NOTICE

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

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

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

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

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

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

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

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

Canada

THE UNIVERSITY OF ALBERTA

THE STANFORD-BINET INTELLIGENCE SCALE: FOURTH EDITION: A DESCRIPTIVE STUDY

by

RIVVY MELOFF

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

IN COUNSELLING PSYCHOLOGY

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

FALL 1987

ISBN 0-315-41094-9

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: RIVVY MELOFF

TITLE OF THESIS: THE STANFORD-BINET INTELLIGENCE SCALE: FOURTH EDITION: A DESCRIPTIVE STUDY.

DEGREE: MASTER OF EDUCATION

YEAR THIS DEGREE GRANTED: FALL 1987

(Signed) Rivvy Meloff

PERMANENT ADDRESS:
71 Quesnell Road
Edmonton, Alberta

DATE: June 26, 1987

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled THE STANFORD-BINET INTELLIGENCE SCALE: FOURTH EDITION: A DESCRIPTIVE STUDY, submitted by RIVVY MELOFF in partial fulfillment of the requirements for the degree of Master of Education.

.....

(Supervisor)

.....

.....

Date: June 26, 1987

ABSTRACT

The Stanford-Binet Intelligence Scale: Fourth Edition (Binet IV) is a revision of the former widely recognized valid and reliable Stanford-Binet Intelligence Scale: Form L-M. This descriptive study undertook to determine whether or not this new test is measuring what the authors have claimed it is capable of accomplishing; the comparison of the Canadian and American data and the discriminative and convergent validity of Binet IV.

The Binet IV, Wide Range Achievement Test-Revised, (WRAT-R), Developmental Test of Visual-Motor Integration, (VMI) and Perception of Ability Scale for Students (PASS) were administered by supervised graduate students to subjects referred to the Education Clinic at the University of Alberta. The 153 subjects ranged in age from three to 27 with the mean age of the population being 10.1. This sample was divided into five designated groups for analysis purposes. These subgroups were classified as learning difficulties (learning disabled): N=72, mentally retarded: N=15, pre-school: N=9, gifted: N=22 and parental interest (normal): N=35. Due to the very small number of pre-schoolers, this subgroup was eliminated from discussion in the study.

The SPSS/PC statistical software package was used in this study to generate t-tests, Anovas, correlations and crosstabulations for analytic purposes.

Results of the study indicate similar mean scores on Binet IV between American and Canadian samples but show higher standard deviation scores for the Canadian sample. Significant differences between the two populations are reported in eight of the twelve individual subtests, Verbal Reasoning IQ and the Composite IQ scores.

Results of this study confirmed the author's contention that Binet IV is able to differentiate learning disabled from mentally retarded populations; and gifted from normal populations.

Significantly high correlations between Binet IV Composite, Area Scores and individual subtests and WRAT-R are reported. More moderate, yet significant correlations are reported between Binet IV and the VMI and Binet IV and PASS. These results demonstrate the convergent validity of Binet IV.

Administrative and interpretive strengths and weaknesses of Binet IV were discussed in light of refinements deemed necessary for improving this revision of Form L-M.

Results of this study were discussed in relation to the findings of past researchers. Implications for future research and educational practice were discussed.

ACKNOWLEDGEMENTS

I am greatly indebted to Dr. Henry Janzen for his considerable expenditure of time and effort in the supervision of this study. Without his continuous positive support and encouragement this project might never have been completed.

I would like to thank Dr. Fred Boersma and Dr. Robert Silverman, committee members for their patience and encouragement throughout this endeavor. I would also like to thank Dr. Robert Silverman for his time and assistance in helping me maneuver through the unfamiliar territory of statistics and discover the world of computers.

An expresssion of gratitude goes to all the faculty and staff of the Educational Psychology Department for their help and encouragement throughout my graduate program. Special thanks go to Dr. George Fitzsimmons, my academic advisor, Mrs. Aggie Ganchev, Mrs Bobby McLaughlin and Mrs. Deb Black, all of whom were particularly helpful and supportive.

I would like to acknowledge the support and assistance of my three children, Lauren, Rob and Liann who have been patient participants throughout my academic career.

And last, but not least I would like to thank Bill Meloff for believing in me and for helping me see this program through to completion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

For hundreds of years man has been fascinated by the noticeable differences in mental capacity that has existed among individuals in society (Thompson, 1984). The concept that individuals possess different levels or degrees of intelligence had its origins in ancient philosophy. Numerous writings predating the birth of Christ, such as those of the Greek philosophers Plato and Aristotle, include discussions of human differences. (Walsh and Betz, 1985).

Ideas relating to intelligence remained a philosophical issue until the late nineteenth century when psychologists began the systematic investigation of intelligence. The Industrial Revolution was the catalyst for the development of more specialized and complex societal systems. In turn, job definitions became more specific, demanding a better understanding of an individual's capacities to perform work in more precise ways.

Because of the exigency of these pragmatic needs of society, psychologists began to systematically utilize scientific methods in an attempt to measure individual differences among people. Prior to the contributions of many theoretical and practicing psychologists in the early nineteen hundreds, the concept of intelligence as it is understood worldwide today was unknown. Thus, the change in focus began unfolding. From its initial prescientific and philosophical roots, the study of intelligence changed drastically.

Perhaps the most revolutionary contribution of all the theorists of their time was that of Binet and his young associate Simon. In 1905, Alfred Binet and Theodore Simon, commissioned by the French Ministry of Education, developed an objective test to differentiate mentally retarded from normal children. The development of this test is

commonly cited as the major cornerstone in the history of intellectual assessment. (Thorndike, 1975, in Thompson, 1984, p. 1). Scores of individual and group tests and procedures used in the measurement of mental abilities are to a certain degree descendents of the original 1905 Binet test (Thondike, 1975, cited in Thompson, 1984).

The Binet-Simon tests, considered by many to be the first practically useful intelligence tests, attracted wide attention and were soon translated and adapted for use in various countries throughout the world.

In 1916, L.M. Terman, a psychologist at Stanford University revised, expanded and standardized the Binet-Simon test for use in the U.S. Terman divided the child's mental age by his chronological age, multiplied it by one hundred and obtained an intelligent quotient (IQ). A subsequent revision of the Stanford-Binet conducted by Terman and Merrill in 1937 produced two alternate forms of the test - L and M. The third revision occurred in 1960 resulting in the single form (L-M). In 1972 a new standardization of Form L-M resulted. All these revisions of the Binet by Terman resulted in the Stanford-Binet Intelligence Scale Form L-M, considered to be the most widely used intelligence test in the world (Walsh and Betz, 1985).

After twenty-six years of extensive testing and research conducted with Form L-M of the Binet, a revision has occurred resulting in the Stanford-Binet Intelligence Scale: Fourth Edition. The Fourth edition- or Binet IV as it will be referred to in this study is a well packaged, attractive looking new test with many new features to attract potential users. The area of assessment and evaluation has always been of concern to practitioners in the field. It is not surprising then that the new intelligence tests as well as revisions of long-standing existing tests are always in the throes of discussion by psychologists. With the latest revision of the Binet being available for use with clinical and school populations in June, 1986, the necessity for a descriptive study became evident.

Although a technical manual is available, little else exists for Binet IV as a psychometric measure. At this time, there is a great need for empirical data to support the use of the test as a major new instrument. Not only is there an interest in what changes have occurred since Form L-M, but substantial support is required to state exactly what this new test is measuring. One would expect changes not only in content and format from the previous revision, but also that different processes are being tapped by the addition of new tests. Again, the technical manual does not provide this kind of information for its users. The authors of the revision do make numerous claims about the capacities of this new test. One such statement is that Binet IV is able to discriminate between the three exceptional populations of the learning disabled, mentally retarded and gifted from "normal" populations. The only information we have in this area is from the limited evidence presented in the manual. This is but one of the many questions that abound regarding Binet IV.

In order to answer some of these questions regarding Binet IV, this study was undertaken. The new test was initially introduced in a graduate level assessment course during the Spring/Summer sessions of 1986, at the University of Alberta's Educational Psychology department. Test reports were generated by that particular group of supervised graduate students, as well as reports submitted by graduate students enrolled in the Fall/Winter session of the same assessment course. Data were collected from these reports to address three major questions which will be the focus of this study.

QUESTION 1

Are there differences between American and Canadian data on Binet IV? Do scores from this study look similar to the manual's standardization group on Binet IV 'normals'? Are we to assume from the manual's reporting of five thousand American children that our Canadian children's performances within the same age ranges will be

the same or very similar to our American counterparts? Since we have become so accustomed to utilizing American tests of intelligence, achievement, interest, aptitude, etc. is it always in our best interests to assume that Canadian children will perform according to the same patterns reported in the manual based on the American standardized sample?

To summarize question one - Are there similarities and/or differences between Canadian data from this study with American data reported in the standardized group from Binet IV's technical manual?

QUESTION 2

Does the Binet IV and other measures discriminate among special populations? In this study three exceptional populations - learning disabled, mentally retarded and the gifted will be studied to see if in fact significant differences occur between their performances. The learning disabled subgroup will be compared to 'normal' groups and mentally retarded subjects on Binet IV in terms of Composite IQ scores, Area Scores, individual subtest scores as well as scores from the Wide Range Achievement Test- Revised (WRAT-R), Developmental Test of Visual-Motor Integration (VMI), and the Perception of Ability Scale for Students (PASS).

Similar comparisons will be made between gifted groups and 'normal' groups. Various statistical tests will be applied to determine if there are any significant differences between groups, and whether or not performances on Binet IV and the other instruments will discriminate among these special populations. In other words, does the Binet IV have discriminative validity with regard to comparing these special populations?

QUESTION 3

How do other measures (WRAT-R, VMI, and PASS) correlate with Binet IV? Are correlations with these other measures significant? In this study, different aspects of Binet IV will be compared to scores attained by the subjects on the other measures. Composite IQ, Area scores and individual subtest scores will be compared with the scores obtained on the WRAT-R, VMI and PASS. Various statistical operations will be analyzed to determine significant correlations between Binet IV and the other measures. In addition, intercorrelations will be calculated between the subtests, Area Scores and Composite IQ score for this Canadian sample. A comparison will be made between this sample and the results reported for the standardized sample and three exceptional samples discussed in the technical manual.

To summarize, question three will deal with Binet IV and how it correlates with other reliable and valid measures. Underlying this question will be an exploration into the convergent validity of other measures to Binet IV.

II. Thesis Outline

This first chapter provides a statement of purpose for this study. A very short summary of Binet's history and the need for this kind of descriptive study will be discussed.

The second chapter will review in detail the development of the Stanford-Binet Intelligence Scale Form L-M, describe the salient features of The Stanford-Binet Intelligence Scale: Fourth Edition as presented in the technical manual, and review the limited but current literature available for the Binet IV. Hypotheses will be generated and presented.

The third chapter will outline the research design and procedures involved in this study. Procedures, sample characteristics, tests administered, analysis of the data and limitations of the study will be discussed in detail.

The results of the study will be presented in chapter four. Sample characteristics, clinic subgroups and descriptive statistics of all the tests will be included in the results. The proposed hypotheses of chapter two will be discussed at length in light of the information generated from the statistical analysis. Tables and figures will be representative of t-Tests, Anovas, Correlations and Crosstabulations analyzing the total sample as well as analysis of the four clinic subgroups. Intraindividual differences will be reported for this study as well as interindividual comparisons with the American data from the technical manual.

Chapter five will discuss the results of the study that were presented in chapter four and will make recommendations for future research. References and appendices will follow this chapter.

CHAPTER II

REVIEW OF THE LITERATURE

History of the Binet

In 1905 Alfred Binet was commissioned to develop a test by the French Minister of Public Instruction. The rationale for developing the test was to differentiate mentally retarded children from normal children within the French school system. Binet worked with V. Henri and later Theodore Simon. Binet and Simon collaborated in devising thirty objective tests. They focused "on the execution of simple commands, coordination, recognition, verbal knowledge, definitions, picture recognition, suggestibility and completion of sentences." (Sattler, 1982, p. 99). These tests, since termed the 1905 scale, were developed to measure intellectual capacities rather than to treat mentally retarded or brain damaged children. This scale was ethnically homogeneous (French children), but obviously not statistically representative. The sample was only fifty children. (Kaplan & Saccuzzo, cited in French, 1986).

In 1908, Binet and Simon's scale changed such that tests were now grouped according to age levels. This revision of the 1905 scale introduced the concept of mental age. Placing the tests into specific age groups, "represented the most significant advance over the 1905 scale" (Pinter, 1931, cited in Sattler, 1982, p. 99). The 1908 scale generated its standardization norm from a sample of approximately two hundred children. A later revision in 1911, introduced such new testing concepts as "general mental ability," mental age," "age scale," and "basal age/testing ceiling." (French, 1986, p. 65).

A significant revision of the Binet-Simon scale appeared in 1916 when Lewis M. Terman "Americanized" the scale. Terman revised and expanded the Binet-Simon scale while at Stanford University, hence the test's new name - The Stanford-Binet Intelligence Scale. Terman and H.G. Childs collaborated on four years of revision,

producing a comprehensive standardization of Goddard's earlier work. This 1916 scale utilized Stern's development of the mental quotient derived by the division of a person's mental age by their chronological age. Terman coined the term intelligence quotient from this formula simply by divicing the mental age by the chronological age, and multiplying by 100.

In the manual for the third revision, The Stanford-Binet Intelligence Scale, "attempted to provide standards of intellectual performance for average American born children from age three to young adulthood . . . Tests were arranged in order of difficulty by age levels. . .The intellectual ability of an individual determined by his performance on the scale was judged by comparison with the standards of performance for normal children of different ages." (Terman & Merrill, 1973, p. 5)

Terman increased Binet's original fifty-four tests to ninety tests. Tests that had inadequate discriminative value were not included. Approximately one thousand children and four hundred adults were used in the standardization sample and conscious efforts were made to derive a representative sample of the U.S. population. With this revision came specific instructions for the administration and scoring of each of the ninety tests.

> Although affording a satisfactorily valid and reliable measure over a fairly wide intermediate range, abilities below the mental age of four and at the adult levels were very inadequately sampled. Certain tests of the scale were found to have low validity, instructions for administering and scoring were still lacking in the precision necessary to insure objectivity and comparability of results and no alternative form was available. (Terman & Merrill, 1973, p. 5).

These factors prompted a second revision of the Stanford Scale. The 1937 revision incorporated the extensive and comprehensive results of a ten year standardization task by Terman and Merrill. This revision included more items; more performance tests for earlier levels; greater variety of tests at preschool and adult levels; better standardization;

and two forms (L and M) were now available. Other changes reflected improvements in memory tests and semantic improvements in the formulation of the questions. According to Sattler (1982, p.105)

> The scales in [illegible]vision had excellent reliability ranging fr[illegible]m .98 [illegible]those with IQ's below 70 to .90 for those w[illegible] IQ's above 129 and acceptable validity (correlations of .40 to .50 with school success).

Although this revision met with positive results from reviewers, critics still echoed concerns. Verbal material was too heavily weighted; discontent with the age-scale format was expressed; dissatisfaction with procedures involving attainment of basal and ceiling levels; placement of items; too much emphasis on rote memory; insufficient measurement of "g"; and concerns regarding adult usage.

These criticisms did not go unheeded by Terman and Merrill and even though great strides had been accomplished with their 1937 revision, more research gave rise to the 1960 revision. Forms L and M were combined into one form - Form L-M. Essentially very few intrinsic changes were apparent from the 1937 revision. The major impact of this 1960 revision was the usage of the Pinneau tables from the conventional ratio IQ tables that had been used previously. Pinneau's tables introduced Deviation IQ scores for individuals aged two to eighteen. Deviation IQ is defined as a standard score having a mean of 100 and standard deviation of 16. Other minor changes included using the age of eighteen as a ceiling level rather than the previously used age of sixteen, and greater clarification of scoring principles.

Critics of this revision still found the L-M form to be too heavily weighted with verbal materials and the ceiling still was detrimental to very superior students. Specific subtest criticisms suggested that abstract verbal tests were present at a level that was too low and that tests tapping rote memory suggested too high a level.

In 1972, the norms for the Stanford-Binet Intelligence Scale - Form L-M were revised again, this time including white and non-white children in the standardization

group. No changes were made regarding the use of the Deviation IQ, thus maintaining the ability to make comparisons between the 1960 and 1972 norms. It has been found that the 1972 norms most often produce lower IQ scores than the 1960 norms (Sattler, 1982). These differences are related to both the child's age and ability level. The greatest fluctuations occur with younger children (aged two - five) at all ability levels and with children above the age of five with at least average or above average ability. Sattler in his discussion of the Stanford-Binet succinctly summarizes Form L-M in all of its growth, rebirth and changes laying the groundwork for the new Stanford-Binet's Fourth Edition. He states:

> The 1937 and 1960 forms of the Stanford-Binet have proved to be extremely reliable and valid instruments. However, the Stanford-Binet like any measuring instrument, is far from perfect. The scales have been criticized for (a) placing too heavy emphasis on verbal and rote memory tests, (b) providing too few tests of 'g', (c) providing only one score (the IQ) to represent the complex nature of cognitive functions, (d) failing to measure creative abilities, and (e) being unsuitable for testing adults. Technical criticisms of all forms include the cumbersomeness of the age-scale format, scoring and administration difficulties, and the low ceiling for gifted adolescents. In the 1960 form, revision procedures were found to be inadequate, especially with respect to the construction of the Deviation IQs and the determination of difficulty levels for the tests. The Binet-Simon scale was influential in stimulating the development of clinical psychology in the United States and in many other countries. The scales demonstrated that mental measurement was possible, and by so doing, led to the development of many other types of tests and to an acceptance by the public of testing. The Binet-Simon scales and their successors have had an important impact on Western society. (Sattler, 1982, p. 112)

The Stanford-Binet Intelligence Scale: Fourth Edition

In the last twenty-six years considerable changes have occurred in all areas of society in the United States and Canada. These social, political, economic and cultural

changes, together with a redirection and refocus towards cognitive psychology, inspired the careful scrutiny of the Stanford-Binet Intelligence Scale (Form L-M) and the resultant development of the Stanford-Binet, Fourth Edition in 1986 by R. L. Thorndike, E.P. Hagen and J. M. Sattler. For the purposes of this study this new test will be referred to as Binet IV, and the former Binet Intelligence Scale (Form L-M) as Form L-M.

Form L-M has contributed to the field of intelligence testing by its two unique features as an intelligence test. One is its adaptive-testing format - that is, no subject is required to attempt all items, and all subjects who are the same age chronologically do not necessarily respond in the same way to identical tasks. The other major contribution of Form L-M is that, according to the authors, "the Stanford-Binet Intelligence Scale provides a continuous scale for appraising cognitive development from age two to adult." (Thorndike, Hagen & Sattler, 1986A, p. 8).

These two features have been maintained in Binet IV. The variety of content and tasks is also retained but a substantial change is the absence of the age-scale format. Taking the place of the former framework, the Binet IV groups items of the same type in order to evaluate an individual's cognitive functioning more aptly. Other differences that are noted between Form L-M and Binet IV include the following:

1. Subtests:

Items are grouped into fifteen subtests. The examiner is never required to give all fifteen to any one subject. "However, at least eight of the following subtests should be given, regardless of age . These are Vocabulary, Bead Memory, Quantitative, Memory for Sentences, Pattern Analysis, Comprehension, Absurdities, Copying, Memory for Digits (For 1+ level entry) and Memory for Objects (For 1+ level entry)." (Janzen, 1986, p.1). Naturally, the more tests that are given to any subject the greater the reliability of the Area and Composite IQ scores.

According to Janzen (1986), certain subtests should be given for exceptional students such as the gifted and learning disabled. Since some subtests are less subject to ceiling effects (problems encountered by Form L-M and WISC-R scales) the following Binet IV subtests are suggested for administering to potential gifted students:

> For two to five year olds:
> Vocabulary, Bead Memory, Quantitative, Memory for Sentences, Pattern Analysis, Comprehension and Absurdities.
>
> For six to ten year olds:
>
> Vocabulary, Bead Memory, Quantitative, Pattern Analysis, Comprehension, Matrices and Number Series
>
> For eleven and up:
>
> Vocabulary, Bead Memory, Quantitative, Matrices, Number Series Equation Building, Paper Folding and Cutting and Verbal Relations. (p.2).

In the identification of learning disabled children where uneven patterns of cognitive development exist, Janzen (1986) points out that it is important that tests chosen consist of a balance of verbal, abstract/visual and memory areas. The following Binet IV tests are suggested for administering to students with learning difficulties.

> For two to five year olds:
>
> Vocabulary, Bead Memory, Memory for Sentences, Pattern Analysis, Quantitative, Comprehension and Copying.
>
> For six to ten year olds:
>
> Vocabulary, Bead Memory, Quantitative, Memory for Sentences or Memory for Digits, Pattern Analysis, Comprehension or Absurdities, Copying and Matrices.
>
> For eleven and up:
>
> Vocabulary, Bead Memory, Quantitative, Memory for Sentences or Memory for Digits, Pattern Analysis, Comprehension, Matrices, Paper Folding and Cutting.(p. 3).

Janzen (1986) outlines the salient features of each of the fifteen subtests which the researcher feels is necessary to include for a better understanding of Binet IV.

1. Vocabulary: measures the subject's ability to use words in association with both concrete and abstract material. An understanding of verbal concepts and an ability to attach meaning to specific words is necessary. Two levels of the vocabulary subtest are included in Binet IV for different age levels.

a. Picture Vocabulary - requires the subject to have sufficient concrete language to express nouns. The subject must be able to distinguish objects from their function. This subtest requires the subject to be able to recognize objects belonging to a particular category.

b. Vocabulary- appears to be the most important subtest on the Binet IV. Vocabulary measures aspects of abstract thought, verbal fluency and development of temporal concepts as well as a meaning-symbol association. Long-term memory for linguistic symbols is a necessary requisite for this subtest.

Routing test for entry utilizes the Vocabulary subtest and chronological age. This procedure locates the entry for a subtest and the level at which to begin that particular subtest. The Vocabulary subtest was chosen as the routing test for three valid reasons: (1) it functioned well across all age groups; (2) it could be graded more steeply with regard to difficulty; and (3) it had the highest correlation with all the other subtests.

Bead Memory: measures the ability to attend and retain information. Both motivation and attention are necessary as well as retention of visual information in short-term memory store. Visual and spatial relations, in addition to spatial orientation are measured. Simultaneous short-term memory functions are tapped.

