	0	0
RLD/AN	5	0	2	2	1	0	1	1	1	1	1
R/ALD	2	0	0	0	2	0	0	0	1	0	1
ALD/RN	3	0	0	0	3	0	0	0	0	0	3
	120	8	39	46	27	5	19	33	24	19	20

		Five Cluster Solution					Seven Cluster Solution						
Group	n	1	2	3	4	5	1	2	3	4	5	6	7
VS/S	26	5	15	6	0	0	5	10	11	0	0	0	0
NL	82	0	11	36	21	14	0	8	21	20	18	4	11
SL	2	0	0	0	0	2	0	0	0	0	0	0	2
MR	0	0	0	0	0	0	0	0	0	0	0	0	0
RLD	0	0	0	0	0	0	0	0	0	0	0	0	0
RLD/AN	5	0	1	2	1	1	0	1	1	1	1	1	0
R/ALD	2	0	0	0	0	2	0	0	0	1	0	0	1
ALD/RN	3	0	0	0	0	3	0	0	0	0	0	2	1
	120	5	27	44	22	22	5	19	33	22	19	7	15

Note: VS/S=Very Superior/Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; RLD=Reading Disabled; RLD/AN=Reading LD with Math Normal; R/ALD=Reading/Arithmetic Disabled; ALD/RN=Arithmetic Disabled/Reading Normal.

Table 71

4-7 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 7-11-11: Model One

		Four Cluster Solution				Six Cluster Solution					
Group	n	1	2	3	4	1	2	3	4	5	6
VS	9	9	0	0	0	9	0	0	0	0	0
S	23	23	0	0	0	15	8	0	0	0	0
HA	46	20	26	0	0	18	24	0	4	0	0
A	168	0	65	103	0	0	16	76	50	26	0
LA	33	0	0	17	16	0	0	1	0	32	0
SL	15	0	0	0	15	0	0	0	0	3	12
MR	7	0	0	0	7	0	0	0	0	0	7
	301	52	91	120	38	42	48	77	54	61	19

		Five Cluster Solution					Seven Cluster Solution						
Group	n	1	2	3	4	5	1	2	3	4	5	6	7
VS	9	9	0	0	0	0	9	0	0	0	0	0	0
S	23	23	0	0	0	0	5	10	8	0	0	0	0
HA	46	22	1	23	0	0	0	26	17	3	0	0	0
A	168	0	87	48	33	0	0	7	8	52	76	25	0
LA	33	0	3	0	30	0	0	0	0	0	2	31	0
SL	15	0	0	0	0	15	0	0	0	0	0	3	12
MR	7	0	0	0	0	7	0	0	0	0	0	0	7
	301	54	91	71	63	22	14	43	33	55	78	59	19

Note: VS=Very Superior; S=Superior; HA=High Average; A=Average; LA=Low Average; SL=Slow Learner; MR=Mentally Retarded.

Table 72

4-7 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 7-11-11: Model Two

		Four Cluster Solution				Six Cluster Solution					
Group	n	1	2	3	4	1	2	3	4	5	6
VS/S	28	28	0	0	0	21	7	0	0	0	0
NL	168	17	63	78	10	15	27	50	38	38	0
SL	15	0	0	0	15	0	0	0	0	3	12
MR	7	0	0	0	7	0	0	0	0	0	7
LD	83	7	28	42	6	6	14	27	16	20	0
	301	52	91	120	38	42	48	77	54	61	19

		Five Cluster Solution					Seven Cluster Solution						
Group	n	1	2	3	4	5	1	2	3	4	5	6	7
VS/S	28	28	0	0	0	0	13	8	7	0	0	0	0
NL	168	21	59	47	41	0	0	28	16	50	37	37	0
SL	15	0	0	0	0	15	0	0	0	0	0	3	12
MR	7	0	0	0	0	7	0	0	0	0	0	0	7
LD	83	5	3	[illegible]	22	0	1	7	10	28	18	19	0
	301	54	[illegible]	7	63	22	14	43	33	78	55	59	19

Note: VS/S=Very Superior/Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; LD=Learning Disabled.

Table 73

4-7 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 7-11-11; Model Three

		Four Cluster Solution				Six Cluster Solution					
Group	n	1	2	3	4	1	2	3	4	5	6
VS/S	28	28	0	0	0	21	7	0	0	0	0
NL	168	17	63	78	10	15	27	50	38	38	0
SL	15	0	0	0	15	0	0	0	0	3	12
MR	7	0	0	0	7	0	0	0	0	0	7
SLD	15	3	6	6	0	3	4	6	1	1	0
ALD	14	0	4	9	1	0	2	5	3	4	0
RLD	7	0	4	3	0	0	2	1	2	2	0
R/SLD	22	1	8	12	1	1	3	8	5	5	0
R/ALD	4	0	0	2	2	0	0	1	1	2	0
S/ALD	9	1	4	3	1	1	2	3	2	1	0
R/S/ALD	12	2	2	7	1	1	1	3	2	5	0
	301	52	91	120	38	42	48	77	54	61	19