3. Quantitative: measures one's ability to count one -to-one relationship of numbers to objects. Processes involved in this subtest include visual-motor abilities, short-term visual memory, sequencing of information on a short term basis and eye-hand

coordination. In order to solve the problems, mathematical reasoning, short-term memory, sequencing of steps, language and visualization of linguistic materials are all necessary. Additionally, mental visualization of the problems is involved. One needs to have mastered the concepts of volume and quantity. Processes of inductive reasoning, sequencing and even trial and error-learning are involved with this subtest. Highly abstract analagous forms of reasoning are necessary to be competent in this subtest.

4. Memory For Sentences: measures auditory attention span, visualization and association, auditory memory for language structured stimuli and auditory semantic sequencing. Subjects also require knowledge of words, articulation skills and good auditory reception.

5. Pattern Analysis: measures visual-motor skills, spatial orientation, eye- hand coordination, non-verbal and visual-spatial reasoning. It involves the ability to decipher form perception, part-whole relationships and gestalt perception appropriate for one's chronological age. The subjects must be able to attrack problem solving utilizing a trial and error method when necessary.

6. Comprehension: measures a subject's ability to make sense from verbal and non-verbal material. Through auditory, visual and other sensory modality input, messages must be integrated by the subject utilizing common sense. Reception of information and use of vocabulary in association with concrete stimuli is measured. An individual's social confidence and ability to verbally express himself in a fluent fashion is measured via this subtest. The subject must deal with questions involving social insight, empathy, appropriateness to one's age level - in other words common sense of the individual. There is a need for an alertness to the environment, visualization and the capacity for mental imagery. Long-term memory for language (similar to Vocabulary) is necessary as well as an individual's capacity to shift a mental set.

7. Absurdities: is a subtest measuring one's visual acuity, awareness and verbal reasoning from visual data. Visual reception, gestalt perception and understanding social and physical conventions are all necessary. Being able to see the subtleties and inappropriateness of- for example - writing with a fork, or taking a fish for a walk, or seeing a cat in a cage involves these processes.

8. Memory for Digits:

a. Digits Forward - measures immediate recall and short-term auditory rote memory of non-meaningful information - such as digit sequences. Subjects need the ability to organize and group digits.

b. Digits Backward - measures the same skills required in digits forward but also taps mentally reordering stimuli without losing the memory trace.

9. Copying: measures the subject's ability to manipulate materials in problem-solving. It also involves the integration of visual-motor skills, eye- hand coordination, gestalt perception, figure-ground perception, grasping the concept of matching a measurement of form constancy; and being flexible in trial and error learning.

10. Memory for Objects: measures one's familiarity with and then the sequencing of visual information. It involves short-term memory of meaningful stimuli and the simple recognition of common objects. It is a process of recognition memory, not simply recall.

11. Matrices: is a subtest in which a subject must have the capacity for non-verbal, visual-spatial reasoning. Clear thinking, observation, visual acuity and visual perception are all necessary for positive results in this subtest. It taps abstract visual reasoning, attention to sequence, form perception and part-whole visual-spatial reasoning. This is one of the few tests measuring inductive reasoning and requires the capacity of the subject to shift their mental-set and be able to visualize the particular pattern.

12. Number Series: involves not only counting in sequence but understanding basic rules or principles in visual and spatial arithmetic reasoning. The subject must mentally deal with concepts of numbers and the number line. Knowledge of numerical relations, fractions and multiplication is a prerequisite to successful completion of the items on the subtest. Analytical thinking is the underlying process.

13. Paper Folding and Cutting: is another subtest measuring inductive visual reasoning. It requires analytic thinking, sequencing of visual data and visual-motor sequencing. The subject needs to be able to focus his skills on a specific orientation, spatial relationships, visual closure and gestalt perception.

14. Verbal Relations: measures one's verbal reasoning, associative thinking, and concept formation. It involves thinking in terms of analagous functions, parts or qualities and requires good long-term semantic memory. The subject must have the capacity to shift mental sets, be aware of opposites and have a good basic word knowledge and meaning of words.

15. Equation Building: involves abstract/visual and mathematical reasoning and concept formation. There must be a sound knowledge of basic arithmetic operations and functions and a capacity to mentally reorganize and sequence the numerical data. The subject must have the capacity to shift a mental set, have immediate recall and have a good long-term memory for arithmetic facts and operations.

Janzen (1986) points out that nine tests on the Binet IV have evolved from item types in Form L-M and that six tests are of an original nature. The six essentially new tests include: Quantitative, Number Series, Equation Building, Pattern Analysis, Matrices and Copying. He states that Paper Folding and Cutting was in Form L-M, but in Binet IV it evolves into a visual-spatial reasoning task of escalating complexity.

Items in each subtest are organized in levels and are assigned letters 'A' through 'Y'. Each level consists of two items of approximately equal difficulty.

11. Area Scores :

Subtests are grouped into four area scores.

a. Verbal Reasoning Area: Vocabulary, Comprehension, Absurdities and Verbal Relations

b. Abstract/Visual Reasoning Area: Pattern Analysis, Copying, Matrices Paper Folding and Cutting.

c. Quantitative Reasoning Area: Quantitative, Number Series and Equation Building.

d. Short-Term Memory Area: Bead Memory, Memory for Sentences, Memory for Digits and Memory for Objects.

These four area scores yield a Composite IQ score which is the measure of "g".

To summarize, the Binet IV is comprised of fifteen individual subtests which are grouped into four areas of cognitive ability which in turn yield the Composite IQ score.

111. Theoretical Model of Binet IV:

The theoretical model of Binet IV is based on the hierarchical model of intelligence as depicted in Figure 2.1, p. 9 of the technical manual (Thorndike, Hagen & Sattler, 1986A).

There has been much written criticizing or applauding this particular model. Other models or theories of intelligence have been developed by such theorists as Spearman who proposed a two-factor theory of intelligence with an emphasis on a general factor "g" and one or more specific factors (s). Thorndike on the other hand described three different kinds of intelligence - social, concrete and abstract. In Guilford's (Sattler, 1982, p. 46) three-dimensional 'Structure of Intellect Model', one hundred and twenty possible factors are postulated. Vernon's hierarchical theory emphasizes "g" and below "g" are Verbal-Educational and Spatial-Mechanical group factors, which are further categorized into minor group factors. Many other theorists such as Thurstone, Jensen, Das, Piaget, etc. have discussed at length which model of intelligence is right. But,

this thesis is not involved with debating the issue, but rather is an attempt to describe what the authors consider to be the model of intelligence upon which the Binet IV is based.

1. "g"

"g" has been defined in many different ways by different theorists, but for the purposes of this study is interpreted to mean 'general reasoning'.

In reviewing Wechsler's intelligence definition emphasizing global capacity, Matarazzo (1972, p. 79) discusses the point that functional intelligence is not necessarily the sum of one's abilities to act purposefully, think rationally or to deal effectively with one's environment. He states,

> 1. The ultimate products of intelligent behavior are a function not only of the number of abilities or their quality but also of the way in which they are combined, that is, their configuration.
>
> 2. Factors other than intellectual ability, for example, those of drive and incentive, are involved in intelligent behavior.
>
> 3. Finally, whereas different orders of intelligent behavior may require varying degrees of intellectual ability, an excess of any given ability may add relatively little to the effectiveness of the behavior as a whole. (Matarazzo, 1972, p. 79).

2. Fluid and Crystallized Intelligence

The Cattell-Horn theory presented by Horn (1968 in Matarazzo, p. 55) discusses two broad factors, with each factor representative of a different type of intelligence.

A. Fluid Intelligence "corresponds to and reflects a pattern of neural-physiological and incidental learning influences." Horn (cited in Matarazzo, 1976). It is a natural thing for an individual - what we would call a 'god-given talent'. Cattell (in Matarazzo, 1976) states that, "fluid intelligence is conceived as a general (independent of sensory area) relation-perceiving capacity which is determined by each individual's unique endowment in cortical, neurological connection count development." This theory affirms that fluid intelligence continues to grow within each person until the age of

fourteen, levels off and then drops off dramatically after the age of twenty-two. Horn talks about fluid intelligence as being 'formless', essentially independent of one's academic endeavors and life experiences and can 'flow into' a wide variety of intellectual abilities. With brain damage or disease and advancing age, fluid intelligence declines.

Fluid intelligence is particularly influenced by the physical and psychological health of an individual as well as any biological conditions or diseases incurred by an individual.

B. Crystallized Intelligence is affected strongly by an individual's own cultural, educational and environmental experiences. Although highly related to fluid intelligence, it normally taps a unique component of each individual's performance capacity. That component which reflects material normally taught in school and which manifests itself in ability tests of vocabulary, synonyms, numerical skills, mechanical knowledge, a well-stocked memory, and even habits of logical reasoning. . .The extent to which an individual takes or leaves what he is taught, that is, gains from these cultural experiences, depends on (1) his underlying fluid ability, (2) his years of formal education and (3) his motivation to learn." (Matarazzo, 1972, p. 56).

Given time to grow and develop, one's crystallized intelligence reflects not only his fluid intelligence but also his positive exposure culturally. Crystallized intelligence is not the same as one's scholastic ability. It is a much more complex commodity. Unlike fluid intelligence that declines after age twenty-two, crystallized ability continues flourishing and growing to at least age forty. Horn reiterates this phenomenon by stating,

> Crystallized intelligence is a precipitate out of experience. It results when fluid intelligence is 'mixed' with what can be called 'the intelligence of the culture.' Crystallized intelligence increases with a person's experience and with the education that provides new methods and perspectives for dealing with this experience. (Horn, 1967, p. 23 in Matarazzo, p. 57).

Verbal Reasoning Area subtests and Quantitative Reasoning Area subtests are considered to be measuring Crystallized Intelligence while the Abstract/Visual Reasoning Area subtests would be tapping one's Fluid Intelligence.

3. Short-Term Memory

The third group factor in the hierarchical model is that of short-term memory. One of the major changes in the format of the Binet IV is the importance attached to memory. It is no longer integrated into other subtest areas, but rather is recognized as a unique and important area on its own. This is similar to the WISC-R's third factor - Freedom from Distractibility. According to cognitive theorists regarding the area of memory (Atkinson and Schiffrin, 1968, 1971; Shiffrin and Atkinson 1969, in Hill, 1985) three memory stores exist whereby information can be stored for some period of time. The first of the three stores is called the 'sensory register' where all information entering the system through the senses is initially registered. It is stored in exactly the form in which it was received - much like a copy of the image from the retina of the eye. The sensory register is like a television picture containing a great deal of information but retains the information for a very short period of time. The information is either moved to short-term memory or is not retained in one's system any longer. If we committed all sensory stimuli from the sensory register to short-term store our system would become overloaded. We select what we want to retain in short-term store. A very limited number of items can be kept at any one time. Actually, as soon as additional items are added, some stimuli will have to leave the short-term stores. Essentially items can be kept in short-term memory almost indefinitely as long as the individual concentrates on those particular items. According to Atkinson & Schiffrin, 1979, in Hill, 1985) "the capacity of the short-term store is almost seven times, not always exactly seven units, but rarely less than five or more than nine." From the short-term store either the information is discarded or it becomes part of the long-term memory bank. One's short-term memory affects the individual's selection of what is to

be stored, whether it remains in short-term store on a temporary basis or moves into long-term storage. Long-term store has unlimited capacity and in reality nothing is lost from . When we forget something that is in the long-term store we must then rely upon the techniques and strategies we select to retrieve the information. Thus, a good memory would depend upon the relationship between the way information is labelled as it goes into long-term store and the way it is searched for when the time comes to retrieve it.

4. Verbal Reasoning, Quantitative Reasoning, Abstract/Visual Reasoning

The third level of the hierarchical system includes the three specific factors of Verbal Reasoning, Quantitative Reasoning and Abstract/Visual Reasoning.

It is important to consider the specific factors by looking at each individual subtest. Each subtest has a certain amount of specificity associated with it. According to Janzen, (1986), for all ages, two of the subtests have the lowest specificity of .22. These are Vocabulary and Pattern Analysis. The "g" loadings on these two subtests are .76 and .67 respectively. Bead Memory, depending on the age of the client and the different strategies used, seems to measure different things for different individuals. Copying, similar to Beery's VMI, is very high on specificity because it is measuring a high level of visual/spatial ability. Paper folding and cutting is a subtest which requires an individual to visualize mentally and have a good concept of form constancy. A number of different psychological constructs are associated with this subtest and it too has a high (.69) specificity rating. Most of the other subtests fall into the 40 range for specificity. Detailed information on subtest "g" loadings, group factor loadings and specificity of each subtest for all ages can be found in table 6.2 of the technical manual (Thorndike et al. 1986A).

1V Administration of the Test:

Test administration is approximately one to one and one-half hours for two to eight year olds, and one and one-half hours for eight years and up. Janzen (1986) outlines

how to begin testing in stages. One must first find out the correct age of the client, remembering to add an extra month if fifteen or more days are included with the chronological age. The first test administered is Vocabulary to determine the entry level for the other tests. When the ceiling (three to four failures at two levels) has been reached on Vocabulary, the entry level is located at the back of the protocol booklet. By entering the child's chronological age and the child's highest vocabulary item given, the intersection provides the alphabet letter for the appropriate entry level. The rest of the tests are administered in the order given in the booklet and one begins at the entry level, after the appropriate sample item has been given.

V. Range of IQ Scores:

According to the manual, IQ's can range from approximately 36 to 164. With the increase in the range, Binet IV is suitable for testing moderate retardation through to giftedness.

V1. Standard Age Score (SAS):

Individual subtests have a mean of 50 and a standard deviation of 8. To show significant differences between subtests $\pm$ 8 should be used as a guideline. Composite IQ has remained unchanged from the 1960 version maintaining a mean of 100 and a standard deviation of 16.

V11. Confidence Intervals:

Intervals are reported in tables F.1 and F.2 of the technical manual (Thorndike et al. 1986A). Confidence intervals are reported at the 15% and 5% levels of confidence for individual subtests; between areas by age; and between area and composite score by age.

V111. Standardization of the Sample:

Standardization included five thousand subjects selected on five variables of geographic region, community size, ethnic group, age and gender.

IX. Reliability and Validity:

a. Reliability: for the Binet IV was reported on the basis of internal consistency and test-retest reliability. The most reliable score is the Composite Standard Age Score with reliability indices ranging from .95 to .99 as the age of the subject increases, followed by the Area Scores. Each Area Score's reliability depends on the number of tests administered. Reliability levels increase with more tests given and vary according to age. Reliability indices for Area Scores range from .91 (Short-term memory) to .97 (Verbal Reasoning). Reliabilities for individual subtests are in the .80's and .90's except for Memory for Objects where the range is .66 to .78. Tables 5.1 to 5.3 of the technical manual (Thorndike et al. 1986A) give further information on internal consistency. In looking at the tables, it seems that there is a general tendency for reliabilities to be somewhat higher for the older age group.

Test-retest reliability data was obtained from retesting 112 children aged five and eight. Test-retest correlations for Area Scores ranged from .71 to .88; and .91 for the preschooler's Composite score. The range was lower for the eight year olds - being .51 to .87 for Area Scores and .91 for the Composite. Consult tables 5.4 and 5.5 of the technical manual (Thorndike et al. 1986A) for further information.

b. Validity: Generally, test validity is reported in terms of content, construct and predictive validity. Binet IV reports only construct validity based on three specific areas:

> 1. the internal structure of the test as evidenced by the correlations among the tests and as clarified by factor analytic procedures
>
> 2. the correlation of the test with scores on other tests deemed to measures of the same or a similar construct; and
>
> 3. the performance of groups identified by indices other than Binet IV as presumably high or low on the construct(s) the test is designed to measure. (Thorndike et al. 1986A, p. 52)

When analyzing the internal structure of Binet IV, the manual reports intercorrelations between tests and are reported in Appendix B of the manual (Thorndike et al. 1986A). "g" loadings, specificity and group factors (residual correlation among tests that are assigned to a given content area) are also reported. Most tests received fairly high "g" loadings - the range being from .51 (Memory for Objects) to .79 (Number Series). Group factor loadings were in all cases lower than the general factor loadings, yet substantiation was received for each of the four Area Scores. Specificity of each test was reported by the use of KR-20 values with Copying showing the highest value at .69. Further information can be found in Table 6.2 of the technical manual (Thorndike et al. 1986A).

Correlations with other tests are reported for the Stanford-Binet L-M, WISC-R, WAIS-R, WPPSI, and K-ABC. Correlations range from .56 to .76 for Area Scores and .81 for Binet IV Composite and Form L-M. Correlations between Binet IV's Composite and WISC-R Verbal - .78, WISC-R Performance - .73 and WISC-R Full Scale - .83. For the WPPSI and Binet IV - correlations between the Composite and Verbal - .78, Performance - .71 and Full Scale .80. Binet IV Composite and the WAIS-R - Verbal - .90, Performance - .85 and Full Scale - .91. Binet IV Composite and K-ABC - correlations ranged from .82 to .89. All correlations with other intelligence tests are relatively high (see tables in manual for all correlations).

Eight individual studies were performed and reported comparing the performance of gifted, learning disabled and mentally retarded children on the Binet IV and other individualized intelligence tests.

Two studies of gifted students follow. The first study compared 82 subjects on Binet IV with Form L-M. As anticipated the range of scores was very narrow with standard deviations reaching approximately 9 for both tests. There was a significant difference of 13.5 points between the mean Total score of Form L-M - 135.3 and Binet IV Composite - 121.8. The authors attempt to provide a logical explanation for this

major difference. They believe that in addition to the difference in standardization dates, the emphasis on verbal skills on Form L-M, with few tasks measuring quantitative and abstract/visual reasoning skills gives students with highly developed verbal skills an advantage on Form L-M. Yet, when one examines Table 6.11 of the technical manual (Thorndike et al. 1986A) a verbal reasoning mean of 117.7 as compared to an abstract/visual reasoning mean of 121.8 is reported. Given the author's explanation of the Binet IV and Form L-M differences, these scores do not logically fit with their reasoning. Also given the strong emphasis of the verbal content of Form L-M, one would expect mean scores for Binet IV's Verbal Reasoning area and Form L-M Total IQ score to be much closer than the reported scores of 117.7 and 135.3 respectively.

It is interesting to note that in the second study reporting of gifted children being compared on the results of Binet IV and the WISC-R that the Composite mean and the Full Scale mean are almost identical. Mean scores of 116.3 and 117.7 respectively are reported.

Three studies are cited for designated learning disabled students. A comparison was made between Binet IV scores of learning disabled subjects and their performances on Form L-M, WISC-R and K-ABC. In the first study, the mean Composite score was 79.9 and Form L-M was 76.9. Correlations of Area Scores with Total score of the L-M ranged from .54 for Abstract/Visual Reasoning to .86 for Verbal Reasoning. The authors caution that although these correlations are acceptable, readers must be conscious of the very small sample.

The second study compared Binet IV with the WISC-R. The Composite mean score for learning disabled children was 84.8 and Full Scale mean 87.8. In this study the mean scores of Verbal Reasoning and Quantitative Reasoning plus Short-Term Memory were averaged to be more comparable to WISC-R's verbal scale. When these calculations were completed, the average of the two means was 87.5 and 86.6 for the

three areas, with the verbal scale mean being 85.2 - a difference of 2.3 and 1.4 points respectively. The authors believe the difference between the two instruments is due mainly to the standard deviations used. The largest difference in mean scores for the two scales was between Abstract/Visual Reasoning and the WISC-R Performance Scale with a difference of 4.6 points.

The third study reported compared Binet IV and K-ABC. Binet IV's Composite mean score was 92.5 and the mean K-ABC Mental Processing Composite was 94.2. Correlations ranged from .28 for Abstract/Visual Reasoning and Sequential Processing to .74 for the Composite and Achievement. Although the size of the correlations of the learning disabled subjects was smaller than non-exceptional samples reported earlier the pattern of correlations was very similar.

Three studies using designated mentally retarded subjects were examined comparing performances attained on Binet IV with Form L-M, WISC-R and WAIS-R. The subjects lowest scores were on Short-Term Memory subtests where the mean score was three standard deviations below the population mean. Composite mean compared to Form L-M mean was 50.9 and 49.5 respectively. Correlations were high ranging from .84 for Form L-M and Verbal Reasoning to .91 for Form L-M Total and Binet IV Composite.

In the second study the mean Composite was 66.2 and WISC-R Full Scale mean 67. As in the first study cited, the subjects worst performances were in the Short-Term Memory Area.

The third study was a comparison of scores attained on Binet IV and the WAIS-R. Means differed by 9.3 points with subjects attaining a mean Composite score of 63.8 and 73.1 on WAIS-R Full Scale. Explanation for this significant difference was expected by the authors because:

1. Binet IV has been designed to have a lower floor than WAIS-R.
2. Binet IV has the most recent standardization date; and

3. The WAIS-R tends to yield higher scores than do other intelligence tests. (Thorndike et al. 1986A, p. 81)

To give further evidence on the construct validity of the Binet IV examinees designated as gifted, learning disabled and mentally retarded by their schools were used. Sample numbers were approximately 220 for each group. As expected, the means of the gifted examinees on Area Scores and the Composite were significantly above the means of the standardization group as were the means of the learning disabled and mentally retarded children significantly lower than standardization sample. These results also showed significant differences in higher mean scores of the learning disabled subjects compared to those of the mentally retarded subjects on all Area Scores and the Composite of the Binet IV. It is unfortunate that individual subtest means and standard deviations were not reported for these two groups. The authors feel that enough substantiation has been offered to state that the Binet IV is able to discriminate reliably between learning disabled and mentally retarded subjects.

Current Literature

To date only four articles have been published regarding the Fourth Edition of the Stanford-Binet Intelligence Scale. These articles appeared in September, 1986 and November 1986 editions of an American publication titled, "Communique", National Association of School Psychologists.

Editor, Tom Fagan (1986) reported on a meeting held in the summer of 1986 in Seattle by the National Association of School Psychologist's Executive Board which discussed Binet IV. Concerns were such that a resolution by different parts of the country considered non-usage of the scale. Robert Thorndike, one of the authors was present at the meeting to answer questions regarding the test. The Executive board recommended the following motion:

> The National Association of School Psychologists does not currently recommend the use of the Revised Stanford-Binet for purposes of determining student eligibility and/or placement. This resolution would remain in effect until: (1) the serious problems involving notification of test users regarding table errors are resolved, (2) the distinctions between the original and revised manual (with table corrections) are made in a manner that will assure destruction and non-use of the original manual, and (3) it can be assured that all purchases of the test have copies of the technical manual which has not yet been distributed. (1986).

The essence of the entire resolution was not that the Binet IV was inadequate technically as an intelligence test, but rather certain technicalities unrelated to the test itself needed to be resolved.

The second article in the September issue dealt with more salient issues of the test itself. J.R. Slate (1986) discussed the revision of the Binet as being new but not necessarily an improvement over the Stanford-Binet Form L-M. After attending a workshop on Binet IV, Slate (1986) raised his concerns regarding the instrument. His criticisms include the following points:

1. The technical manual does not include the standard error of measurement score for the Composite IQ, although it does for the individual subtests and Area Scores. However, if one recalls the first appearances of Form L-M, WISC-R and WAIS-R manuals, they did not include this information until later.

2. Slate (1986) questioned the author's statement regarding Binet IV "essentially" measuring "g". He asks what percentage of the variance does "g" actually account for? His interpretation is such that if "g" accounts for at least eighty-two percent variance (the same as the WISC-R) then factor areas of verbal reasoning, abstract/visual reasoning, quantitative reasoning and short-term memory are not being measured. Slate is effectively saying that Binet IV is measuring only "g" and nothing else. However, if one reads the technical manual with a discerning eye to what the authors are saying they discuss "g" in its relationship to percent of the variance. For example, they suggest that for ages two to eleven forty-two percent of the variance is accounted

for by "g". They also discuss each test's specificity - that is how much a factor (such as memory or verbal reasoning etc.) contributes to its own essence. This second criticism does not seem to be particularly valid.

3. In the sampling procedures there is an overselection of college graduates (36.7 %) although the national population is only 19%. The authors were aware of this overrepresentation. This was also carried over into an overrepresentation of managerial/professional occupations. The authors utilized a weighting procedure to adjust for the discrepancy. The sampling process, however, should have taken this point into consideration while sample procedures were being conducted to try and match the sample population with the national population. Slate may have very legitimate concerns with the lack of representativeness not reflecting national norms. Future problems could result particularly in dealing with adults.

4. Slate (1986) questioned why the authors still retained the standard deviation of sixteen when essentially all other intelligence, achievement and psychometric devices use fifteen as a standard deviation. Continuity with Form L-M for comparative purposes was probably one of the main reasons. This is a contentious issue and is discussed in further articles.

5. The cost of Binet IV is almost prohibitive. Slate (1986) has a valid point on this issue. If the authors would like to see this new edition become an important assessment instrument for the future, costs will have to be within reasonable parameters for both trainees as well as professionals in the field.

6. The flexibility that permits examiners to pick and choose which set of subtests to be given to their subjects can have a major effect on IQ scores as well as on important placement decisions. For example, the manual suggests that a six-test battery can give reliable results - but the major question is "which six tests?" How does an examiner decide? The focus of the subtests could very well be either on the subject's strengths or weaknesses, and depending on the referral reason could be the factor swaying the

pendulum in one direction or another regarding an important placement decision. Slate (1986) is very criticial of Binet IV on this point. However, he does acknowledge that this problem is not unique to the Binet IV.

7. A number of subtests do not load on the factors that they are supposed to represent. For example, Bead Memory on the Short-Term Memory Area loads very low, except for the older age group. This suggests that perhaps Bead Memory should not be included in short-term memory, but rather another process is being tapped. Time and use of Binet IV will perhaps clarify Bead Memory's inclusion as a memory scale. Future adjustments could in fact change its definition or remove it entirely from the subtests.

8. According to its authors Binet IV is based on the three-level hierarchical model as discussed earlier. Slate (1986) felt that the factor analytic data does not support this model fully and questions why the theory remains as its basis when a number of subtests in essence do not validate its existence in its present form. This criticism can not be addressed at this time. Use and research of the Binet IV will help to clarify the hierarchical constructs upon which the test is built or else nullify what the authors purport the test's construct basis to be.

J.R Slate (1986) does have some legitimate concerns with Binet IV, although most of them are concerns that can be applied to any intelligence test. However, certain criticisms are of concern and the authors do have a responsibility of attempting to rectify them at some future date. Examiners should certainly be aware of these problems with the Binet IV. When placement decisions are made, caution regarding specific areas of concern as well as confirmation from other instruments should be considered.

In the November issue of "Communique" an update regarding the prior resolution was presented. In essence the Association now fully accepts Binet IV since the

publisher had followed suggestions ensuring that corrected manuals, technical manuals and new protocol forms were widely distributed.

Also in this issue, David W. Osberg (1986) has a rebuttal to J.R. Slate's (1986) article. Osberg (1986) discussed the lack of essential technical data. He feels that now with the second printing of the Guide for Administering and Scoring the Fourth Edition (Thorndike, Hagen & Sattler 1986B) and the publication of the technical manual (Thorndike et al. 1986A) this concern has been eliminated.

He points out that the former criticisms regarding the lack of a Standard Error of Measurement score for the Composite score is not provided in detail in the Administration Guide (Thorndike et al. 1986B) but is printed in the technical manual (Thorndike et al. 1986A).

Slate (1986) was concerned with what Binet IV attributes to "g" at the expense of what the Area Scores measure. Osberg (1986) feels that Slate's argument is based on Kaufmann's "g" in relationship to the WISC-R, but is not in fact based on the amount of variance attributed to "g" for Binet IV as is reported in the technical manual (Thorndike et al. 1986A).

Osberg (1986) deals with the overrepresentation of higher socioeconomic classes and higher occupation levels quite simplistically. His argument is that the overrepresentation was due mainly to the fact that parents from higher socioeconomic categories were more willing to sign permission forms allowing their children to be tested than were parents from lower socioeconomic categories. His reason does not change the issue of overrepresentation. Perhaps for a representative sample to be reported different methods of recruiting candidates should have been considered.

The issue of the standard deviation being 15 or 16 seems like a moot point. It does not seem advantageous for the authors of Binet IV to have kept the standard deviation at 16 other than being able to compare data with Form L-M. It would be an advantage to be able to compare results from Binet IV with other intelligence and achievement tests

whose standard deviations are 15. There certainly seem to be advantages and disadvantages for having kept the standard deviation at 16.

Osberg (1986) discusses the cost of Binet IV and appears to make allowance for the exorbitant price, stating that "each purchaser must determine the value of a particular product." (1986, p. 3). He further explains the fact that the cost has been misrepresented and is in fact commensurate with WISC-R and WPPSI tests combined. At this point in his defense of Binet IV and criticism of Slate's comments it sounded as if Osberg worked for the publishing company. In checking the author's title at the end of the article, it was found that he indeed does work for the Riverside Publishing Company. It is hardly an unbiased defence of any critical statements directed toward the test.

Further points by Osberg (1986) of Binet IV defends the selection of specific tests by examiners who are perhaps unintentionally guilty of biasing the scores. Osberg's arguments are extremely weak - not really addressing the issue on any solid ground. He states, " neither is probably the 'true' score since all test batteries are estimates. Each test or subtest is designed to include specific attributes that differ from those measured by other tests or subtests." (1986, p. 3.). He does not address the whole issue of major placement decisions which, based on choices of selected tests and subtests would certainly have a major impact on a child's entire life. Rather, Osberg gets caught up in the semantics instead of the real issues.

In refuting Dr. Slate's comments about Bead Memory's place on the memory tests, Osberg (1986) points out that the other three memory tests are indicative of sequential short-term memory whereas Bead memory is the only one which taps simultaneous memory functions. It seems that a more in depth explanation of this area would have been more beneficial to his readers.

Osberg (1986) states that the purpose of the hierarchical model was in providing the authors with a conceptual framework for the test construction and development.

The fact that the factor analytic data does not give complete assurance that the hierarchical model fits completely with the test is an area for further research.

It is too early for trained professionals to give either positive or negative feedback about Binet IV. However, when enough time and experience with the test has elapsed perhaps those who administer and interpret the test can give us a more comprehensive understanding of the test than the portrayal presented by the publishing company that obviously has a vested interest in his comments.

As yet, there has been very little published on Binet IV. No research data is available on this test. Many questions have not been dealt with in any depth by the technical manual. Certainly more modern psychometric methods have been used in this current revision, and Binet IV shows promise in examining a broader coverage of an individual in terms of abilities and age. But many questions still persist. Does Binet IV really fit the hierarchical model of intelligence? Does it matter if it does or does not fit this particular model? Do the subtests actually differentiate learning disabled children from "normal" children or mentally retarded children? Would the norms of their standardization sample be the same had a more representative sample been polled? Many more questions need to be asked.

Until the test has been utilized by trained clinicians in the field these questions can not be answered adequately. Research is clearly necessary for us to know more about Binet IV. Certainly the exterior packaging, the less complex administration and its flexibility for a highly competent examiner is making it more attractive to prospective clinicians than was its predecessor. However, Dr. Slate's (1986) statement in the Communique article does focus users concerns when he says that "new is not necessarily better." We need more time and experience with Binet IV to learn all of the nuances of the test. Only with more research in the field can a more complete picture become apparent.

As with any intelligence test, such as the extensive studies done with Form L-M, WISC-R, WAIS-R, etc., much research is necessary to explore in depth what a test really measures and how it can be applied in a diagnostic, clinical or educational manner.

Research Questions and Hypotheses

The revision of a major assessment tool typically elicits considerable interest among practicing psychologists and educators, as well as psychometric theoreticians. The present study addresses those interests. Research hypotheses under three major headings have been formulated after an analysis of Binet IV's manual.

i. Canadian Data vs. American Data on Binet IV

Hypotheses:

1. [illegible]s and standard deviations of the total sample of this study for Binet IV [illegible] Area Scores, and individual subtest scores will be similar to the [illegible]rmal standardization sample.

2. There will be significant correlations between Verbal Reasoning subtests and the Verbal Reasoning IQ score.

3. There will be significant correlations between Abstract/Visual Reasoning subtests and the Abstract/Visual Reasoning IQ score.

4. There will be significant correlations between Quantitative subtests and the Quantitative Reasoning IQ score.

5. There will be significant correlations between Short-Term Memory subtests and the Short-Term Memory IQ score.

6. There will be significant correlations between individual Binet IV subtest scores, Area Scores and the Binet IV Composite IQ score.

ii. The Discriminative Ability of Binet IV, Wide Range Achievement Test-Revised, Developmental Test of Visual-Motor Integration, and Perception of Ability Scale for Students among Special Populations.

Hypothesis:

7. Scores from IQ, achievement, visual-motor and self concept tests will be significantly different for learning disabled, mentally retarded, gifted and parental interest (normal) groups.

111. Convergent Validity of The Wide Range Achievement Test-Revised, The Developmental Test of Visual-Motor Integration and The Perception of Ability Scale for Children to Binet IV.

Hypotheses:

8. The WRAT-R Reading and Spelling subtests will correlate significantly with Binet IV's Composite IQ, Verbal Reasoning IQ and Verbal Reasoning subtests.

9. The WRAT-R Arithmetic subtest will correlate significantly with Binet IV Composite IQ, Quantitative Reasoning IQ, Short-Term Memory IQ, Quantitative and Number Series subtests.

10. The VMI will correlate significantly with Binet IV's Composite IQ, Abstract/Visual Reasoning IQ score, Copying, Pattern Analysis, and Matrices subtests.

11. The PASS will correlate significantly with Binet IV Composite IQ, Area Scores and specific individual subtests.

CHAPTER 111

RESEARCH DESIGN AND PROCEDURES

Procedure

Data for this thesis were collected by graduate students enrolled at the Master's Level at the University of Alberta who attended the Spring/Summer and Fall/Winter sessions of the Educational Psychology 524 assessment course, 1986-1987.

Each of the subjects in the study was administered the Stanford-Binet Intelligence Scale: Fourth Edition and additional tests that comprised the test battery. The graduate students were competently trained in the administration, scoring and interpretation of all of the tests in the respective batteries, and as a standard procedure of the Educational Psychology 524 course were supervised in all aspects of their work by a trained senior psychologist.

The subjects for the study were clients from the Education Clinic at the University of Alberta who are either self referred, or referred by parents, teachers, administrators or other professionals. There is no fee charged for the services provided by the clinic. However, the clients utilize the services of the clinic with the understanding that student clinicians will perform assessments or counselling services as necessary. There is no age restriction for clients to avail themselves of the clinic services. Information about the clinic is made available to the public through radio programs hosted by University professors from the Educational Psychology Department; via schools and other public sources.

Clients are usually accepted to the clinic on a first- come-first served basis. Exceptions are made, particularly when school personnel or other professionals specifically request assessments. In addition to clients coming to the Education Clinic for services, graduate students go outside the agency to assess clients in homes,

schools, hospitals, or other institutions. These assessments are conducted both within Edmonton's city limits and its outlying cities and communities.

The data for this study were collected by reviewing the subjects' test reports. These reports are held within confidential files in the clinic and are available only for students registered in the appropriate courses. Test report information was entered on data sheets, key punched on computer cards, and then a computational analysis of scores was performed using SPSS/PC, a statistical package for the IBM micro-computer. Information was thus gathered and compiled in a systematic way and put into a format which comprised the basis for this study.

Sample

The initial sample consisted of 119 subjects. These were subjects who had been seen by student clinicians attending the Spring/Summer session of the Educational Psychology 524 assessment course. Thirty-four additional subjects were added to the sample from Fall referrals to the clinic. Appendix A reports the distribution of clients in terms of demographics. The final sample consisted of 153 subjects ranging in age from 3 years, 6 months to 27 years, 5 months, ($\bar{X}$ age = 10.1). This range compares similarly to the American sample used for the Binet IV technical manual (Thorndike et al. 1986A) which had an age range of 2 years, 11 months to 23 years, 11 months. No percentages or mean age are reported in the manual. Grade levels of the subjects in this study ranged from pre-school to grade 12 with seven subjects categorized as special education students. Although no grade levels are reported in the manual it is assumed from the ages that all grade levels were included in their sample. They do point out that an oversampling for ages 5 ($\underline{N}$= 460), 7 - 8 ($\underline{N}$= 415), and 12 - 13 ($\underline{N}$= 313) was deliberately conducted for potential future research. In this study there were 91 male (59.5%) and 62 female (40.5%) subjects. The American sample was more evenly distributed with 48.3% males and 51.7% females. From the University of Alberta's

Psychological Testing Center's "Report Forwarding Form" (Appendix B) fourteen categories emerged as referral reasons. These reasons included: readiness for kindergarten, pre-school assessment, learning disabilities (child), school learning problems (general), behavior problems, developmental delay/maturation, self-referral-general interest, reading/language problems, neuropsychological, giftedness/enrichment, parental interest (in child's ability), special class or program placement, special school placement and mental retardation. For the purpose of this study, these fourteen categories were regrouped into five subgroups, (see Table 3.1) and categorized in the following manner:

1. Learning difficulties (Learning disabled) included the following referral reasons: learning difficulties (child), school learning problems (general), reading/language problems and developmental delay/maturation.

For the purpose of this study, the operational definition of learning disabilities was provided by a Learning Disabilities handbook compiled by the Lethbridge and District Association for Children and Adults with Learning disabilities. They utilize the definition adopted by the Learning Disabilities Association of Canada LDAC, October 18, 1981.

> 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 delay in early development and/or difficulties in any of the following areas: attention, memory, reasoning, coordination, communicating, reading, writing, spelling, calculation, social competence, and emotional maturation.
>
> Learning disabilities are intrinsic to the individual and may affect learning and behaviour in any individual, including those with potentially average, average, or above average intelligence.

> Learning disabilities are not due primarily to visual, hearing, or motor handicaps; to mental retardation, emotional disturbance, or environmental disadvantage; although they may occur concurrently with any of these.
>
> Learning disabilities may arise from genetic variations, bio-chemical factors, events in the pre to perinatal period, or any other subsequent events resulting in neurological impairment. (C.A.C.L.D., 1986).

2. Mental retardation included the following referral reasons:

   mental retardation, special class or program placement, special school placement.

The operational definition of mental retardation for the purposes of this study is the one established by The American Association on Mental Deficiency (AAMD), which defines Mental Retardation as:

> . . . significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior, and manifested during the developmental period. (Grossman,1973, in Sattler).

Sattler in his discussion of mental retardation points out the AAMD classification of mental retardation consists of four categories: mild, moderate, severe and profound. The level of retardation arrived at by a particular intelligence test is dependent upon the standard deviation of the test. Usually a minus two standard deviation criterion is used, thus giving different cut-off scores for the Binet in comparison to the WISC-R, WPSSI and WAIS-R. The Range in IQ for the Stanford-Binet, Form L-M as reported in Sattler (1982) gives the following levels and range:

Level of Mental Retardation	IQ Range For Stanford-Binet Form L-M
Mild	67 - 52
Moderate	51 - 36
Severe	35 - 20
Profound	<20

3. Pre-School included the following referral reasons: readiness for kindergarten (4-5), pre-school assessment. Due to the very small N, this category was eliminated from discussion in this thesis.

4. Gifted included the following referral reason: giftedness/ enrichment.

The operational definition of "gifted" for the purposes of this study was taken from the 1983 Report of the Minister's Task Force on Gifted and Talented Pupils commissioned by Alberta Education. Their two-part definition is as follows:

> Gifted and talented pupils are those who by virtue of outstanding abilities are capable of exceptional performance. These are children who require differentiated provisions and/or programs beyond the

regular school program to realize their contribution to self and society.

Children capable of exceptional performance include those with demonstrated achievement and/or potential ability in one or several areas:

a. general intellectual ability,
b. specific academic aptitude,
c. creative or product thinking,
d. visual and performing arts..... and since 1983, have added
e. leadership ability
f. psychomotor ability. (Alberta Education. Special Education Services, 1986)

5. Parental Interest (Normal) included the following referral reasons: parental interest (in child's ability), self-referral or behavior problem.

a. Parental interest: 27 out of the sample of 35 (77%) gave general measurement of the child's current level of intellectual functioning with an interest in investigating their child's strengths and weaknesses as reason for referral, whereas three parents (8.5%) listed parental concern regarding future school success and establishment of reasonable expectations for the child as the referral reason.

b. Behavior problems: Four out of 35 (11.4%) listed general behavior problems in school - such as inability to control temper, frustration or lack of motivation as the referral reason.

c. Self referral: One self referral (2.8%) was initiated by a high school graduate wishing a current assessment of strengths and weaknesses in relation to his overall level of functioning to assist planning future academic routes

TABLE 3.1

REASON FOR REFERRAL BY SUBGROUP

SUBGROUP	NUMBER	PERCENT	MEAN AGE
Learning Difficulties	72	47.1	10.05
Mentally Retarded	15	9.8	13.63
Pre-School	9	5.9	4.87
Gifted	22	14.4	11.31
Parental Interest (Normal)	35	22.9	9.59

Four ethnic groups were identified in the Canadian sample with "whites" comprising almost the entire population - over 88%. Representation of the American standardization sample by ethnicity included the five categories: Native American (1.7 %), Hispanic (6.3 %), Black (14.4%), white (74.6%), Asian/Pacific Islander (2.5%) and other (.5%). The American sample closely approximates the U.S. population as was intended by the researchers. The largest proportion of this American sample is also comprised of "whites".

Approximately 65 percent of the population in this study came from an urban background. The American standardization sample is categorized into large cities and cities which combined are representative of 54.7% of the sample; with small towns and rural areas making up 45.3% of the sample.

Of the 153 subjects, thirteen (8.5%) were only children and 67 (43.8%) were first-born children. With regard to parent's marital status the greatest proportion of the sample (83.7 %) are currently married. No information is given regarding birth order or marital status for the American sample.

Occupations were listed within seven categories - six of which were the same occupational categories used in the technical manual (Thorndike et al. 1986A). One additional category - "Unemployed" was added to this study. Both father's occupation

and mother's occupation are included to ensure the inclusion of those situations where the mother is the sole wage earner or when the mother's occupation is ranked at a higher occupational level than the father's. As can be seen by Table V111 in Appendix A, the Managerial-Professional category of the father's occupations was represented by the largest number of subject's parents, (27.5%), followed by the Technical-Sales category (13.1%). This overrepresentation by the Managerial-Professional category and Technical-Sales categories for parent occupation is very similar to the figures presented in the standardization sample in the technical manual. Managerial-Professional (45.9%) and Technical-Sales (26.2%) constitute approximately three-quarters of the entire American sample. The Canadian sample appears to be more similar to the weighted sample than the actual sample. The only category which differs is the farming/forestry which one would assume to be higher for this study because of Alberta's agricultural basis. Presumably, if a representative Canadian sample was available the farming/forestry percentage would in all likelihood be lower than what this prairie province sample produced, and perhaps would more closely approximate the American sample.

Unfortunately, there was insufficient information available for a "Parent's education level" category so no information can be reported on those figures. It would be interesting to note if the numbers followed the same pattern as those listed for parental education levels in the technical manual's (Thorndike et al. 1986A) standardization sample.

Other information pertaining to the subjects that was available and was of interest to the researcher was that of handedness and whether the subjects had repeated a grade. One hundred and thirty-one subjects were right handed, 17 were left-handed, one alternated and no information was available on four subjects for this variable. Thirty-two subjects (20%) repeated at least once during their schooling; 108 (70%) had not

repeated at this point in their schooling. There was no information on 13 subjects in this area.

Tests Administered

Stanford-Binet Intelligence Scale: Fourth Edition, (Binet IV); Wide Range Achievement Test-Revised (WRAT-R); Beery's Developmental Test of Visual Motor Integration (VMI); and the Perception of Ability Scale for Students (PASS). All scores cited are standard scores except in the case of the PASS where only raw scores were available for reporting.

Stanford-Binet Intelligence Scale: Fourth Edition

Refer to chapter two for a detailed discussion of this scale.

Wide Range Achievement Test - Revised (WRAT-R)

Joseph J. Jastak was responsible for the development of the original Wide Range Achievement Test (WRAT) standardized in 1936. The WRAT was developed as a quick-screening instrument to evaluate children's achievements in the areas of reading (word recognition, naming letters, and pronunciation of words out of context); spelling (copying marks similar to letters, writing one's name and spelling words from an oral dictation); and arithmetic (counting, reading numbers symbolically, and solving oral and written computations).

Five revisions have occurred over the past fifty years with the most recent revision completed in 1984. Sarah Jastak and Gary S. Wilkinson were responsible for the latest major revision which redefined and restandardized the WRAT using six thousand subjects.

The 1984 WRAT-R includes a number of changes from the former editions of the test. These include the following thirteen changes as outlined in the WRAT-R manual:

1. Separate test forms for Level 1 and Level 2 with larger print and more space on each form.
2. National stratified sampling by age.
3. Rasch item analysis and scaling.
4. Standard errors of measurement at multiple scale and age levels.
5. Revised standard scores based on new age norms.
6. New non-decimal grade equivalent scores which are on an ordinal scale rather than an interval scale.
7. Person separation reliability coefficients at all ages.
8. Item separation reliability coefficients at all ages.
9. Test-retest reliability coefficients at four age levels.
10. White/non-white item difficulty comparisons.
11. Minor item changes to Arithmetic Levels 1 and 2.
12. Extension of Level 2 norms to include individuals from 65 years 0 months to 74 years, 11 months.
13. Bridge of Level 1 and 2 for two age groups above and below cut off age. (Jastak & Wilkinson, 1984).