		Five Cluster Solution					Seven Cluster Solution						
Group	n	1	2	3	4	5	1	2	3	4	5	6	7
VS/S	28	28	0	0	0	0	13	8	7	0	0	0	0
NL	168	21	59	47	41	0	0	28	16	50	37	37	0
SL	15	0	0	0	0	15	0	0	0	0	0	3	12
MR	7	0	0	0	0	7	0	0	0	0	0	0	7
SLD	15	2	6	6	1	0	1	2	3	6	2	1	0
ALD	14	0	7	4	3	0	0	0	2	6	3	3	0
RLD	7	0	2	3	2	0	0	2	0	1	2	2	0
R/SLD	22	1	8	8	5	0	0	1	2	8	6	5	0
R/ALD	4	0	1	0	3	0	0	0	0	1	1	2	0
S/ALD	9	1	3	3	2	0	0	1	2	3	2	1	0
R/S/ALD	12	1	5	0	6	0	0	1	1	3	2	5	0
	301	54	91	71	63	22	14	43	33	78	55	59	19

Note: VS/S=Very Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; SLD=Spelling Disabled; ALD=Arithmetic Disabled; RLD=Reading Disabled; R/SLD=Reading/Spelling Disabled; R/ALD=Reading/Arithmetic Disabled; S/ALD=Spelling/Arithmetic Disabled; R/S/ALD=Reading/Spelling/Arithmetic Disabled.

Table 74

4-7 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 7-11-11: Model Four

		Four Cluster Solution				Six Cluster Solution					
Group	n	1	2	3	4	1	2	3	4	5	6
VS/S	30	30	0	0	0	23	7	0	0	0	0
NL	187	19	71	84	13	17	32	55	42	41	0
SL	15	0	0	0	15	0	0	0	0	3	12
MR	7	0	0	0	7	0	0	0	0	0	7
RLD	8	1	4	3	0	1	2	1	1	3	0
RLD/AN	24	0	9	14	1	0	3	9	6	6	0
R/ALD	16	2	3	9	2	1	1	5	4	5	0
ALD/RN	14	0	4	10	0	0	3	7	1	3	0
	301	52	91	120	38	42	48	77	54	61	19

		Five Cluster Solution					Seven Cluster Solution						
Group	n	1	2	3	4	5	1	2	3	4	5	6	7
VS/S	30	30	0	0	0	0	14	9	7	0	0	0	0
NL	187	22	64	56	45	0	0	30	20	55	42	40	0
SL	15	0	0	0	0	15	0	0	0	0	0	3	12
MR	7	0	0	0	0	7	0	0	0	0	0	0	7
RLD	8	1	2	2	3	0	0	2	0	1	2	3	0
RLD/AN	24	0	9	9	6	0	0	1	2	9	6	6	0
R/ALD	16	1	7	1	7	0	0	1	1	5	4	5	0
ALD/RN	14	0	9	3	2	0	0	0	3	8	1	2	0
	301	54	91	71	63	22	14	43	33	78	55	59	19

Note: VS/S=Very Superior/Superior: NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; RLD=Reading Disabled; RLD/AN=Reading LD with Math Normal; R/ALD=Reading/Arithmetic Disabled; ALD/RN=Arithmetic Disabled/Reading Normal.

Table 75

4-7 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 12-23-11: Model One

		Four Cluster Solution				Six Cluster Solution					
Group	n	1	2	3	4	1	2	3	4	5	6
VS	2	2	0	0	0	2	0	0	0	0	0
S	14	14	0	0	0	14	0	0	0	0	0
HA	28	26	2	0	0	24	4	0	0	0	0
A	78	4	56	18	0	2	31	29	11	5	0
LA	29	0	0	29	0	0	0	0	16	13	0
SL	19	0	0	9	10	0	0	0	7	6	6
MR	16	0	0	0	16	0	0	0	0	0	16
	186	46	58	56	26	42	35	29	34	24	22

		Five Cluster Solution					Seven Cluster Solution						
Group	n	1	2	3	4	5	1	2	3	4	5	6	7
VS	2	2	0	0	0	0	2	0	0	0	0	0	0
S	14	14	0	0	0	0	14	0	0	8	0	0	0
HA	28	26	2	0	0	0	24	4	0	0	0	0	0
A	78	5	52	15	6	0	2	31	29	11	5	0	0
LA	29	0	0	17	12	0	0	0	0	16	13	0	0
SL	19	0	0	6	6	7	0	0	0	4	4	11	0
MR	16	0	0	0	0	16	0	0	0	0	0	11	5
	186	47	54	38	24	23	42	35	29	31	22	22	5

Note: VS=Very Superior; S=Superior; HA=High Average; A=Average; LA=Low Average; SL=Slow Learner; MR=Mentally Retarded.