One of the WRAT's main purposes was to be utilized in conjunction with intelligence tests and behavior adjustment. According to the WRAT manual,

> The method of measuring the basic codes was advisedly chosen to achieve the following ends: (1) to study the sensory-motor and coding skills involved in learning to read, spell, write, and figure, (2) to provide simple and homogeneous content, (3) to avoid duplication and overlapping with tests of comprehension, judgment, reasoning, and generalization studied by means other than reading, spelling and arithmetic tests, (4) to free diagnostic inferences from common confusions due to operational semantics, and (5) to permit validity analyses by the method of internal consistency. (Jastak & Wilkinson, 1984).

The WRAT-R's simplicity in design in conjunction with its meaningful norms accounts for its very popular usage for school and clinic populations.

The WRAT-R has two forms: Level 1 for subjects aged 5-0 to 11-11, and Level 2 for subjects aged 12-0 to 74-11. Being an age-normed test, each individual is able to

have their score compared to a similar aged group of individuals, representative of the general population. Each level takes approximately twenty to thirty minutes to administer.

The WRAT-R will generate four kinds of scores: raw, grade equivalents, standard, and percentiles. A short description follows for the four types of scoring.

> Raw Scores: are an ordinal scale score and are usually utilized in a transformation process in attaining an interval or ratio score.
>
> Grade Equivalents: include the opportunity for an individual's scores to be compared with samples of individuals which represent a specific educational level. Grade equivalent scores are shown via whole numbers with a smaller part of the grade indicated by the letters B (beginning), M (middle) and E (end) for each grade level. These grade ratings are especially useful to utilize when explaining test results to parents.
>
> Standard Scores: are deviation quotients with a mean of 100 and a standard deviation of 15. They are the kind of score which can be utilized in comparisons within and between individuals (inter and intra-individual comparisons). According to the manual standard score ratings include the following classifications:

Classification	Score Range
Very Superior	130 and up
Superior	120 to 129
High Average	110 to 119
Average	90 to 109
Low Average	80 to 89
Borderline	70 to 79
Deficient	69 and below (Jastak & Wilkinson, 1984, p.29)

Standard scores of the WRAT-R can be used to compare results of other standardized tests - such as IQ. From the standpoint of the four scores that are reported the most useful score for comparative purposes is the standard score which was utilized for this study.

> Percentiles: are ordinal ranks that range from 0 to 100. The manual cautions percentile use in comparative research.

Reliability and Validity:

Test-retest coefficients for the WRAT-R Level 1 Reading, Spelling and Arithmetic are reported to be .96, .97 and .94 respectively. Test-retest coefficients for Level 11 are somewhat lower - .90, .89 and .79 for Reading, Spelling and Arithmetic. (Jastak & Wilkinson, 1984, p. 64).

Concurrent validity cited in the manual for the WRAT indicate that when comparing the WRAT with other achievement and ability tests the following correlations existed - high .60's to the .80's when compared with the Peabody Individual Achievement Test, California Achievement Test and the Stanford Achievement Test.

Tests of construct and concurrent validity are cited for the WRAT rather than the WRAT-R as information for the WRAT-R in these areas is not currently available.

The Developmental Test of Visual-Motor Integration: (VMI)

The VMI, a tool basically for educational assessment, is a perceptual-motor ability test in which children aged two through fifteen are required to copy twenty-four geometric designs. These forms involve skills relating to sensory perception and motor functions. They are arranged in increasing order of difficulty in that each successive figure has a higher 'estimated age' for successful completion. Scoring is on a pass-fail basis. When sizable discrepancies occur between a child's visual motor integration age equivalent score and their chronological age, then diagnostic significance is often assigned.

Reliability and Validity:

Dyckman and Rentfrow report test-retest correlations on children from the second, fourth and sixth grades of .62 to .34 respectively. (Dyckman & Rentfrow, 1971, cited in Fisk & Janzen, 1981, p. 256). The manual (Beery, 1967) reports correlations between VMI scores and chronological age to be .89 for the two to 15 age range, with

VMI correlations being higher with mental age than with chronological age. The manual also reports that VMI correlations with reading achievement at the first grade were higher than those between IQ and reading achievement. No specific IQ or reading achievement tests were mentioned. The manual seems to indicate that the VMI is more a measure of a child's coordinating abilities. Coordination is most definitely involved in writing where a child must coordinate both visual perception and motor behavior. Thus the functions of integration and coordination are essential aspects of visual and motor skills.

The VMI may be administered individually or in groups. The subjects are to copy the forms with a pencil, without erasing and without being timed. Only one attempt on each geometric form is allowed. Time to administer the test will depend on the age of the subject as well as the level of competence. The VMI is a very easy test to administer and requires only paper and pencil.

Perception of Ability Scale for Students (PASS)

The PASS is a short, easy to administer tool to measure academic self concept in elementary and junior high school children. Seven scales make up the PASS and include: (1) general ability, (2) arithmetic, (3) reading and spelling, (4) school satisfaction, (5) penmanship and neatness, (6) confidence and (7) a full scale score which according to PASS authors, "is an index of general academic self-confidence." (Boersma, Chapman & Maguire, 1979).

These scales are represented by raw scores and are grouped in the following table as descriptive classifications for PASS scores:

Classification

Scale	Weak	Bel. Aver.	Average	Ab. Aver.	Strong
Full Scale	≤ 35	36 - 40	41 - 51	52 - 56	57 ≥
General Ability	≤ 4	5 - 6	7 - 9	10 - 11	12
Arithmetic	≤ 5	6 - 7	8 - 10	11	12
Sch. Satisfaction	≤ 4	5 - 6	7 - 9	10 - 11	12
Read/Spell.	≤ 5	6 - 7	8 - 10	11	12
Pen./Neatness	≤ 4	5 - 6	7 - 9	10 - 11	12
Confidence	1	2	3 - 5	6 - 7	8 ≥

(Boersma & Chapman, 1985, p. 37).

Reliability and Validity

> Internal consistency estimates utilizing Cronbach's alpha yielded a full scale alpha of .92. Test-retest reliability over a four to six week period was .83. (Fisk & Janzen, 1981).

The authors of PASS discuss validity studies using the following measures: a. School achievement based on report card grades - some specific PASS subscales had moderate correspondence with report card grades ranging from .27 to .37; b. Standardized tests (Canadian Test of Basic Skills) - intercorrelations ranged from .12 to .44 with a number of negative correlations reported; c. Intelligence measures - the PASS has a very low almost non-existant relationship with IQ scores. This may suggest that a student's perception of school ability may be relatively independent of individual and group intelligence test measures.

Analysis of the Data

Frequencies and percentages were generated for the total sample in terms of the demographic variables of age, grade, gender, ethnicity, locale, child's position in the family, parent's marital status, parent occupation, handedness and grade repetition. These variables are reported in Appendix A, Tables I to XI.

The variable "Reason for Referral" (Table 3.1) was reported in terms of frequencies, percentages and mean ages for the five subgroups of learning difficulties (learning disabled), mentally retarded, pre-school, gifted and parental interest (normal) subjects.

Summary statistics for the total sample reported in Appendix C, Table X111 include means, standard deviations, kurtosis, skewness, minimum and maximum scores and the standard error of the mean for WRAT-R subtests, VMI, PASS subtests, Binet IV individual subtests, Area Scores and the Composite IQ score.

Pearson Product Moment Correlations were calculated to determine the relationships between Binet IV Composite IQ, Area Scores and individual subtests. Pearson correlation coefficients (r) will give an indication whether there is strong or weak linear association between two variables. Correlations were calculated to determine Binet IV's relationship with the WRAT-R, VMI and PASS. A 27 by 27 correlation matrix (Table 4.36) was devised to give the pertinent information regarding significant correlations and intercorrelations.

All the instruments in this study had some missing data due to certain tests not being administered to all of the subjects. Table 3.2 includes the list of the variables and the percentage of data missing.

TABLE 3.2

Percentage of Missing Data

VARIABLE	PERCENTAGE OF DATA MISSING
WRAT-R subtests	22.9
VMI	36.7
PASS subtests	53.0
BINET IV subtests	
Comprehension	2.0
Absurdities	7.9
Pattern Analysis	.7
Copying	25.0
Matrices	54.5
Quantitative	2.0
Number Series	55.0
Bead Memory	1.4
Memory for Sentences	3.3
Memory for Digits	40.6
Memory for Objects	43.8
Quantitative IQ	1.4
Short-Term Memory IQ	.7

Three of the Binet IV subtests - Verbal Relations, Paper Folding and Cutting and Equation Building were not included in the reported statistical data. Only 9% of the entire sample were administered the Verbal Relations and Equation Building subtests; and 12% the Paper Folding and Cutting subtest. These three subtests are also not included in the technical manual's descriptive raw score statistics by age until at least 12+. The numbers reported are smaller than for other subtests until the age of 14 and up where they approximate similar number of cases reported for the other Binet IV subtests.

Since missing data were present in this study a specific procedure defined as "pairwise missing-value treatment" was used. (Norusis, 1984, p. 124b). This technique involves the calculation of the correlation coefficient between a pair of variables based on all cases with complete information for the two variables regardless of whether the cases have missing data on any other variable. This procedure is only

useful when one is satisfied that the missing data are random. One must exercise caution in viewing significance levels utilizing this particular method of calculation. However, given the circumstances that not all subjects were administered all of the tests in the battery this seemed to be the most satisfactory alternative. The other alternative was to use only those subjects who had taken every subtest of every instrument in the battery. This however reduced the sample size from 153 to eight cases thus limiting the feasibility of utilizing that method with a sample of such miniscule size.

A oneway analysis of variance (ANOVA) was used to test the hypotheses that at least two population means in a set were not equal. Appendix D, Tables XIV to XL1 report these results. Tables 4.7 to 4.34 report Scheffe significant differences. ANOVA was done to determine whether differences existed between the mean scores attained by learning disabled, mentally retarded, gifted and normal subjects. Anovas were used to determine differences on measures of Binet IV Composite IQ scores, Area scores, individual subtest scores, academic achievement, visual-motor integration development and self concept.

In analysis of variance, variability in the sample consists of two areas - the variability within group - measured by the within groups sum of squares; and the between-groups sum of squares which measures the variability of the group means. When a significant F statistic is present this is an indication that population means are probably not equivalent. A significant F does not indicate which means differ. A multiple comparison test is used to determine which population means differ from one another. Multiple comparison tests protect against indicating too many differences as significant. They have more stringent criteria for declaring differences significant than the usual t-tests. The Scheffe test, a posteriori or post hoc comparison test was used in this study. It was chosen over the Tukey because group sizes were not equal in this study. Although the Scheffe is probably one of the most conservative of the multiple comparison tests it is the least likely to make a Type 1 error (Moore, 1983). The

Scheffe method requires larger differences between means for significance than most of the other multiple comparisons tests (Norusis, 1984).

Crosstabulation analysis was used to determine whether there were relationships between selected variables of gender, only child, child's position in the family, parent marital status, parent's occupation, repeated grade and locale with each of the four subgroups. Table 4.35 reports these crosstabulation results.

Figures 4.1 and 4.2 were plotted to see if differences on the Binet IV Composite, Area and subtest scores showed parallel patterns for the four designated subgroups.

Limitations of the study:

1. The data were collected only from subjects referred or self-referred to the Education Clinic, thus limiting the generalizability of the results to a special clinic referral population.
2. Not all the tests administered in the battery were administered to all of the subjects. (See Table 3.2).
3. Referral reasons were not always precise or specific.
4. When dealing with reasons for referral, subjective judgment calls were made by the researcher in cases where more than one referral reason was listed.

CHAPTER IV

RESULTS

The results of the study are presented in three main sections corresponding to the research questions and the hypotheses formulated earlier (see Chapter 11). The intent of the questions was to focus on the similarities and differences between the data from this Canadian sample and the American standardization sample on the Binet IV; the discriminative ability of Binet IV, WRAT-R, VMI and PASS among special populations; and the convergent validity of the WRAT-R, VMI and PASS to the Binet IV.

In comparing Canadian data with the American sample, total sample means and standard deviations for all variables in this study are reported in Table 4.1. This table also includes variable names and numbers assigned to each variable for referral purposes for other tables discussed later in the study.

Means and standard deviations for the non-exceptional sample of this study and the American sample for Binet IV's Composite IQ and Area Scores are presented in Table 4.2. Table 4.3 reports the results of T-Tests conducted on the Canadian and American samples for Binet IV subtests, Area Scores, and Composite IQ scores.

Table 4.4 summarizes the intercorrelations between Binet IV Composite Score, Area Scores and individual subtests. Comparisons between these statistics and those comparable statistics from the technical manual (Thorndike et al. 1986A) will be reported.

TABLE 4.1

Total Sample Frequencies, Means and Standard Deviations for All Variables

VAR. #	VARIABLE NAME	N	MEAN	ST. DEV
V18	WRAT Arithmetic	118	96.6	19.1
V20	WRAT Reading	118	98.7	19.5
V22	WRAT Spelling	118	94.4	20.4
V24	VMI	97	10.2	9.2
V26	PASS Full Scale	72	44.3	13.4
V27	PASS General Ability	72	6.8	3.2
V28	PASS Arithmetic	72	8.8	2.7
V29	PASS School Satisfaction	72	7.9	2.7
V30	PASS Reading/Spelling	72	8.4	3.3
V31	PASS Penmanship/Neatness	72	7.8	3.0
V32	PASS Confidence	72	4.7	2.5
V38	BINET Vocabulary	153	51.3	9.2
V39	BINET Comprehension	150	51.7	9.4
V40	BINET Absurdities	131	52.7	8.3
V42	BINET Pattern Analysis	152	51.5	8.9
V43	BINET Copying	115	48.5	8.3
V44	BINET Matrices	85	52.6	8.5
V46	BINET Quantitative	150	50.9	10.0
V47	BINET Number Series	70	53.6	8.3
V49	BINET Bead Memory	151	50.5	9.3
V50	BINET Memory For Sentences	148	49.2	9.1
V51	BINET Memory For Digits	91	52.2	8.1
V52	BINET Memory for Objects	86	52.0	7.6
V53	BINET Verbal Reasoning IQ	153	104.3	19.1
V54	BINET Abstract/Vis. Reasoning IQ	153	102.2	17.3
V55	BINET Quantitative Reasoning IQ	151	103.2	20.3
V56	BINET Short-Term Memory IQ	152	101.8	18.3
V57	BINET Composite IQ	153	103.4	19.1

*Please note: The following variables were not included in this sample due to very few examinees being administered V41 (Verbal Relations); V45 (Paper Folding and Cutting): and V48 (Equation Building).

Table 4.5 reports means and standard deviations for Binet IV's individual subtests, Area Scores and Composite IQ scores for the learning disabled, mentally retarded, gifted and normal subjects.

Data regarding hypotheses one through six will be discussed in the first section.

The second major section will deal with the discriminative ability of the Binet IV, WRAT-R, VMI and PASS among special populations. Table 4.6 compares the Canadian and American samples on Binet IV's Composite IQ, and Area Scores for the three populations of gifted, learning disabled and mentally retarded subjects. One way ANOVA of the group means on all the variables for the four populations of learning disabled, mentally retarded, gifted and normal are presented in Appendix D. Results of the Scheffe post hoc tests at the .05 significance level are presented in tables 4.7 to 4.34. Figures 4.1 and 4.2 are line graphs representative of the mean scores attained by the four exceptional populations on Binet IV individual subtests, Area Scores, and Composite IQ scores. Table 4.35 reports significant levels of the crosstabulations between characteristics other than performance on IQ, achievement, visual motor or self-concept for the four exceptional populations. Data regarding hypothesis number seven will be discussed in this second section.

The last section examines the convergent validity of the WRAT-R, VMI, and PASS to the Binet IV. Table 4.36, a 27 by 27 correlation matrix, summarizes all the significant correlations of these measures. Hypotheses eight through 11 will be examined.

CANADIAN DATA VS. AMERICAN DATA

Hypothesis 1

The means and standard deviations of the total sample of this study for Binet IV Composite IQ, Area Scores, and individual subtest scores will be similar to the manual's normal standardization sample.

Mean scores and standard deviations were calculated for all variables and have been recorded in Table 4.1 of this study.

Table 4.3 of the technical manual (Thorndike et al. 1986A) presents Standard Age Score means and standard deviations for each Binet IV subtest. Means reported in the manual are all very close to the expected 50.0 and the standard deviations are approximately eight. The manual means range from 49.5 for Copying to 50.5 for Paper Folding and Cutting. Standard deviations from the standardization sample range from 7.3 for Equation Building to 8.5 for Memory for Sentences.

In this study there is a wider range of mean scores and standard deviations. Mean scores range from 48.5 for Copying to a high of 53.6 for Number Series. Copying had the lowest mean score for both Canadian and American samples. Standard deviations for this sample ranged from a low of 7.6 for Memory for Digits to 10 for Quantitative. Although this study's means and standard deviations approximated scores of 50 and 8, they tended to be somewhat higher than the American sample.

TABLE 4.2

Comparison of Canadian Vs. American Data On Binet IV Composite IQ and Area Scores for Non-exceptional samples.

	Canadian Sample			American Sample*		
	N	$\bar{X}$	SD	N	$\bar{X}$	SD
Comp. IQ	153	103.4	19.1	5013	99.7	16.1
Verb. Reas. IQ	153	104.3	19.1	5013	100.1	16.3
Ab./Vis. Reas. IQ	153	102.2	17.3	5013	99.5	16.0
Quant. Reas. IQ	151	103.2	20.3	5013	99.8	15.9
Sh.-Term Mem. IQ	152	101.8	18.3	5013	99.5	16.3

*Table 4.4, adapted from - (Thorndike et al. 1986A, p. 33).

Table 4.4 (Thorndike et al. 1986A, p. 33) of the technical manual reporting Binet IV Area Scores and Composite IQ scores is utilized to compare this sample with the American sample. Table 4.2, in reporting the mean scores and standard deviations of this study, indicates similar scores.

As can be seen by Table 4.2, differences in means between the two samples are as follows: Composite IQ=3.7; Verbal Reasoning IQ = 4.2; Abstract/Visual Reasoning IQ = 2.7; Quantitative Reasoning IQ = 3.4; and Short-Term Memory = 2.3. The differences are slight between the two samples of non-exceptional children except for the 4.2 point discrepancy between the Verbal Reasoning IQ scores. Standard deviations for the Canadian sample are considerably higher than the American sample which more closely approximates a score of 16. All of the Canadian sample standard deviations are above 16, with Quantitative Reasoning IQ the highest at 20.3 and the Abstract/Visual IQ standard deviation the lowest at 17.3. The largest difference between the two samples occurs with the Quantitative Reasoning IQ standard deviations where there is a 4.4 point difference. In comparing these samples in Table 4.2 it appears that mean Area Scores and mean Composite IQ scores for the two samples are very similar. However, in order to determine if there are any significant differences between the Canadian and American samples, t-Tests were conducted not only on mean Area Scores and mean Composite IQ scores, but also on the mean scores of Binet IV subtests as well. Table 4.3 reports results at the .05 level of significance.

TABLE 4.3

T-Tests

Canadian Sample Vs. American Sample According to Stanford-Binet Intelligence Scale 4th Edition - Individual Subtests, Area Scores and Composite IQ Scores

	CANADIAN SAMPLE			AMERICAN SAMPLE*				
	NO.	MEAN	SD	NO.	MEAN	SD	df	t	p
VAR									
V38	153	51.3	9.2	5013	50.0	8.1	5164	1.96	.05
V39	150	51.7	9.4	5013	50.1	8.4	5161	2.29	.05
V40	131	52.7	8.3	3824	50.0	8.2	3953	3.70	.05
V42	152	51.5	8.9	5013	49.6	8.1	5163	2.84	.05
V43	115	48.5	8.3	3381	49.5	8.1	3494	1.69	NS
V44	85	52.6	8.5	3030	50.0	7.9	3113	2.98	.05
V46	150	50.9	10.0	5013	49.6	8.4	5161	1.86	NS
V47	70	53.6	8.3	3020	49.9	7.8	3088	3.92	.05
V49	151	50.5	9.3	5013	49.9	8.5	5162	.85	NS
V50	148	49.2	9.1	5013	49.5	8.5	5159	.42	NS
V51	91	52.2	8.1	3054	50.2	8.0	3143	2.35	.05
V52	86	52.0	7.6	3034	49.8	7.8	3118	2.58	.05
V53	153	104.3	19.1	5013	100.1	16.3	5164	2.32	.05
V54	153	102.2	17.3	5013	99.5	16.0	5164	1.55	NS
V55	151	103.2	20.3	5013	99.8	15.9	5162	1.93	NS
V56	152	101.8	18.3	5013	99.5	16.3	5163	1.26	NS
V57	153	103.4	19.1	5013	99.7	16.1	5164	2.18	.05

*Subtest information adapted from Table 4.3 of the technical manual (Thorndike et al. 1986A, p. 32); Area Scores and Composite IQ adapted from Table 4.4 of the technical manual (Thorndike et al. 1986A, p.33).

As can be seen by Table 4.3 eight of the twelve Binet IV subtests for both samples are significantly different at the .05 level of significance. T-Test results indicate that there are differences between the Canadian and American samples for Vocabulary, Comprehension, Absurdities, Pattern Analysis, Matrices, Number Series, Memory for Digits and Memory for Objects. Only Copying, Bead Memory and Memory for Sentences show no significant differences when comparing the means of individual subtests. For the four Area Scores and Composite IQ scores there are no significant differences reported between the two samples for Abstract/Visual Reasoning IQ, Quantitative Reasoning IQ or Short-Term Memory IQ scores. Significant differences are reported for Verbal Reasoning IQ and the Composite IQ score.

Therefore, Hypothesis1 does not hold for total similarities of mean scores and standard deviations between the two samples. The results only partially support Hypothesis1.

Hypotheses 2

There will be significant correlations between Verbal Reasoning subtests and the Verbal Reasoning IQ Score.

The intercorrelations for the 17 variables are shown in Table 4.4. The upper off-diagonal elements only are presented since the matrix is symmetrical. Decimal points have been omitted. All correlations (r) are positive and are significant at the .001 level except for Copying and Number Series, and Memory for Objects and Number Series which are significant at the .01 level.

As expected, in this sample, correlations between the Verbal Reasoning subtests were moderately high to high ranging from .56 for Comprehension and Absurdities to .67 for Absurdities and Vocabulary; with the highest r of .76 occurring between Vocabulary and Comprehension. This compares almost identically with the American

sample which reported r's of .59, .62 and .73 respectively. (Thondike et al. 1986A, p. 53).

The correlations between this study's Verbal Reasoning subtests and the Verbal Reasoning Area Scores are all very high - the highest being Vocabulary and Verbal Reasoning IQ at .92, then Comprehension and Verbal Reasoning IQ = .89, and Absurdities and Verbal Reasoning IQ = .84. Again these scores are very similar to the American sample values for these Verbal Reasoning variables being .90, .88 and .84 respectively.

Therefore, Hypothesis 2 is accepted.

Hypothesis 3

There will be significant correlations between Abstract/Visual Reasoning subtests and the Abstract/Visual Reasoning IQ score.

These correlations fall within the moderate range (Pattern Analysis and Copying at .43) to the very low range (Copying and Matrices at .18). This r of .18 differs a great deal from the American sample where the r between Copying and Matrices was at a moderate range of .41. The r's for other Abstract/Visual subtests in this study were in the .50 range as were those for the American sample. The r's between Abstract/Visual subtests and Abstract/Visual Reasoning IQ scores are high, Pattern Analysis and Abstract/Visual Reasoning IQ = .85, Copying and Abstract/Visual Reasoning IQ = .77; and Matrices and Absract/Visual Reasoning IQ = .71. These findings are similar to the American samples r's of .87 for Pattern Analysis and Abstract/Visual Reasoning IQ. However, the American sample r's are higher for Matrices and Abstract/Visual IQ (.85) than for Copying and Abstract/Visual IQ (.80).

Therefore, Hypothesis 3 is accepted.

TABLE 4.4*

Intercorrelations Between Binet IV Composite Score, Area Scores and Individual Subtests

	V38	V39	V40	V42	V43	V44	V46	V47	V49	V50	V51	V52	V53	V54	V55	V56	V57
V38		76	67	57	40	57	67	63	60	75	51	48	92	66	68	75	85
V39			56	56	38	47	60	67	54	62	50	44	89	61	64	66	80
V40				45	40	53	56	51	52	56	44	48	84	52	56	63	74
V42					43	40	48	60	52	45	54	46	61	85	53	60	73
V43						18	35	40	33	37	39	27	45	77	38	42	58
V44							68	63	62	48	49	53	60	71	70	67	74
V46								72	57	56	52	47	70	61	97	63	83
V47									63	60	50	37	70	76	90	70	85
V49										52	51	56	65	61	61	82	76
V50											61	49	73	55	56	85	76
V51												54	56	59	53	81	70
V52													53	50	48	78	63
V53														69	72	78	91
V54															67	69	85
V55																66	86
V56																	88
V57																	

*all decimal points have been omitted

**all variables are significant at the .001 level except for V43 and V47; and V47 and V52 which are significant at the .01 level.

Hypothesis 4

There will be significant correlations between Quantitative subtests and the Quantitative Reasoning IQ score.

The Quantitative and Number Series correlation for this study is .72 and is slightly higher than the American sample at .67. Correlations between Quantitative and Number Series and Quantiative Reasoning IQ are very high at .97 and .90 respectively. These r's are almost identical to the American sample's r's which are .91 and .89.

Therefore, hypothesis 4 is accepted.

Hypothesis 5

There will be significant correlations between Short-Term Memory subtests and the Short-Term Memory IQ score.

These moderate r's range from .49 (Memory for Sentences and Memory for Objects) to .61 (Memory for Sentences and Memory for Digits). The other r's fall within the .50 range. These r's are slightly higher than the American sample where r's range from .42 to .56. Correlations between Short-Term Memory subtests and the Short-Term Memory IQ score range from .78 for Memory for Objects to .85 for Bead Memory. A similar range of r's (.73 to .78) are reported for the American sample.

Therefore, Hypothesis 5 is accepted.

Hypothesis 6

There will be significant correlations between individual Binet IV subtest scores, Area Scores and the Binet IV Composite IQ score.

Correlations are moderately high between individual subtests and the Composite IQ score. The range of r's is .58 for Copying to .85 for Vocabulary and Number Series. Other r's are in the .70's to low .80's. Although r's are similar for the American sample, the lowest r in the standardization sample is Memory for Objects at .60, then Memory for Digits at .64 and then Copying at .66. The highest r's are for Quantitative and Number Series at .82 and Vocabulary at .81. The rest of the subtest r's are in the low to high .70's.

Area Score correlations with Composite IQ in this study are in the mid .80's to .90's with Verbal Reasoning IQ and the Composite IQ score having the highest r at .91. The American sample has a similar range with the highest r being Quantitative Reasoning IQ and the Composite IQ score at .90 followed by Verbal Reasoning at .89.

Therefore Hypothesis 6 is accepted.

Conclusion

All in all the intercorrelations for this study are very similar to the American sample reported in the technical manual (Thorndike et al. 1986A). These moderate to high correlations appear to indicate that Binet IV seems to work in this study as it has for the American standardization sample.

THE DISCRIMINATIVE ABILITY OF BINET IV, WIDE RANGE ACHIEVEMENT TEST-REVISED, DEVELOPMENTAL TEST OF VISUAL-MOTOR INTEGRATION, AND PERCEPTION OF ABILITY SCALE FOR STUDENTS AMONG SPECIAL POPULATIONS

Hypothesis 7

Scores from IQ, achievement, visual-motor and self-concept tests will be significantly different for learning disabled, mentally retarded, gifted and parental interest (normal) groups.

Means and Standard Deviations for the Canadian Sample

In order to discuss the IQ results of this hypothesis, means and standard deviations for Binet IV's individual subtests, Area Scores and Composite IQ scores for the four groups in this sample are reported in Table 4.5. There are definite differences on each of these areas for the four groups. The range of subtest mean scores for the learning disabled populations are 47.8 (Copying) to 52.2 for Absurdities. There is little variability among the subtest scores. Subtest mean scores for the mentally retarded subjects are considerably lower with the range being 34.5 for Memory for Sentences to 44.4 for Memory for Objects. The mean score range of approximately ten points for the mentally retarded group tends to be more variable than for the other three groups. The parental interest (Normal) mean subtest scores range from 50.8 for Memory for Sentences to 57.4 for Number Series. It is interesting to note that the lowest mean score for both the mentally retarded and normal populations is Memory for Sentences. As expected the mean scores attained by the gifted population was considerably higher than the other three groups.

TABLE 4.5

Means and Standard Deviations for Learning Disabled, Mentally Retarded, Gifted and Normal Populations on Binet IV Individual Subtests, Area Scores and Composite IQ Score

		LD			MR			GIFTED			NORMAL	
	N	$\bar{X}$	SD	N	$\bar{X}$	SD	N	$\bar{X}$	SD	N	$\bar{X}$	SD
V38	72	49.69	7.17	15	36.40	7.35	22	60.86	7.01	35	54.57	6.12
V39	70	50.62	8.04	15	38.53	8.76	21	59.95	6.22	35	53.65	7.58
V40	62	52.29	7.99	12	40.66	5.51	15	59.26	6.69	33	53.96	7.24
V42	72	50.55	6.59	15	40.26	8.09	21	58.19	9.04	35	54.71	8.44
V43	56	47.89	6.55	11	38.36	7.92	13	54.53	10.43	29	51.20	7.02
V44	44	51.75	6.39	8	39.62	8.00	15	59.53	8.40	17	54.47	6.84
V46	70	48.71	8.10	15	36.66	7.08	22	62.68	6.42	34	52.76	7.10
V47	34	51.23	5.58	6	38.83	6.79	15	61.46	6.06	15	57.40	5.57
V49	70	49.34	7.16	15	38.06	7.89	22	59.54	8.63	35	51.71	8.28
V50	69	48.71	7.95	14	34.50	6.90	21	57.85	7.11	35	50.82	6.51
V51	45	51.1	5.60	9	40.77	6.01	15	58.26	7.99	21	54.61	8.33
V52	41	52.12	6.25	10	44.40	8.32	14	55.21	6.71	21	53.28	8.33
V53	72	101.30	15.10	15	72.30	16.70	22	124.30	12.20	35	109.50	12.00
V54	72	99.80	10.90	15	74.70	14.00	22	120.10	16.20	35	108.40	13.40
V55	70	98.70	15.40	15	72.30	14.40	22	128.00	12.80	35	107.10	13.70
V56	71	99.40	13.30	15	71.80	15.00	22	121.60	13.00	35	106.00	14.70
V57	72	99.60	12.50	15	68.90	15.20	22	127.30	11.80	35	109.40	10.90

The range of mean scores for gifted subjects is 54.5 for Copying to 62.6 for Quantitative. The Copying subtest is reported as the lowest scoring subtest for both the learning disabled and the gifted populations.

Means and Standard Deviations: Comparisons of the Canadian and American Samples

Table 4.6 compares the Canadian and American Samples on Binet IV's Composite IQ and Area Scores for the three populations of gifted, learning disabled and mentally retarded subjects. The American "gifted" sample is taken from table 6.12, p. 72; the "learning disabled" sample from table 6.14, p. 75; and the "mentally retarded" sample from table 6.17, p. 80 of the technical manual (Thorndike et al. 1986A). Table 4.6 indicates much greater differences in mean scores attained by the exceptional samples than the non-exceptional samples reported in Table 4.2.

A. GIFTED

In looking at the "gifted" samples the Canadian mean scores for all the Area Scores and the Composite IQ are considerably higher than the American sample. The following differences occurred: Composite IQ = 11.0 points; Verbal Reasoning IQ = 10.8 points; Abstract/Visual Reasoning IQ = 10.5 points; Quantitative Reasoning IQ = 10.8 points and Short-Term Memory IQ = 5.1 points. It would appear that this Canadian sample more closely represents accepted mean scores for gifted students

TABLE 4.6

Comparison of Canadian Vs. American Data on Binet IV
Composite and Area Scores for Exceptional Populations

	Canadian Sample			American Sample		
A. GIFTED						
	$\bar{X}$ AGE = 11-3			$\bar{X}$ age = 12-11		
	N	$\bar{X}$	SD	N	$\bar{X}$	SD
Comp. IQ	22	127.3	11.8	19	116.3	16.4
Verb. Reas. IQ	22	124.3	12.2	19	113.5	18.3
Ab./Vis. Reas. IQ	22	120.1	16.2	19	109.6	15.1
Quant. Reas. IQ	22	128.0	12.8	19	117.2	16.9
Sh.-Term Mem. IQ	22	121.6	13.0	19	116.1	17.2
B. LEARNING DISABLED						
	$\bar{X}$ age = 10-0			$\bar{X}$ age = 11-0		
	N	$\bar{X}$	SD	N	$\bar{X}$	SD
Comp. IQ	72	99.6	12.5	90	84.8	14.5
Verb. Reas.	72	101.3	15.1	90	88.8	15.0
Ab./Vis. Reas. IQ	72	99.8	10.9	90	87.7	14.6
Quant. Reas IQ	72	98.7	15.4	90	86.3	14.1
Sh.-Term Mem. IQ	72	99.4	13.3	90	84.8	14.5
C. MENTALLY RETARDED						
	$\bar{X}$ AGE = 13.6			$\bar{X}$ age = 13-11		
	N	$\bar{X}$	SD	N	$\bar{X}$	SD
Comp. IQ	15	68.9	15.2	61	66.2	9.8
Verb. Reas. IQ	15	72.3	16.7	61	69.4	10.8
Ab./Vis. Reas. IQ	15	74.7	14.0	61	71.6	10.9
Quant. Reas IQ	15	72.3	14.4	61	73.9	11.1
Sh.-Term Mem. IQ	15	71.8	15.0	61	67.5	12.4

than the reported American sample. Other than the Short-Term Memory IQ score, mean area scores and the Composite IQ scores are close to one standard deviation above the American sample. Standard deviations vary for both samples with the largest difference of 4.6 points occurring between Composite IQs.

B. LEARNING DISABLED

When comparing the performances of designated learning disabled subjects on the four Area Scores and Composite IQ score, the Canadian sample means are all higher with the following differences between samples occurring: Composite IQ = 14.8 points; Verbal Reasoning IQ = 12.5 points; Abstract/Visual Reasoning IQ = 12.1 points; Quantitative Reasoning IQ = 12.4 points; and Short-Term Memory IQ = 14.6 points. It would seem that this Canadian sample of learning disabled subject's mean scores that approximate 100 are more in line with the operational definition of a learning disabled child utilized in this study. Learning disabled children are defined as being of average to above average intelligence. The mean scores reported for the American sample are exceptionally low. This study's learning disabled mean scores are approximately 15 points above the American sample's mean scores. Standard deviations for both of these samples are similar with the largest difference of 3.7 points occuring between the standard deviations of Abstract/Visual Reasoning IQs. The other differences between samples are slight.

C. MENTALLY RETARDED

The two samples are more similar in scores and patterns than the gifted or learning disabled samples. As was the case in the other two exceptional samples all mean scores for the Canadian mentally retarded subjects are higher than the American sample with the exception of the Quantitative Reasoning IQ mean. Mean score differences are as follows: Composite IQ = 2.7 points; Verbal Reasoning IQ = 2.9 points; Abstract/Visual Reasoning IQ = 3.1 points; Quantitative Reasoning IQ = 1.6 points and Short-Term Memory = 4.3 points. The highest mean score for the Canadian mentally retarded sample was the Abstract/Visual Reasoning Area and the lowest score was Short-Term Memory. For the American sample, the highest mean score was the Quantitative IQ mean, and the lowest was also Short-Term Memory. Standard deviations were generally three to seven points higher for the Canadian sample, but approximated a score of 16.

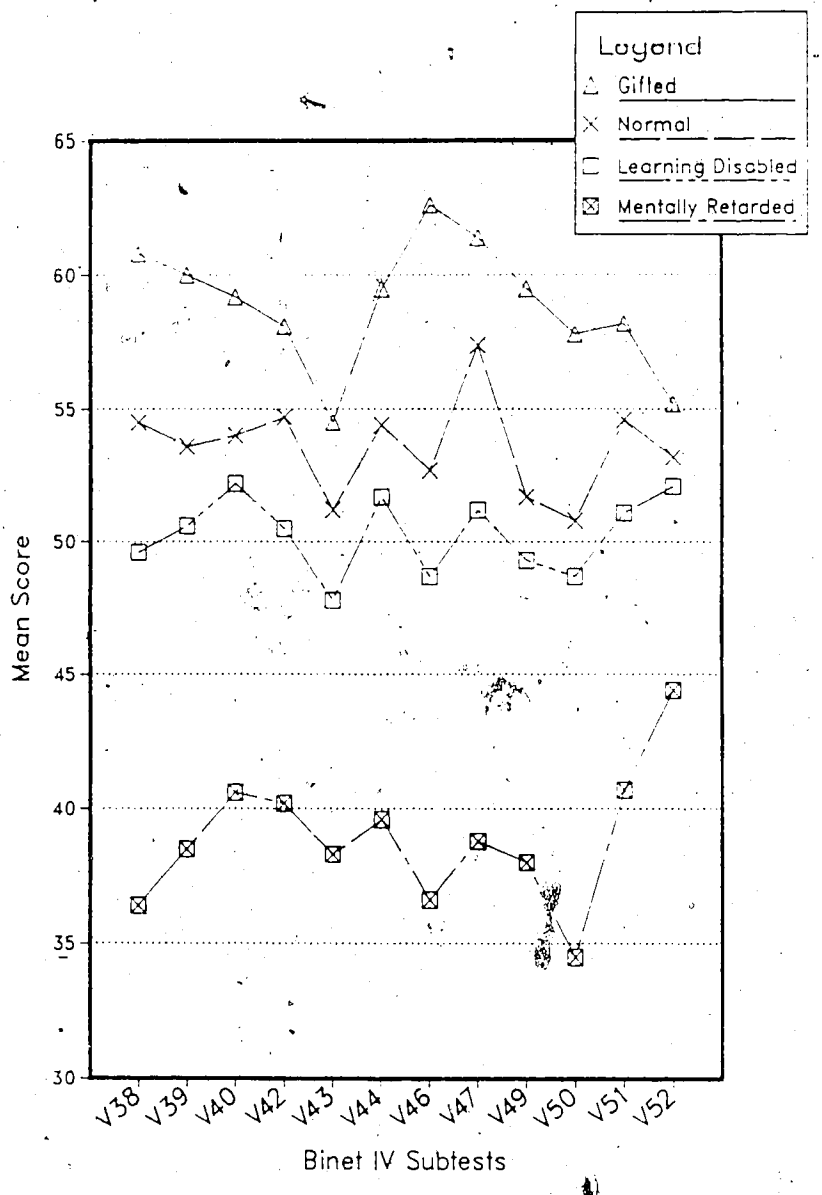

FIGURE 4.1

Graphical Representation of the Four Populations on Binet IV Individual Subtests

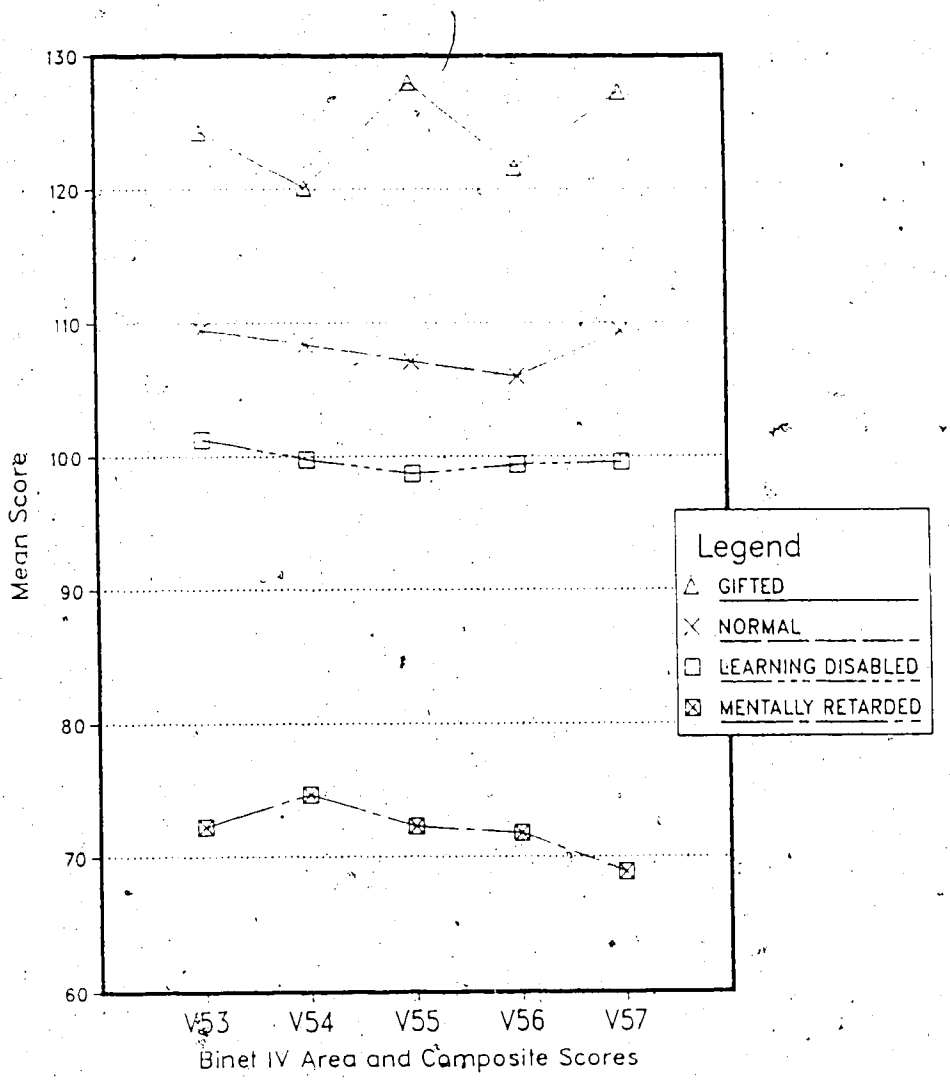

FIGURE 4.2

Graphical Representation of the Four Populations on Binet IV Area Scores and Composite IQ

Graphical Representation

Figures 4.1 and 4.2 give the reader a pictorial summary of the mean scores attained by the four populations.

In Figure 4.1, it can be seen that the Copying subtest is one of the lowest on the graph for all four groups and Matrices is relatively high for all the groups. Quantitative is highest for only the gifted group, whereas Number Series is high for the three groups of gifted, normal and learning disabled.

The peaks and the valleys depicted in this graph are notable - with the gifted group's peaks represented by Quantitative, Number Series and Vocabulary subtests. The normal group's peaks are indicated by Number Series, Vocabulary, Pattern Analysis, Matrices and Memory for Digits. The learning disabled group's peaks are represented by Absurdities, Matrices, Number Series, Memory for Digits and Memory for Objects. The mentally retarded group's peaks include Memory for Objects, Absurdities, Pattern Analysis and Memory for Digits.

The lower points on Figure 4.1 for the gifted group are represented by Copying, Memory for Sentences and Memory for Objects. For the normal group, lows are Memory for Sentences, Bead Memory, and Copying. For the learning disabled group lows are indicated by Copying, Quantitative, Memory for Sentences, Bead Memory and Vocabulary. Mentally retarded subjects scored low on Memory for Sentences, Vocabulary and Quantitative.

It is interesting to see that all four groups had their lowest mean scores on the Memory for Sentences and Copying subtests.

In looking at Figure 4.2, it appears that graphic representation of the four mean Area Scores and the mean Composite IQ score does not follow any particular patterns among the groups. For the gifted group the Quantitative Verbal Reasoning IQ score represents the highest mean score; for the normal and learning disabled subjects, Verbal

Reasoning IQ is the highest mean score, and for the mentally retarded group the Abstract/Visual Reasoning area is the highest mean score of the four Area Scores. The lowest scores represented in Figure 4.2 are Abstract/Visual Reasoning IQ for gifted; Short-Term Memory IQ for normal and mentally retarded and Quantitative Reasoning for the learning disabled group.

It is relevant for the purposes of this study to visually see and discuss mean scores for Binet IV variables according to the four groups. However, it is necessary to determine whether there are in fact significant differences among the mean scores of the four populations.

ONEWAY Anova

In order to determine whether these differences are significant between the four groups, ONEWAY analysis of variance was conducted on each of the variables. When significant F's were obtained on the ANOVA's appropriate Scheffe post hoc comparisons were carried out. The .05 level of significance was the criterion for all tests.