Table 76

4-7 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 12-23-11: Model Two

		Four Cluster Solution				Six Cluster Solution					
Group	n	1	2	3	4	1	2	3	4	5	6
VS/S	16	16	0	0	0	16	0	0	0	0	0
NL	99	28	41	32	0	23	24	21	22	9	0
SL	19	0	0	9	10	0	0	0	7	6	6
MR	16	0	0	0	16	0	0	0	0	0	16
LD	36	4	17	15	0	3	11	8	5	9	0
	186	46	58	56	26	42	35	29	34	24	22

		Five Cluster Solution					Seven Cluster Solution						
Group	n	1	2	3	4	5	1	2	3	4	5	6	7
VS/S	16	16	0	0	0	0	16	0	0	0	0	0	0
NL	99	27	38	25	9	0	23	24	21	22	9	0	0
SL	19	0	0	6	6	7	0	0	0	4	4	11	0
MR	16	0	0	0	0	16	0	0	0	5	0	11	0
LD	36	4	16	7	9	0	3	11	8	5	9	0	5
	186	47	54	38	24	23	42	35	29	31	22	22	5

Note: VS/S=Very Superior/Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; LD=Learning Disabled.

Table 77

4-7 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 12-23-11; Model Three

		Four Cluster Solution				Six Cluster Solution					
Group	n	1	2	3	4	1	2	3	4	5	6
VS/S	16	16	0	0	0	[illegible]	0	0	0	0	0
NL	99	26	41	32	0	[illegible]	4	21	22	9	0
SL	19	0	0	9	10	[illegible]	0	0	7	6	6
MR	16	0	0	0	16	[illegible]	0	0	0	0	16
SLD	8	2	3	3	0	2	3	1	1	1	0
ALD	10	1	5	4	0	0	3	2	3	2	0
RLD	2	1	1	0	0	1	1	0	0	0	0
R/SLD	4	0	2	2	0	0	1	1	0	2	0
R/ALD	0	0	0	0	0	0	0	0	0	0	0
S/ALD	4	0	3	1	0	0	1	2	0	1	0
R/S/ALD	8	0	3	5	0	0	2	4	1	3	0
	186	46	58	56	26	42	35	29	34	24	22

		Five Cluster Solution					Seven Cluster Solution						
Group	n	1	2	3	4	5	1	2	3	4	5	6	7
VS/S	16	16	0	0	0	0	16	0	0	0	0	0	0
NL	99	27	38	25	9	0	23	24	21	22	9	0	0
SL	19	0	0	6	6	7	0	0	0	4	4	11	0
MR	16	0	0	0	0	16	0	0	0	0	0	11	5
SLD	8	2	3	2	1	0	2	3	1	1	1	0	0
ALD	10	1	4	3	2	0	0	3	2	3	2	0	0
RLD	2	1	1	0	0	0	1	1	0	0	0	0	0
R/SLD	4	0	2	0	2	0	0	1	1	0	2	0	0
R/ALD	0	0	0	0	0	0	0	0	0	0	0	0	0
S/ALD	4	0	3	0	1	0	0	1	2	0	1	0	0
R/S/ALD	8	0	3	2	3	0	0	2	2	1	3	0	0
	186	47	54	38	24	23	42	35	29	31	22	22	5

Note: VS/S=Very Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; SLD=Spelling Disabled; ALD=Arithmetic Disabled; RLD=Reading Disabled; R/SLD=Reading/Spelling Disabled; R/ALD=Reading/Arithmetic Disabled; S/ALD=Spelling/Arithmetic Disabled; R/S/ALD=Reading/Spelling/Arithmetic Disabled.

Table 78

4-7 Cluster Iterative Partitioning and a priori Groups for the Subsample Age 12-23 11: Model Four

		Four Cluster Solution				Six Cluster Solution					
Group	n	1	2	3	4	1	2	3	4	5	6
VS/S	16	16	0	0	0	16	0	0	0	0	0
NL	113	28	48	37	0	25	29	24	24	11	0
SL	19	0	0	9	10	0	0	0	7	6	6
MR	16	0	0	0	16	0	0	0	0	0	16
RLD	4	0	3	1	0	0	2	1	0	1	0
RLD/AN	2	1	0	1	0	1	0	0	0	1	0
R/ALD	6	0	3	3	0	0	2	2	0	2	0
ALD/RN	10	1	4	5	0	0	2	2	3	3	0
	186	46	58	56	26	42	35	29	34	24	22

		Five Cluster Solution					Seven Cluster Solution						
Group	n	1	2	3	4	5	1	2	3	4	5	6	7
VS/S	16	16	0	0	0	0	16	0	0	0	0	0	0
NL	113	29	45	28	11	0	25	29	24	24	11	0	0
SL	19	0	0	6	6	7	0	0	0	4	4	11	0
MR	16	0	0	0	0	16	0	0	0	0	0	11	5
RLD	4	0	3	0	1	0	0	2	1	0	1	0	0
RLD/AN	2	1	0	0	1	0	1	0	0	0	1	0	0
R/ALD	6	0	3	1	2	0	0	2	2	0	2	0	0
ALD/RN	10	1	3	3	3	0	0	2	2	3	3	0	0
	186	47	54	38	24	23	42	35	29	31	22	22	5

Note: VS/S=Very Superior/Superior; NL=Normal Learner; SL=Slow Learner; MR=Mentally Retarded; RLD=Reading Disabled; RLD/AN=Reading LD with Math Normal; R/ALD=Reading/Arithmetic Disabled; ALD/RN=Arithmetic Disabled/Reading Normal.