ONEWAY Anova results are reported in Appendix D. Tables 4.7 to 4.23 report mean scores and Scheffe post hoc significant differences between the four groups on Binet IV Composite IQ, Area Scores and Individual Subtests. The following abbreviations denote the four groups. MR = Mentally Retarded; LD = Learning Disabled; No. = Normal and Gif. = Gifted.

TABLES 4.7 - 4.23*

SCHEFFE TESTS OF COMPARISONS ON
BINET COMPOSITE IQ, AREA SCORES AND INDIVIDUAL SUBTESTS

TABLE 4.7
Composite IQ

Mean	Group				
68.93	MR				
99.68	LD	*			
109.45	No.	*	*		
127.36	Gif.	*	*	*	
		MR	LD	No.	Gif.

TABLE 4.8
Verbal Reasoning IQ

Mean	Group				
72.33	MR				
101.38	LD	*			
109.51	No.	*			
124.31	Gif.	*	*	*	
		MR	LD	No.	Gif.

*significance at the 0.05 level is denoted by *.

TABLE 4.9
Abstract/Visual Reasoning IQ

Mean	Group				
74.73	MR				
99.80	LD	*			
108.40	No.	*	*		
120.18	Gif.	*	*	*	
		MR	LD	No.	Gif.

TABLE 4.10
Quantitative Reasoning IQ

Mean	Group				
72.33	MR				
98.74	LD	*			
107.17	No.	*			
128.04	Gif.	*	*	*	
		MR	LD	No.	Gif.

TABLE 4.11
Short-Term Memory IQ

Mean	Group				
71.80	MR				
99.49	LD	*			
106.05	No.	*			
121.58	Gif.	*	*	*	
		MR	LD	No.	Gif.

The results indicate there are significant differences on the Binet IV Composite IQ and Abstract/Visual Reasoning IQ. Significant differences are reported for Verbal Reasoning, Quantitative Reasoning, and Short-Term Memory, for all groups, except for no significant difference between learning disabled and normal subjects on these three Area Scores.

TABLE 4.12
Vocabulary

Mean	Group				
36.40	MR				
49.69	LD	*			
54.57	No.	*	*		
60.86	Gif.	*	*	*	
		MR	LD	No.	Gif.

TABLE 4.13
Comprehension

Mean	Group				
38.53	MR				
50.62	LD	*			
53.65	No.	*			
59.95	Gif.	*	*	*	
		MR	LD	No.	Gif.

TABLE 4.14
Absurdities

Mean	Group			
40.66	MR			
52.29	LD	*		
53.96	No.	*		
59.26	Gif.	*	*	
		MR	LD	No. Gif.

TABLE 4.15
Pattern Analysis

Mean	Group			
40.26	MR			
50.55	LD	*		
54.71	No.	*		
58.19	Gif.	*	*	
		MR	LD	No. Gif.

TABLE 4.16
Copying

Mean	Group			
38.36	MR			
47.89	LD	*		
51.20	No.	*		
54.53	Gif.	*	*	
		MR	LD	No. Gif.

TABLE 4.17
Matrices

Mean	Group			
39.62	MR			
51.75	LD	*		
54.47	No.	*		
59.53	Gif.	*	*	
		MR	LD	No. Gif.

TABLE 4.18
Quantitative

Mean	Group			
36.66	MR			
48.71	LD	*		
52.76	No.	*		
62.68	Gif.	*	*	*
		MR	LD	No. Gif.

TABLE 4.19
Number Series

Mean	Group			
38.83	MR			
51.23	LD	*		
57.40	No.	*	*	
61.46	Gif.	*	*	
		MR	LD	No. Gif.

TABLE 4.20
Bead Memory

Mean	Group				
38.06	MR				
49.34	LD	*			
51.71	No.	*			
59.54	Gif.	*	*	*	
		MR	LD	No.	Gif.

TABLE 4.21
Memory for Sentences

Mean	Group				
34.50	MR				
48.71	LD	*			
50.82	No.	*			
57.85	Gif.	*	*	*	
		MR	LD	No.	Gif.

TABLE 4.22
Memory for Digits

Mean	Group				
40.77	MR				
51.17	LD	*			
54.61	No.	*			
58.26	Gif.	*	*		
		MR	LD	No.	Gif.

TABLE 4.23
Memory for Objects

Mean	Group				
44.40	MR				
52.10	LD	*			
53.28	No.	*			
55.21	Gif.	*			
		MR	LD	No.	Gif.

For the reported twelve Binet IV subtests only the Vocabulary subtest reported significant differences for all four groups. Comprehension, Quantitative, Bead Memory, and Memory for Sentences indicated significant differences between all four groups except between learning disabled and normal subjects on the four subtests. Absurdities, Pattern Analysis, Copying, Matrices and Memory for Digits report no significant differences between the learning disabled and normal groups as well as no differences between the gifted and normal groups. These above mentioned subtests only indicate significant differences occurring between the mentally retarded and the other three groups as well as differences between the learning disabled and gifted populations. Number Series differentiated three groups, but did not differentiate between the gifted and normal populations. Memory for Objects only differentiated the mentally retarded group from the other three populations. No significant differences were noted for learning disabled with normal, learning disabled and gifted or gifted and normal populations.

ONEWAY Anova's were conducted on the four groups and their performances on the WRAT-R, VMI and PASS. Tables 4.24 to 4.34 report these results.

TABLES 4.24 - 4.34*

SCHEFFE TESTS OF COMPARISONS ON

WRAT-R, VMI and PASS

TABLE 4.24
WRAT-R Arithmetic

Mean	Group				
72.83	MR				
89.73	LD	*			
103.19	No.	*	*		
123.63	Gif.	*	*	*	
		MR	LD	No.	Gif.

TABLE 4.25
WRAT-R Reading

Mean	Group				
68.00	MR				
92.68	LD	*			
110.30	No.	*	*		
121.57	Gif.	*	*	*	
		MR	LD	No.	Gif.

*significance at the 0.05 level is denoted by *

TABLE 4.26
WRAT-R Spelling

Mean	Group				
72.66	MR				
88.83	LD	*			
99.34	No.	*			
119.26	Gif.	*	*	*	
		MR	LD	No.	Gif.

TABLE 4.27
VMI

Mean	Group				
6.00	MR				
8.33	LD				
11.22	No.	*	*		
12.83	Gif.	*	*		
		MR	LD	No.	Gif.

TABLE 4.28
PASS Full Scale

Mean	Group				
37.00	MR				
42.10	LD				
51.20	No.				
44.16	Gif.				
		MR	LD	No.	Gif.

TABLE 4.29
PASS General Ability

Mean	Group
4.00	MR
6.39	LD
8.25	No. *
7.71	Gif.

MR LD No. Gif.

TABLE 4.30
PASS Arithmetic

Mean	Group
8.71	MR
8.36	LD
9.45	No.
9.71	Gif.

MR LD No. Gif.

TABLE 4.31
PASS School Satisfaction

Mean	Group
7.28	MR
7.71	LD
8.85	No.
7.50	Gif.

MR LD No. Gif.

TABLE 4.32
PASS Reading/Spelling

Mean	Group
5.50	MR
8.08	LD
10.00	No.
8.71	Gif.

MR LD No. Gif.

TABLE 4.33
PASS Penmanship/Neatness

Mean	Group
7.55	MR
8.14	LD
9.15	No.
5.85	Gif.

MR LD No. Gif.

TABLE 4.34
PASS Confidence

Mean	Group
4.14	MR
4.36	LD
5.45	No.
5.42	Gif.

MR LD No. Gif.

As reported in Tables 4.24 and 4.25, all four groups were significantly different on WRAT-R's Arithmetic and Reading subtests. WRAT-R Spelling differentiated between all the groups except no significant difference was indicated between learning disabled and normal subjects.

The VMI produced some interesting results and Table 4.27 reports there was no significant differences between learning disabled and mentally retarded populations; or the gifted and normal populations. There was however, significant differences reported between the mentally retarded and normal populations as well as differences between the learning disabled and normal populations.

On the PASS there were no significant differences among the four groups for the Full Scale score, or the following scales - Arithmetic, Penmanship/Neatness and Confidence. The only significant differences are reported for mentally retarded and normal subjects on General Ability (Table 4.29); and mentally retarded and normal groups on Reading/Spelling (Table 4.32).

Analysis of Subjects within the four Groups: Crosstabulations and Chi Square Analysis

To determine if other characteristics besides performance on IQ, achievement, visual-motor and self-concept tests defined subjects within the four groups, chi-square analyses were calculated on the following variables: gender, only child, order in family, parent marital status, father's occupation, mother's occupation, repeated grade and locale. Table 4.35 summarizes the results of the crosstabulations. The only two variables that showed significant differences among the four groups were Father's Occupation and Repeated Grade. None of the other variables were significantly different.

TABLE 4.35
Crosstabulation Variables

Reason for Referral	Chi Square	df	Significance
Gender	4.06	4	0.396
Only Child	.15	4	0.997
Order in Family	38.59	28	0.087
Parent Marital Status	14.32	24	0.938
Father's Occupation	63.40	28	0.000
Mother's Occupation	16.44	20	0.688
Repeated Grade	34.63	4	0.000
Locale	3.22	4	0.520

Conclusion

In this second section of this chapter the discriminative ability of Binet IV, WRAT-R, VMI and PASS among special populations was discussed utilizing different tables and figures to emphasize the results. Means and standard deviations were reported for the four groups of this sample by means of comparative tables and graphical representations. Canadian and American samples were compared on the basis of means and standard deviations of gifted, learning disabled and mentally retarded subjects. ONEWAY Anova's were reported showing the significant differences determined by Scheffe comparisons. Finally; crosstabulations were reported regarding characteristics other than IQ performance, achievement, visual- motor or self-concept for the four groups. Hypothesis seven is both accepted and rejected on the basis of this statistical information as some scores from the IQ, achievement, visual-motor and self-concept tests differentiated the four groups, but other scores were unable to do so.

CONVERGENT VALIDITY OF THE WRAT-R, THE VMI, AND THE PASS TO BINET IV

In order to evaluate the convergent validity of the WRAT-R, VMI and PASS to Binet IV correlations are reported. Hypotheses eight through 11 are examined in this last section.

TABLE 4.36

Correlation Matrix of WRAT-R subtests, VMI, PASS subtests; and Binet IV individual subtests; Area Scores and Composite IQ*

	V18	V20	V22	V24	V26	V27	V28	V29	V30	V31	V32	V38	V39	V40	V42	V43	V44	V46	V47	V49	V50	V51	V52	V53	V54	V55	V56
V20	70																										
V22	64	77																									
V24	64	70	67																								
V26																											
V27					70																						
V28					68	42																					
V29					68		44																				
V30			33		78	46	44	51																			
V31					74	41	38	47	41																		
V32					75	50	44	45	53	48																	
V38	62	66	56																								
V39	54	57	46									78															
V40	45	44	37									67	56														
V42	53	59	49			31						57	56	45													
V43	40	50	42	61		39						40	38	40	43												
V44	57	49	47	44								57	47	53	40												
V46	74	62	58	31								67	60	56	48	35	66										
V47	79	65	56	56								63	67	51	60	40	63	78									
V49	57	46	46									60	54	52	52	33	62	57	63								
V50	48	56	46									75	67	56	46	37	46	56	66	75							
V51	62	57	47	36								61	60	44	54	36	49	62	60	61	61						
V52	38	29	27	35								48	44	48	46		53	47	57	56	49	54					
V53	63	66	55									82	69	64	61	45	60	70	70	65	73	56	53				
V54	64	68	58			38						66	61	52	48	27	71	61	76	61	55	56	60	69			
V55	77	64	56	30								68	64	56	52	56	70	67	60	61	56	53	46	72	67		
V56	62	62	53									75	65	63	60	42	67	63	70	62	65	61	76	78	68	66	
V57	75	74	64	28								65	60	74	73	56	74	83	65	76	78	70	63	81	65	66	66

*decimal points have been omitted from this correlation matrix
**correlations of .40 and above are significant at the .001 level; correlations between .27 and .40 are significant at the .01 level; correlations with values less than .27 were omitted from this table.

Pearson Product Moment Correlations

Pearson Product Moment correlations were calculated between the 27 variables. Correlation coefficients with values less than .27 were omitted from the results. Correlations of .40 and above are significant at the .001 level and correlations between .27 and .40 are significant at the .01 level. The results are presented in Table 4.36, a 27 x 27 correlation matrix. The lower off-diagonal elements only are presented since the matrix is symmetrical. Decimal points have been omitted.

Hypothesis 8

The WRAT-R Reading and Spelling subtests will correlate significantly with Binet IV's Composite IQ, Verbal Reasoning IQ and Verbal Reasoning Subtests.

a. WRAT-R Reading and Binet IV Composite IQ, Verbal Reasoning IQ and Verbal Reasoning subtests: The data presented in Table 4.36 indicate WRAT-R Reading correlates positively with the Binet IV measures. It correlates quite highly with the Composite IQ (.74), relatively highly with the Verbal Reasoning IQ score (.66) and moderately with Vocabulary (.66), Comprehension (.57) and Absurdities (.46).

b. WRAT-R Spelling and Binet IV Composite IQ, Verbal Reasoning IQ and Verbal Reasoning subtests: WRAT-R Spelling correlations with the three main areas of Binet IV are all more modest correlations than WRAT-R Reading and Binet IV. WRAT-R Spelling correlates highest with the Composite IQ at .64 with the other r's ranging from .37 (Absurdities) to .55 (Verbal Reasoning IQ score). Therefore Hypothesis 8 is accepted.

Hypothesis 9

WRAT-R Arithmetic will correlate significantly with Binet IV Composite IQ, Quantitative Reasoning IQ, Short-Term Memory IQ, Quantitative and Number Series Subtests.

WRAT-R Arithmetic correlations with Binet IV are all high correlations. Number Series at .79 and Quantitative Reasoning IQ at .77 are the highest correlations, followed by Composite IQ (.75) and Quantitative (.74). WRAT-R Arithmetic and the Short-Term Memory IQ is a relatively lower correlation at .62 than the Quantitative IQ and subtest correlations but still significant.

Therefore Hypothesis 9 is accepted.

Hypothesis 10

The VMI will correlate significantly with Binet IV's Composite IQ; Abstract/Visual Reasoning IQ score; Copying, Pattern Analysis, and Matrices subtests.

The VMI correlates highest with Copying at .61, followed by Matrices with a modest correlation of .44, and a very low correlation with the Composite IQ at .28. Although expected, there was no significant correlation with Abstract/Visual Reasoning IQ (.24) or with Pattern Analysis (.17).

Therefore Hypothesis 10 is accepted on the basis of VMI correlating with Binet IV Composite, Copying and Matrices, but rejected on the basis of no significant correlation with Abstract/Visual Reasoning IQ or Pattern Analysis.

Hypothesis 11

The PASS will correlate significantly with Binet IV Composite IQ, Area Scores, and specific individual subtests.

A very modest correlation for the General Ability Scale and Binet IV's Abstract/Visual Reasoning IQ at .35 is reported. Correlations between General Ability and the subtests Copying at .39 and Pattern Analysis at .31 are also reported. No other

statistically significant correlations are reported between PASS and Binet IV Composite, Area Scores or individual subtests.

Therefore, Hypothesis 11 does not hold for correlations with Binet IV Composite, Verbal Reasoning IQ, Quantitative Reasoning IQ or Short-Term Memory IQ but does hold for correlations with Abstract/Visual Reasoning IQ, Copying and Pattern Analysis.

Conclusion

A number of general trends become evident as indicated by the data in Table 4.36. WRAT-R Arithmetic and Reading have consistently high correlations with Binet IV Composite, Area Scores and individual subtests with slightly lower correlations for WRAT-R Spelling. VMI correlates moderately with the Binet IV Copying subtest and has low correlations with the Composite IQ Score and Abstract/Visual Reasoning subtests. No significant correlation is present between VMI and the Abstract/Visual Reasoning IQ area. The PASS has fewer and lower statistically significant correlations with Binet IV Composite, Area Scores and individual subtests than the other instruments administered in the battery.

CHAPTER V

DISCUSSION

As a new psychometric measure, Binet IV is currently being used by qualified practitioners and supervised students in the field. The paucity of research regarding this assessment tool has led to this descriptive study being undertaken.

Restatement of the Study's Objectives

In Chapter One the framework for this descriptive study was outlined. Three major research questions were formulated as the basis of this study's orientation.

The historical development of the Stanford-Binet Intelligence Scale Form L-M was presented followed by a description of the Stanford-Binet Intelligence Scale-Fourth Edition as detailed in the technical manual. The very limited current literature was reviewed and hypotheses relating to the three major research questions were presented.

The design and procedures of this study were presented in the third chapter, with the results of the data analysis presented in chapter four.

Finally, this chapter will provide a discussion of the study's results, and their implications for use of the test and for future research.

Since Binet IV is a new testing instrument, to date little research has been reported other than the technical manual's few studies. Thus, when relevant studies are referred to in the following discussion there are no references for Binet IV. Rather, references will be made to other reliable and valid intelligence tests with which Binet IV correlated highly (Thorndike et al. 1986A). These authors report Binet IV Composite IQ correlations with the WISC-R Full Scale IQ at .83, WPPSI Full Scale at .80, WAIS-R Full Scale at .91 and K-ABC (Mental Processing Composite) at .89. These respectively high correlations serve to reinforce validation for Binet IV when references are made to these tests as well as the WRAT-R, VMI and PASS instruments used in this study.

Comparison of Canadian and American Data

There appeared to be many similarities between the Canadian and American samples with regard to the subject's performance on the Binet IV. The data indicated mean scores for both samples were very similar. The three groups of LD, MR and normal Canadian mean scores tended to approximate those of Naglieri (1985) who found that when comparing the performances of learning disabled, borderline mentally retarded and normal children on the WISC-R and K-ABC that the WISC-R scores for the normal population were FSIQ = 108.3; VIQ = 105.5, and PIQ = 109.9. K-ABC Composite Score = 108.8. For the LD population, scores were as follows: FSIQ = 96.8, VIQ = 92.6, and PIQ = 102.8. K-ABC Composite score for the LD sample: = 96.1. The MR scores were as follows: FSIQ = 75.1, VIQ = 70.8, and PIQ = 83.5. K-ABC Composite score for the MR group: = 77.8. These scores are similar to Fisk and Janzen (1981) who reported a mean score of 108.3 for the normal population and 97.2 for the LD population on the Lorge Thorndike, in their comparative study of learning disabled and normal populations utilizing numerous measures.

Reported standard deviations for the Naglieri (1985) and Fisk & Janzen (1981) samples were in the 11.0 to 12.0 range, whereas the Canadian sample standard deviations from this study were much higher, indicating much more variability than the above studies or the reported American standardization sample.

Further analysis of mean scores indicated there were similarities between Canadian and American samples but also pointed out significant differences between the two samples. Specific individual Binet IV subtests were significantly different from one another. These included - Vocabulary, Comprehension, Absurdities, Pattern Analysis, Matrices, Number Series, Memory for Digits and Memory for Objects. Significant differences were reported for Verbal Reasoning IQ and the Composite IQ between the Canadian and American samples. In analyzing these differences according to the

technical manual's (Thorndike et al. 1986A) factor structure, there did not appear to be a pattern of crystallized or fluid abilities as the basis for the differences in the two populations. Both significant and non-significant differences occurred on individual subtests, Area Scores and the Composite IQ that measure crystallized, fluid analytic and short-term memory factors. These reported differences may in part be due to differing testing procedures or the nature of the Canadian sample itself which was a more homogeneous sample than the American sample. Because of the homogeneity of the Canadian sample higher scores were more likely to occur. The differences may indicate actual differences between Canadian and American children. It is clear that much more research is needed in this area before any conclusions can be drawn.

There were no statistically significant differences between the four individual subtests of Copying, Quantitative, Bead Memory, or Memory for Sentences, or the Area Scores of Abstract/Visual Reasoning, Quantitative Reasoning or Short-Term Memory IQ. Reported differences between subtest, Area and Composite IQ scores were statistically significant. What is relevant to this study is that clinically the differences did not appear significant, thus suggesting the use of Binet IV with Canadian children and the use of the American standarization sample as being sufficiently equivalent. However, underlying all of this should be the practitioner's awareness that differences suggested in this study do exist. Only through replications of these kinds of studies employing larger, and more representative samples will further evidence indicate whether or not the differences are significant enough to affect Canadian children's ultimate IQ scores.

Intercorrelations between Verbal Reasoning subtests and Verbal Reasoning IQ scores; Abstract/Visual Reasoning subtests and Abstract/Visual Reasoning IQ scores; Quantitative Reasoning subtests and Quantitative Reasoning IQ scores; and Short-Term Memory subtests and Short-Term Memory IQ scores were all very similar to the American sample. Correlations between each of the Four Areas, as well as the

intercorrelations between subtests, Area Scores and the Composite IQ were relatively high as expected, and indicated a very high concordance with the American standardization group.

On the basis of this research, the Canadian and American samples appear to be very similar. The reported differences that occurred seem to be slight and perhaps could be attributed to sampling procedures or any number of reasons not necessarily related to the Binet IV itself. It is clear that more research is necessary in this testing area before any valid conclusions can be made.

Discriminative Validity

The technical manual (Thorndike et al. 1986A) discussed studies which indicated that the Binet IV can reliably discriminate between learning disabled and mentally retarded subjects. It also stated that the means of the gifted subjects on Area and Composite scores were significantly above the means attained by the standardization sample.

The results of this study indicated that there were significant differences between the three exceptional populations- learning disabled, mentally retarded and gifted on Binet IV. This study also examined these three populations and normal subjects, utilizing the other instruments from the test battery.

The claim that Binet IV can effectively discriminate between learning disabled and mentally retarded subjects was substantiated by this study. Significant differences were reported between the learning disabled and mentally retarded populations on all Binet IV subtests, Area Scores and Composite IQ scores. This is consistent with Kaufman's (1982) findings that learning disabled children have a unique profile, but contradict Naglieri's (1985) findings when performances of learning disabled, borderline mentally retarded and normal children were compared on the WISC-R and

K-ABC. Reported WISC-R scores for this were FSIQ: MR $\bar{X}$ = 75.1, SD = 11.5; FSIQ: LD $\bar{X}$ = 96.8, SD = 11.5. K-ABC Mental Processing Composite scores were: MR $\bar{X}$ = 77.8, SD = 10.0; LD $\bar{X}$ = 96.1, SD = 10.0. He found that the WISC-R and K-ABC subtest profiles differed from the normal subjects, but was unable to distinguish between the learning disabled and borderline mentally retarded children. Naglieri (1985) suggests that:

> the LD and borderline children performed best on subtests that have similar content and cognitive requirements. On the WISC-R, these samples earned their highest scores on Picture Arrangement and Object Assembly (which have been described as simultaneous tasks by Naglieri, Kamphaus, and Kaufman, (1983) and on the K-ABC subtests Gestalt Closure and Photo Series (both Simultaneous Scale tasks). Similarly, the exceptional children performed poorly on WISC-R and K-ABC subtests with sequential aspects (Digit Span, Number Recall, Word Order, and Hand Movements) and academic content (Information, Vocabulary, and Arithmetic on the WISC-R and K-ABC, and Reading), as well as on Matrix Analogies. This configuration of strengths and weaknesses illustrates the complexity of the exceptional children's performance and why a simple simultaneous-sequential discrepancy did not adequately describe the samples.

One possible explanation for this study's results in comparison to Naglieri's (1985) study is that Binet IV in structure and material content is measuring different facets than the WISC-R or K-ABC and is thus able to adequately distinguish between LD and MR populations. It seemed to be particularly convincing that in this study all the subtest, Area and Composite scores differentiated the LD and MR populations. One must be cautious in making a definitive statement regarding scores for the mentally retarded populations as a number of studies point out major discrepancies for mentally retarded children. Cummings & Sanville (1983) compared the means for a sample of EMR children between the Woodcock Johnson Tests of Cognitive Ability (WJTCA) and the WISC-R. WISC-R's Full Scale mean of 69.4 was 16.7 points higher than the WJTCA Broad Cognitive mean of 52.7. This difference of over one standard deviation

between the tests not only has statistical significance, but according to the authors "the magnitude of this discrepancy also has important practical implications. Relatively speaking, the WJTCA would result in substantially more children being classified as eligible for placement in a class for mentally retarded children." In another study by Kaufman & Van Hagen (1977), 80 mentally retarded children were evaluated on the WISC-R and Stanford-Binet to assess the continuity of measurement between the old and new WISC's. Correlations of Verbal, Performance and Full Scale IQ's with Binet Form L-M (1972 norms) were .73, .65 and .82 respectively. These values were very similar to the former correlations. What was of extreme importance however, was that the 1972 Binet Form L-M of 53.8 was seven points higher than the WISC-R Full Scale mean of 46.8 and showed statistical significance. This result was not the same as Wechsler's results with normal children which showed Binet and WISC-R 's having comparable IQ scores. Kaufman & Van Hagen (1977) in determining whether a discrepancy exists between Binet Form L-M and WISC-R FS IQ point out that "if a difference as large as one half of a standard deviation really does exist, practical problems will arise regarding the classification of children s intellectual abilities . . . all of the disagreements were in the same direction, with the Binet classifying the child at a higher level than the WISC-R." It would appear that a comparison of scores between Binet IV and WISC-R, particularly with retarded subjects as well as other exceptional and non-exceptional populations, would shed more light on the discrepancies outlined in the previous studies. This research is currently being undertaken by graduate students in the Educational Psychology department, University of Alberta (1987, personal communication) and results of their studies will certainly be a welcome addition to this area which tends to reflect so many conflicting results.

It is unfortunate that the authors of Binet IV were unable to cite any studies reporting any significant differences between LD and normal populations. The findings of Epps, McGue & Ysseldyke (1982) found that there is considerable doubt that

school personnel are able to accurately or reliably identify LD students. With the entire classification system for learning disabled children being such a grey area it would have been beneficial to have information from the authors of Binet IV to indicate that the test was capable of differentiating an LD population from a normal population. This information however was not provided.

In this study significant differences were noted between LD and normal subjects on Binet IV's Composite IQ, Abstract/Visual Reasoning IQ, Vocabulary, and Number Series subtests. Other Area Scores and subtests gave no indication of significant differences between these two groups. It was interesting that Short-Term Memory tests did not discriminate between LD and Normal populations. It would be of interest to speculate that Fluid/Crystallized Factors were evident in LD/Normal differences but from this study insufficient evidence was available. Too little information was available to warrant any conclusions in this area, and a great deal more research is necessary to make any general statements regarding Binet IV patterns between LD and normal populations. In a study referred to earlier by Naglieri (1985) comparing the performances of LD, borderline MR and normal children on the WISC-R and K-ABC it was noted that LD and borderline MR children were not distinguished. It is interesting to note that in this study, Naglieri suggests that LD children had a WISC-R subtest profile that distinguished them from the normal sample. This finding is consistent with Kaufman's (1982) conclusions that LD children do have a characteristic profile but contradicts Goh & Simons (1980) who report no specific pattern as typical for LD children.

The technical manual (Thorndike et al. 1986A) claimed discriminatory power of the Composite and Area Scores between gifted and normal populations. This was substantiated by this study. In addition to these reported mean scores, this study further delineated gifted from normal populations on Binet IV individual subtests. There were significant differences reported between the gifted sample and the other

three populations on six Binet IV subtests in this study. Although expected, no significant differences were noted between gifted and normal populations on Absurdities, Pattern Analysis, Copying, Matrices, Number Series or Memory for Digits. It was particularly interesting that Memory for Objects was unable to differentiate the gifted population from either LD or normal populations. One might question exactly what Memory for Objects is measuring and why it does not seem to have the same discriminative ability as other Short-Term Memory subtests - such as Bead Memory or Memory for Sentences. Referring back to Chapter two, Osberg (1986) had pointed out that Bead Memory is the only memory subtest tapping simultaneous memory functions, whereas, the other three memory subtests were indicators of sequential short-term memory. The results of this study indicate an unclear picture as to the kind of memory functions each one of the memory subtests are tapping and their discriminative ability does not appear to be as straightforward and clear cut as the other Area Scores. More investigation in this area needs to be done before any conclusive resolutions can be made.

The LD population means from this study with regard to the WRAT-R subtests were consistently higher than Brock's (1982) findings in determining the factor structure of intellectual and achievement measures for LD children. Reported WRAT means from Brock's (1982) study were: Spelling $\overline{X}$ = 80.8, and Arithmetic $\overline{X}$ = 85.9. In another study Haddad (1986) reports mean scores for LD subjects on the WRAT as follows: Spelling $\overline{X}$ = 81.3, Arithmetic $\overline{X}$ = 86.8 and Reading $\overline{X}$ = 82.7. These compare to this study's LD means on the WRAT-R which are Spelling $\overline{X}$ = 88.8 , Arithmetic $\overline{X}$ = 89.7 and Reading $\overline{X}$ = 92.6. Differences between the learning disabled and mentally retarded groups were noted on all three of the WRAT-R subtests. These findings are in contrast with Eno & Woehlke (1980) who determined that WRAT scores for LD and MR subjects did not differentiate the two groups.

Significant differences at the .05 level were reported between the LD and normal groups on WRAT-R Arithmetic and Reading. No significant differences were noted for WRAT-R Spelling. These findings were higher than Fisk & Janzen's (1981) findings which compared the two exceptional populations on an entire test battery. Their findings indicate significant differences of $p < .10$ to .30 found on WRAT Spelling and Arithmetic between the LD and normal groups. Breen's (1983) results show that all three WRAT subtests significantly discriminated between MR and LD students as well as between MR and regular education students but does not report differences between LD and regular education students.

As expected, the gifted population showed significant differences from the other three populations on WRAT-R Reading, Spelling, and Arithmetic at the .05 level of significance. Mean scores attained by the gifted population on the WRAT-R subtests in this study were somewhat lower for Reading and Spelling but higher on Arithmetic than Karnes, Edwards & McCallum's (1986) WRAT means of 132.5, 122.4, and 113.6 respectively.

No significant differences between the LD and MR populations were noted on the VMI. These findings contradict the results of Crofoot & Bennett (1980) who suggest that the VMI does give some indication of a child's performance on verbal/visual-motor and receptive/expressive dimensions, but fails to tap verbal expression and attention/concentration. Similar to Crofoot & Bennett's (1980) findings it is of particular interest that in this study Abstract/Visual Reasoning IQ scores and Abstract/Visual subtest scores on Binet IV were statistically different for learning disabled and mentally retarded groups. From these results one might conjecture that the VMI is measuring different skills than Binet IV. It is interesting to note that no significant differences were reported on the VMI between the LD and MR groups, but significant differences were noted between the LD and normal groups.

On the PASS no significant differences were reported between the four designated groups except between MR and normal populations on General Ability and Reading/Spelling scales. These findings concur with Neufeld & Cozac's (1980) study in which they compared the self concept of gifted students with average students on TIPS (Teacher's Inventory of a Pupil's Self-Concept), WISC-R and CTBS (Canadian Test of Basic Skills). Analysis of their data indicated that no significant relationship existed between self concept and IQ performance, reading comprehension, math achievement and overall composite achievement.

CONVERGENT VALIDITY

WRAT-R

In this study correlations between the WRAT-R Reading, Spelling and Arithmetic subtests and Binet IV Composite and Area Score IQ's ranged from the high 70's to mid 50's. These correlations were consistently higher than the results of Wright and Dappen (1982) whose correlations between WRAT subtest scores with WISC-R IQ's ranged from .56 between Full Scale IQ and Arithmetic to a low of .30 between Performance IQ and Reading. Smith & Smith (1986) found that in their study correlations ranged from between .64 between Verbal IQ and Arithmetic to .21 between Performance IQ and Spelling. Haddad (1986) reported correlations that ranged from .46 between Full Scale IQ and Arithmetic to .33 between Performance IQ and Spelling.

In this study individual Binet IV subtest correlations ranged from a high of .79 between Quantitative and Arithmetic, to a minimum of .27 between Memory for Objects and Reading. All of the individual subtests had correlations of between .40 and .79 with the exception of Memory for Objects where correlations were all below .40, and

Absurdities and Spelling had a correlation of .37. This range of findings tends to be higher than the research of Brock (1982) whose results showed a maximum correlation of .44 between WISC-R's Vocabulary and WRAT Spelling and a minimum correlation between WISC-R's Picture Completion and Spelling at .01.

It seems indicated from this study that the significantly high correlations reported between the WRAT-R subtests and Binet IV Composite IQ and Area Scores are not only consistent with other studies but with the higher correlations reported gave even stronger validation for the convergent validity between WRAT-R and Binet IV.

VMI

In this study the VMI had moderate (Copying) to modest (Matrices) correlations with Binet IV subtests and a non significant correlation of .24 with the Abstract/Visual Reasoning IQ score. It was interesting to note that an unexpected high correlation between VMI and Number Series occurred. It would appear from these findings that the .24 correlation between the VMI and Abstract/Visual Reasoning IQ is low and thus the researcher would not recommend that an Abstract/Visual Reasoning IQ score be used as a substitute for the Visual Motor Integration Test. These findings are similar to Breen, Carlson & Lehman as cited in Siewert & Breen (1983) whose scores yielded significant correlations between WISC-R Performance IQ and VMI, of .33 through to .42. Correlations with WISC-R Verbal IQ and the VMI did not generate any significant values.

PASS

The PASS General Ability scale and Binet IV's Abstract/Visual Reasoning IQ, Pattern Analysis and Co[illegible]ng reported very modest positive correlations. (.31 to .39). No other subscale on the PASS indicated any significant correlations with Binet IV. These few low significant correlations are consistent with the research of Simon & Simon (1975) who explored the relationship between self esteem or self concept as

measured by Coopersmith's Self-Esteem Inventory (SEI), academic achievement as measured by the SRA Achievement Series and the Lorge-Thorndike Intelligence Test. Correlations between SEI scores and SRA were found to be .333 ($p < .01$). Correlation coefficient between SEI scores and Verbal IQ was .299 ($p < .01$) and .232 for non verbal IQ scores. A number of studies as cited in Simon & Simon's (1975 study, Fink, 1962, Stevens, 1956, and Williams & Cole, 1968) consistently reported significant positive relationships between self esteem and academic achievement. Primavera, Simon & Primavera's, (cited in Simon & Simon, 1975) findings indicated positive correlations between self-esteem and each of the five subtests of the Stanford Achievement Tests.

Conclusion

On the basis of the data presented, the convergent validity of Binet IV with WRAT-R, VMI and PASS was demonstrated.

Administration and Interpretation of Binet IV.

While the presentation thus far focused on an analysis of the data relating to the research hypotheses, it is relevant at this point to examine the positive and negative aspects of the administration and interpretation of the test as an additional facet of the total evaluation of this instrument.

Administration

Binet IV has an attractive, colorful, pictorial format which seems to pique the client's attention promoting rapport and interest. The wide variety of activities and tasks motivates the client's interest. The examiner can find items on some subtests to challenge a wide range of intellectual abilities. The layout of visual stimuli seems to assist the client with both visual and auditory questions. The triangular stand-up books

facilitate ease in administering as the examiner has the understandable and well-written directions, answers and scoring criteria directly in front of him. Questions can be read verbatim from the book in a natural manner. In addition to the facilitating style of the four books they are easy to follow, sequential and well organized. Similarly the protocol is well laid out allowing ample space for answers, comments and observations. Since specific entry levels are at the client's ability level rather than their chronological level, in most cases unnecessary items are not administered. At any time a person's entry level can be adjusted given the person's performance.

There appear to be adequate sample, demonstration and practice items as well as visual aids. In most of the subtests this facilitates ease in administration.

Administratively there are a number of general criticisms as well as specific difficulties with subtest administration.

The "entry level" as decided by Vocabulary is not consistent and can be an unreliable estimate of where to begin testing. The backtracking that occurs in order to ascertain the appropriate entry level for skills other than Vocabulary make for confusion and uncertainty. The basic assumption of using Vocabulary as an entry level needs more research to determine its appropriateness and suitability as well as its reliability and validity.

The lengthy administration time clearly can fatigue both client and examiner. Particularly when administering the test to gifted, high achievers and verbose clients excessive time can be spent reaching the desired level of performance.

Even with continued practice, there are difficulties establishing basals and ceilings. Once a basal has supposedly been established it is difficult to back up if on another subtest the basal is missed on the entry level. The choice of which samples to administer when backing up in these particular circumstances can be very confusing.

At times it is difficult to find the correct pages to start each test and much page flipping increases the lack of ease in testing.

The criteria for the length of time allowed on specific items is vague, creating unbalanced time allowances.

A problem related to administration, yet unrelated to testing is the extremely heavy carrying case. Clinicians who travel to do assessments find this problem to be an unnecessary burden which could probably be easily rectified.

There are specific American content items which pose difficulties for this Canadian sample specifically and generally for Canadian subjects. Revision or supplementary items for Canadian consumption include the following:

A. Verbal Reasoning Area

Vocabulary: Item No. 8 - American Flag.

Comprehension: Item No. 38 - Bill of Rights

Absurdities: Item No. 16 - Address is American

Item No. 26 - Imperial measure rather than Metric.

B. Quantitative Reasoning Area

Quantitative : Item No. 23 - American coins.

Item No's 16, 24, 27, 32,33,34,35,36,37,38,39 - All Imperial measures rather than Metric.

In addition to American content in the above examples being a disadvantage to Canadian children there are some other items which seem to pose problems for any clients being administered the Binet IV.

In Quantitative items 22 and 30 the wording is very awkward, making the correct response unclear. Quantitative item number 17 is placed too early in the test. There is a major jump from item16 (looking at a pencil that measures 6 inches) to having a working knowledge of fractions or percentages.

Verbal Relations item number 1 has "scarf, tie, muffler, not shirt" as the group of words to analyze. Muffler is a very uncommon term with few people today knowing it as something to wear as opposed to a piece of equipment on a car.

In Memory for Sentences, Level L, item 24 and Level M, item 26 utilize complex words unfamiliar to a young child, yet at the same levels L and M, items 23 and 25 seem to be suitable for measuring a child's memory skills, rather than their semantic proficiency. There are inconsistencies again when looking at Level N - item 27 : "Chris chased the dog around the house but did not catch it", and item 28: "The birds were flying and singing when Lynn got up this morning"; in which the content is at a simpler level than item 26 of the lower level M - "Sensing defeat, the fighter's manager threw the towel into the ring."

There needs to be an attempt to rectify these general and specific administrative difficulties to improve the quality and suitability of Binet IV for Canadian subjects as well as any clientele.

Interpretation

By giving the clinician four Area Scores and a Composite IQ score, Binet IV provides a very comprehensive, holistic picture of the client. It is most valuable to have the subtest scores and the variety of scores in each Area adds to the credibility of results and interpretation.

The Short-Term Memory Area is very comprehensive and for the most part easy to interpret. It provides much more information on short-term memory than the WISC-R. STM is broken down into visual/auditory clues as opposed to just digit span on the WISC-R.

On the Quantitative subtests, which are untimed, written calculations are permissible, helping to differentiate math ability from the concentration/anxiety factor. The Quantitative area gives a good mathematical sampling particularly for older children.

Binet IV demonstrates its capacity to differentiate academic influenced learning (crystallized intelligence) from that influenced by general experiences (fluid intelligence).

Binet IV is applicable to a wide age range and thus can give much needed input into the assessment of mentally retarded children at one end of the continuum to gifted children at the opposite extreme.

Although Binet IV offers new and varied alternatives to other intelligence measures it too has pitfalls that make for unclear interpretation of the results.

As with any new test, a major area of concern is what the test is really measuring. The technical manual (Thorndike et al. 1986A) provides no information to the clinician and with no other guidelines from any other sources interpretation of subtests and Area Scores is very unclear. In conjunction with this lack of clarity regarding what the tests measure is a difficulty encountered when comparing scores across areas. How many points on SAS differences are required to indicate significant discrepancies between two subtests or among the four Area Scores?

Examiners need to be particularly cautious when a number of estimated scores are present and in need of interpretation. Concern with the validity of estimated scores is indeed a real one. Similarly when only one, for example, out of three tests in an Area is administered, how valid is this one result? Guidelines for scoring in these kinds of circumstances are an absolute necessity.

Although confidence intervals are available from Table F.1 of the technical manual (Thorndike et al. 1986A), it would seem logical to include this information in the Guide for Administering and Scoring the Fourth Edition (Thorndike et al.1986B).

With no distractibility factor as presented in the WISC-R's third factor, there is some question as to the reliability and validity of the scores of an overanxious, highly distractible young child. If for example, STM scores of a pre-schooler of this nature

were low - how could one determine whether results were due to distractibility or memory factors as the two areas do not appear to be one hundred percent related?

A number of the individual subtests of Binet IV seem to have inherent problems that indicate a need for improvement. The Copying subtest from all indications seems to create the most difficulties for interpretation. The scoring is so stringent that it is inconsistent with the more lenient scoring procedures of a reliable and valid visual-motor instrument such as the VMI. A less than perfect design in the Copying subtest may not necessarily indicate visual-motor difficulties, but rather could be an indication of a subject hurriedly copying a design placed in front of him. A "messy" design may only be indicative of a particular approach to tasks rather than a specific visual or perceptual problem. Concerns with this subtest appear particularly obvious when gifted subjects attain low Copying scores and all other scores fall into the superior range. It is questionable as to whether this subtest should remain as it is.

Another aspect of Copying that is of concern is that the copying of the three-dimensional blocks is a hands on visual-motor coordination skill. Levels A-F of the Copying subtest seem to be measuring very different skills than Levels G-N where copying of a design using paper and pencil is required.

In Pattern Analysis, the designs to be copied are reversed in the book for the examiner and examinee making it unnecessarily difficult to ascertain whether the examinee has properly executed the task. Another concern with Pattern Analysis is that the highest SAS for this subtest is 57 at age 14-11 and up, even if all items are answered correctly. This appears to be inconsistent with other subtest scoring.

The inclusion of the Absurdities subtest in the Verbal Reasoning Area is diametrically opposed to a conceptually similar test (Picture Completion) which is included in the Performance score of the WISC-R. Given the longevity, reliability and validity of the WISC-R should Absurdities continue to be considered part of the Verbal Reasoning Area?

The jump in the Matrices subtest from the sample item to Level O seems to be too much for even subjects with a strength in this area.

Difficulties in these general areas and particular subtests are indicative of a need for further research in improving the quality of the overall test.

Binet IV was administered by thirty-three student clinicians attending the Fall/Winter Session of the Educational Psychology course in assessment at the University of Alberta. Sixty-eight percent of them responded to a questionnaire by Janzen, Boersi & Krausher (1987) concerning Binet IV. Their comments confirmed the researcher's opinions regarding the strengths and weaknesses of the test discussed in this chapter. A summary of their responses follows.

A. Four Major Strengths of Binet IV	Percentage of Responses
1. Better sampling of memory area.	72
2. Wide Age range (2-18 years)	68
3. Obtaining four Area Scores plus Composite	63
4. Ease of testing format	59

B. Four Major Weaknesses of Binet IV	
1. Lengthy administration time	68
2. Difficulty in scoring some tests	63
3. American items	50
4. Inadequate SEM and confidence intervals	27

C. Client's Response to Binet IV	
1. Enjoyment of pictorial format	54
2. Moderate enjoyment	50
3. Much enjoyment	40
4. Held attention well	36
5. Found the test lengthy	31

D. Age Range Most Suitable for Binet IV Administration	Percentage of Responses
1. 6-12	77
2. 13-18	22
3. 2-5	9

E. Difficulties Encountered During Administration

1. Backing up after choosing entry level	63
2. Marking the Copying test	59
3. Finding the appropriates pages in books	36

F. Difficulties Encountered While Interpreting

1. Interpreting if significant differences between Area Scores exist	45
2. Knowing what the subtests are measuring	36
3. Interpretation of subtest patterns	31

G. Choice of Binet IV Over Other Individual IQ Tests

1. In order to examine the memory area in more detail	86
2. Being a "power test" with all subtests untimed except for Pattern Analysis	68
3. Use with young children (2-5 years)	54

H. Comparison of Binet IV With The WISC-R

1. Ease of Administration	
Binet IV	18
WISC-R	59
Equal	18
2. Ease of Interpretation	
Binet IV	36
WISC-R	59
Equal	18

H. (Continued) Percentage of Responses

3. Test that gives more Information

Binet IV	45
WISC-R	13
Equal	27

I. Subtests in Need of Revision for a Canadian Population

1. Quantitative	54
2. Comprehension and Quantitative	22

Implications for Future Research

1. The Canadian/American similarities and differences were reflected in this study. The test, as would be expected, is heavily laden with American content. Additionally as indicated by a comparison of the two samples, significant performance differences on specific subtests were reported. For purposes of evaluating Canadian children, the effects of the large number of American items need to be researched further to determine whether they affect the reliability or validity of the IQ scores for the Canadian children.

2. Data from this study indicated that Binet IV does differentiate between clinical groups. It is very significant that Binet IV gives clinicians a reliable tool in differentiating LD children from mentally retarded children. However, there are too few profiles that are strong in identifying LD children. We would need to define the subpopulations of LD children before we could develop tests or subtests to identify these children. Research is needed to be able to reliably differentiate this population from a normal population.

3. This study indicated acceptable correlation coefficients between Binet IV and the WRAT-R. More research with larger samples are necessary to compare results of this study with other studies and verify the results.

Educational Implications

For educational psychologists, there appear to be certain Binet IV subtests and Area Scores that distinguish learning disabled from normal subjects. These subtests included Vocabulary and Number Series; the Abstract/Visual Reasoning Area score as well as the Composite IQ score. Scores that practicing psychologists should be aware of include the following: High subtest scores in this study for LD subjects included Absurdities, Matrices and Memory for Objects. Low subtest scores included Copying, Quantitative and Memory for Sentences.

There were similar patterns of scores for normal subjects except that Absurdities was not a high scoring subtest for normal subjects and Memory for Objects was a low scoring subtest compared to other subtests for normal subjects.

In distinguishing gifted subjects from normal subjects, the gifted sample in this study had very high Quantitative and Number Series scores as well as high scores on Vocabulary, Comprehension and Absurdities of the Verbal Reasoning Area. The two lowest subtests for the gifted sample, although well above average, were on Copying and Memory for Objects.

As previously discussed, all four populations had their lowest score on Copying. Thus, when confronted with a low Copying score one should be cautious in interpretation as it did not appear to discriminate between any of the four populations.

The mentally retarded subjects in this study attained their highest scores on three short-term memory tasks and yet in studies cited in the manual (Thorndike et al. 1986A) STM was the lowest Area Score for MR's. Since the Area Score is the only score reported it is not clear how the STM subtests broke down into scoring for the

MR's. In this study, MR's had their lowest score on Memory for Sentences. Although all the MR subject's scores were significantly lower than the other three populations as one would expect, their test scores tended to follow very similar patterns as noted in Figure 4.1.

These subtest score differences and patterns thus indicated could serve as beginning guidelines for practicing clinicians and educational psychologists in differentiating the four populations. Certainly, further research in this area is warranted before any conclusive statements could be formulated.

Summary

In this descriptive study of the Stanford-Binet Intelligence Scale: Fourth Edition (Binet IV), the major focus centered on three main questions and eleven proposed research hypotheses. The first question compared this Canadian sample with the American standardization sample in terms of similarities and differences based on mean scores, standard deviations and t-tests. Results indicated similar mean scores for both populations in terms of Binet IV's Composite IQ and Area Scores with both population means falling into the 100 range. Standard deviations varied for both samples with the Canadian sample having standard deviations approximately five points higher than the reported American sample.

Significantly high intercorrelations between subtests, Area Scores and Composite IQ scores were reported in this study and reflected very similar results that were reported for the American standardization sample.

T-tests indicated that similarities and differences existed between the American and Canadian samples on Binet IV subtests, Area Scores and the Composite IQ. Significant differences were reported between the two populations on Vocabulary, Comprehension, Absurdities, Pattern Analysis, Matrices, Number Series, Memory for Digits, Memory for Objects, Verbal Reasoning IQ and the Composite IQ score. No

significant differences were reported between the Canadian and American samples on Copying, Quantitative, Bead Memory, Memory for Sentences, Abstract/Visual Reasoning IQ, Quantitative Reasoning IQ, or the Short-Term Memory IQ.

Oneway Anova's and subsequent Scheffe post hoc comparisons were conducted on the designated four groups of learning disabled, mentally retarded, gifted and normal populations to determine if significant differences existed among the four groups on Binet IV and the other instruments utilized in this study. Examination of the results indicated significant differences between learning disabled and mentally retarded subjects on all Binet IV subtests, Area Scores and Composite IQ scores and all three WRAT-R subtests. No significant differences were reported between learning disabled and mentally retarded subjects on the VMI or any PASS subtest.

The gifted population differed from the other three populations on Binet IV Composite, all four Area Scores and Vocabulary, Comprehension, Quantitative, Bead Memory and Memory for Sentences subtests. No significant differences were reported between the gifted and normal subjects on Absurdities, Pattern Analysis, Copying, Matrices, Number Series, Memory for Digits, or Memory for Objects subtests.

WRAT-R Reading and Arithmetic differentiated between all four groups. WRAT-R Spelling and the VMI failed to differentiate between gifted and normal subjects. The PASS did not differentiate the gifted from the other populations on any of the subtests.

The convergent validity of Binet IV was demonstrated by statistically significant correlations reported between all Binet IV scores and the WRAT-R (correlations ranged from the 40's to high 70's). Fewer significant correlations were reported between Binet IV and the VMI. Significant correlations that were indicated by the results fell into the range of high 20's to low 60's. Very moderate or no significant correlations were reported between Binet IV and PASS. All of these correlations are consistent with other studies.

Results of this study were discussed in terms of the proposed research hypotheses, the relation of the findings to past research, and also the positive and negative aspects of Binet IV in terms of administration and interpretation.

To determine whether a new test is necessarily a better test will take much time and a great deal more research. Binet IV seems to offer new and perhaps more inciteful methods of assessing the performance of subjects in a variety of areas. However, as is the case with other instruments that have gone through the metamorphosis of a major revision, refinements are deemed necessary for Binet IV. With further revisions, Binet IV seems to have the potential to be utilized by psychologists with confidence as a valid and reliable choice in the evaluation of an individual's mental ability.

REFERENCES

REFERENCES

Alberta Education. (1986). Review of Issues in Intelligence Test Use in Alberta Schools. Special Educational Services.

Barnard, [illegible] (1987). Considerations Regarding the Use of the New Stanford-Binet. The Alberta School Psychologist, 7 (1), 16.

Beery, K. (1967). Developmental Test of Visual-Motor Integration: Administration and Scoring Manual. Chicago: Follett Publishing Co.

Boersma, F. & Chapman, J. (1985). Manual for the Student Perception of Ability Scale. Edmonton: University of Alberta.

Breen, M., Carlson, M. & Lehman, J. (1983). The Revised Developmental Test of Visual-Motor Integration: Its Relation to the VMI, WISC-R, and Bender Gestalt for a Group of Elementary Aged Learning Disabled Students. Journal of Learning Disabilities, in press.

Brock, H. (1982). Factor Structure of Intellectual and Achievement Measures for Learning Disabled Children. Psychology in the Schools, 19 (3), 297-304.

C.A.C.L.D. (1986). Please Listen, We Have Something to Tell You.

Carvajal, H. & Weyland, K. (1986). Relationships Between Scores on Stanford-Binet IV and Wechsler Intelligence Scale for Children-Revised. Psychological Reports, 59, 963-966.

Crofoot, J. & Bennett, T. (1980). A Comparison of Three Screening Tests and the WISC-R in Special Education Evaluations. Psychology in the Schools, 17 (4), 474-478.

Cummings, J. & Sanville, D. (1983). Concurrent Validity of the Woodcock-Johnson Tests of Cognitive Ability with the WISC-R: EMR Children. Psychology in the Schools, 20 (3), 298-303.

Eno, L. & Woehlke, P. (1980). Diagnostic Differences Between Educationally Handicapped and Learning Disabled Students. Psychology in the Schools, 17 (4), 469-473.

Epps, S. McGue, M. & Ysseldyke, J. (1982). Interjudge Agreement in Classifying Students as Learning Disabled. Psychology in the Schools, 19 (2), 209-220.

Evans, P. & Richmond, B. (1976). A Practitioner's Comparison: The 1972 Stanford-Binet and the WISC-R. Psychology in the Schools, 13 (1), 9-14.

Fagan, T. (1986). Stanford-Binet Draws Board Attention. Communique, 15 (3), 1-2.

Fedun, G. (1981). A Study of the Validity of Using the Lorge-Thorndike and Raven's Standard Progressive Matrices in the Identification of Underachievers. Unpublished Master's thesis, University of Alberta.

Fisk, R. & Janzen, H. (1981). Identifying Learning Disabled Students with a Selected Psychoeducational Test Battery. The Alberta Journal of Educational Research, 27 (3), 252-263.

French, L. (1986). MR Testing and Evaluation: New Dimensions and Old Concerns. Psychology in the Schools, 23 (1), 64-74.

Goh, D. & Simons, M.R. (1980). Comparison of Learning Disabled and General Education Children on the McCarthy Scales of Children's Abilities. Psychology in the Schools, 17 (4), 429-436.

Haddad, F. (1986). Concurrent Validity of the Test of Nonverbal Intelligence with Learning Disabled Children. Psychology in the Schools, 23 (4), 361-364.

Hill, W. (1985). Learning: A Survey of Psychological Interpretations. Fourth Edition. New York: Harper & Row Publishers.

Janzen, H. (1986). The Stanford-Binet IV - Subtests & Administration Guidelines, Unpublished manuscript.

Janzen, H., Boersma, F. & Krausher, R. (1987). Report on the Binet IV. Unpublished manuscript.

Jastak, J. & Wilkinson, G. (1984). The Wide Range Achievement Test-Revised: Administration Manual. Revised Edition. Wilmington, Delaware: Jastak Associates, Inc.

Jensen, A.J. (1980). Bias in Mental Testing. New York: Free Press.

Karnes, F., Edwards, R. & McCallum, S. (1986). Normative Achievement Assessment of Gifted Children: Comparing the K-ABC, WRAT, and CAT. Psychology in the Schools, 23 (4), 346-352.

Kaufman, A. & Van Hagen, J. (1977). Investigation of the WISC-R for Use with Retarded Children: Correlation with the 1972 Stanford-Binet and Comparison of WISC and WISC-R profiles. Psychology in the Schools, 14 (1), 10-15.

Kaufman, A.S. (1982). The Impact of WISC-R Research for School Psychologists. In C.R. Reynolds & T.B. Gutkin (eds.), The Handbook of School Psychology. New York: Wiley.

Knoff, H. & Sperling, B. (1986). Gifted Children and Visual-Motor Development: A Comparison of Bender-Gestalt and VMI Test Performance. Psychology in the Schools, 23 (3), 247-2[illegible].

Matarazzo, J. (1976). Wechsler's Measurement and Appraisal of Adult Intelligence, 5th Edition. Baltimore: The Williams & Wilkins Co.

McKenzie, P. (1983). Three Intelligence Measures: A Comparative Anal; sis. Unpublished Master's thesis, University of Alberta.

Moore, G.W. (1983). Developing and Evaluating Educational Research. Boston: Little, Brown & Company.

Naglieri, J. (1985). Use of the WISC-R and K-ABC with Learning Disabled, Borderline Mentally Retarded and Normal Children. Psychology in the Schools, 22 (2), 133-141.

Neufeld, J. & Cozac, E. (1980). A Study of the Self-Concept of Intellectually Superior Children. Alberta Journal of Educational Research, 26 (3), 149-157.

Norusis, M.J. (1984). SPSS/PC: For the IBM PC/XT. Chicago: SPSS Inc.

Osberg, D. (1986). New Tests: New Options. Communique, 15 (3), 3.

Paananen, N. (1983). Incidence and Characteristics of Depression in Late Childhood: An Exploratory Study. Unpublished Master's thesis, University of Alberta.

Sattler, J. (1982). Assessment of Children's Intelligence and Special Abilities, 2nd Edition. Boston: Allyn & Bacon, Inc.

Siewert J. & Breen, M. (1983). The Revised Test of VMI: Its Relation to the Test of VMI and Bender Visual Motor Gestalt Test for Regular Education Students. Psychology in the Schools, 20 (3), 304-306.

Sim, J. (1987). The Stanford-Binet IV: Recent Criticisms and Issues. The Alberta School Psychologist, 7 (1), 17-21.

Simon W.E. & Simon M.G. (1975). Self-esteem, Intelligence and Standardized Test Achievement. Psychology in the Schools, 12 (1), 97-100.

Slate, J. (1986). New Does Not Necessarily Mean Better. Communique, 15 (1), 3.

Smith, C.R. (1983). Learning Disabilities: The Interaction of Learner, Task and Setting. Boston: Little, Brown & Co.

Smith, T. & Smith B. (1986). The Relationship Between the WISC-R and WRAT-R for a Sample of Rural Referred Children. Psychology in the Schools, 23 (3), 252-254.

Telrow, C. (1987). New Stanford-Binet and Preschoolers. The Alberta School Psychologist, 7 (1), 22.

Terman, L & Merrill, M. (1973). Stanford-Binet Intelligence Scale, 1973 Norms Edition. Boston: Houghton-Mifflin Co.

Thompson, J. (1984). An Empirical Versus Conceptual Model of the Stanford-Binet for Interpretive Purposes. Unpublished Master's thesis, University of Alberta.

Thorndike, R. (1975). Mr. Binet's Test Seventy Years Later. Educational Researcher, 4 (5), 3-6.

Thorndike, R., Hagen, E. & Sattler, J. (1986A). The Stanford-Binet Intelligence Scale: Fourth Edition: Technical Manual. Chicago: The Riverside Publishing Co.

Thorndike, R., Hagen, E. & Sattler J. (1986B). Guide for Administering and Scoring the Fourth Edition. Chicago: Riverside Publishing Co.

Waddell, D. (1980). The Stanford-Binet: An Evaluation of the Technical Data Available Since the 1972 Restandardization. Journal of School Psychology, 18 (3), 203-209.

Walsh, W. & Betz, N. (1985). Tests and Assessment. Englewood Cliffs: Prentice-Hall, Inc.

Wright, D. & Dappen, L. (1982). Factor Analysis of the WISC-R and the WRAT with a Referral Population. Journal of School Psychology, 20 (4), 306-312.

APPENDIX A

DISTRIBUTION OF THE SAMPLE BY DEMOGRAPHIC CHARACTERISTICS

Tables 1 - X1

TABLE 1
DISTRIBUTION OF THE SAMPLE BY AGE

AGE	N	PERCENT	AGE	N	PERCENT
3-0 to 3-11	3	2.1	10-0 to 10-11	14	9.5
4-0 to 4-11	11	7.4	11-0 to 11-11	9	6.1
5-0 to 5-11	13	8.7	12-0 to 12-11	13	8.8
6-0 to 6-11	12	8.1	13-0 to 13-11	7	4.7
8-0 to 8-11	17	11.2	15-0 to 15-11	6	4.2
9-0 to 9-11	11	7.4	16-0 to 16-11	9	6.0
			18-0 to 27-11	3	2.1
TOTAL N = 153			X AGE = 10.1		

TABLE 11
DISTRIBUTION OF THE SAMPLE BY GRADE

GRADE	NUMBER	PERCENT
Special Class	7	4.6
Pre-School	15	9.8
Kindergarten	16	10.5
Grade One	16	10.5
Grade Two	9	5.9
Grade Three	15	9.8
Grade Four	9	5.9
Grade Five	14	9.2
Grade Six	12	7.8
Grade Seven	10	6.5
Grade Eight	13	8.5
Grade Nine	6	3.9
Grade Ten	3	2.0
Grade Eleven	7	4.6
Grade Twelve	1	.7
TOTAL	153	100.0

TABLE 111
DISTRIBUTION OF THE SAMPLE BY GENDER

SEX	NUMBER	PERCENT
Male	91	59.5
Female	62	40.5
TOTAL	153	100.0

TABLE IV
DISTRIBUTION OF THE SAMPLE BY ETHNICITY

ETHNICITY	NUMBER	PERCENT
White	136	88.9
Native	5	3.3
Oriental	7	4.6
Other	5	3.3
TOTAL	153	100.0

TABLE V
DISTRIBUTION OF THE SAMPLE BY LOCALE

LOCALE	NUMBER	PERCENT
Urban	99	64.7
Rural (including small towns)	54	35.3
TOTAL	153	100.0

TABLE V1
DISTRIBUTION OF THE SAMPLE BY BIRTH ORDER IN THE FAMILY

ORDER IN FAMILY	NUMBER	PERCENT
First	67	43.8
Second	57	37.3
Third	16	10.5
Fourth	5	3.3
Fifth	4	2.6
Sixth	1	.7
Seventh	2	1.3
Eighth	1	.7
TOTAL	153	100.0

TABLE V11
DISTRIBUTION OF THE SAMPLE BY PARENT MARITAL STATUS

MARITAL STATUS	NUMBER	PERCENT
Married	128	83.7
Divorced	7	4.6
Separated	7	4.6
Single	1	.7
One Parent Deceased	1	.7
Parent Remarried	4	2.6
Unknown	5	3.3
TOTAL	153	100.0

TABLE V111
DISTRIBUTION OF THE SAMPLE BY FATHER'S OCCUPATION

OCCUPATION	NUMBER	PERCENT
Managerial/Professional	42	27.5
Technical/Sales	20	13.1
Service Occupations	17	11.1
Farming/Forestry	12	7.8
Precision Production	11	7.2
Operators/Fabricators	20	13.1
Unemployed	2	1.3
Unknown	29	19.0
TOTAL	153	100.0

TABLE 1X
DISTRIBUTION OF THE SAMPLE BY MOTHER'S OCCUPATION

OCCUPATION	NUMBER	PERCENT
Managerial/Professional	5	3.3
Technical/Sales	6	3.9
Service Occupations	7	4.6
Farming/Forestry	0	0.0
Precision/Production	0	0.0
Operators/Fabricators	7	4.6
Unemployed	1	.7
Unknown	127	83.0
TOTAL	153	100.0

TABLE X
DISTRIBUTION OF THE SAMPLE BY HANDEDNESS

HAND USED	NUMBER	PERCENT
Right	131	85.6
Left	17	11.1
Alternated	1	.7
Unknown	4	2.6
TOTAL	153	100.0

TABLE X1
DISTRIBUTION OF THE SAMPLE BY GRADE REPETITION

GRADE REPEATED	NUMBER	PERCENT
Yes	32	20.9
No	108	70.6
Unknown	13	8.5
TOTAL	153	100.0

APPENDIX B

PSYCHOLOGICAL TESTING CENTER

FACULTY OF EDUCATION, CLINICAL SERVICES

REPORT FORWARDING FORM

Table X11

TABLE X11
REPORT FORWARDING FORM*

REASON FOR REFERRAL: (Check one or several)

_____Readiness for Kindergarten (4-5)
_____Pre-School Assessment (2-4)
_____Learning Disabilities (Child)
_____Learning Disabilities (Adult)
_____Dyslexia (Child or Adult)
_____School Learning Problems (General)

_____Behavior Problems

_____Developmental Delay/Maturation
_____Social/Emotional Problems (General)
_____Self-Referral - General Interest

_____Career Choices/Planning
_____Personality/Ability Interest

_____Reading/Language Problems
_____Neuropsychological
_____Child Custody
_____Guaranteed Income (Adult)
_____Giftedness/Enrichment
_____Parental Interest (In child's ability)
_____Special Class or Program Placement
_____Special School Placement
_____Mental Retardation
_____Sensory Impairment (Hearing vision, speech)

* *Reason for Referral*, used for this study is only part of the complete "Report Forwarding Form" used in the Faculty of Education, Clinical Services.

APPENDIX C

TOTAL SAMPLE SUMMARY STATISTICS

Table X111

TABLE X111
TOTAL SAMPLE SUMMARY STATISTICS

VAR	MEAN	STD DEV	KURT	SKEW	MIN	MAX	S.E. $\bar{X}$
V18	96.661	19.188	-0.358	0.160	49.000	143.000	1.766
V20	98.746	19.533	-0.104	-0.040	54.000	152.000	1.798
V22	94.490	20.470	0.657	-0.019	26.000	155.000	1.884
V24	10.299	9.214	72.476	7.927	1.000	19.000	0.936
V26	44.306	13.407	-1.180	-0.173	18.000	66.000	1.580
V27	6.806	3.249	-1.115	-0.181	1.000	12.000	0.383
V28	8.833	2.773	-0.027	-0.772	1.000	12.000	0.327
V29	7.972	2.793	-0.765	-0.363	1.000	12.000	0.331
V30	8.471	3.335	-0.665	-0.658	1.000	12.000	0.399
V31	7.889	3.093	-1.055	-0.340	2.000	12.000	0.364
V32	4.750	2.555	-0.897	0.292	1.000	10.000	0.301
V38	51.366	9.247	0.775	-0.171	23.000	79.000	0.748
V39	51.740	9.470	0.461	-0.390	23.000	78.000	0.773
V40	52.718	8.395	-0.228	-0.105	32.000	70.000	0.734
V42	51.566	8.934	0.907	0.133	25.000	82.000	0.725
V43	48.530	8.330	-0.295	0.012	29.000	71.000	0.777
V44	52.635	8.580	0.360	-0.457	30.000	72.000	0.931
V46	50.960	10.089	-0.152	0.015	21.000	75.000	0.824
V47	53.686	8.387	0.394	-0.377	31.000	72.000	1.002
V49	50.536	9.390	0.357	0.126	23.000	77.000	0.764
V50	49.209	9.171	0.140	-0.232	18.000	70.000	0.754
V51	52.209	8.104	-0.172	0.158	31.000	71.000	0.850
V52	52.012	7.609	0.357	-0.122	33.000	70.000	0.820
V53	104.320	19.134	0.965	-0.432	41.000	163.000	1.547
V54	102.209	17.369	0.964	-0.116	49.000	161.000	1.404
V55	103.225	20.357	-0.022	-0.082	42.000	151.000	1.657
V56	101.822	18.390	0.975	-0.207	46.000	155.000	1.492
V57	103.484	19.165	0.945	-0.479	38.000	157.000	1.549

APPENDIX D

ONEWAY ANALYSIS OF VARIANCE

BINET IV, WRAT-R, VMI and PASS SCORES

TABLES X1V - XL1

TABLE X1V
ONEWAY ANALYSIS OF VARIANCE
BINET IV COMPOSITE IQ

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F. Prob.
Between Groups	3	32718.63	10906.21	71.38	0.000
Within Groups	140	21388.36	152.77		
Total	143	54107.00			

TABLE XV
ONEWAY ANALYSIS OF VARIANCE
BINET IV VERBAL REASONING IQ

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	25675.36	8558.45	42.30	0.000
Within Groups	140	28321.96	202.29		
Total	143	53997.32			

TABLE XV1
ONEWAY ANALYSIS OF VARIANCE
BINET IV ABSTRACT/VISUAL REASONING IQ

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F. Prob.
Between Groups	3	20182.55	6727.51	40.81	0.000
Within Groups	140	23075.88	164.82		
Total	143	43258.43			

TABLE XV11
ONEWAY ANALYSIS OF VARIANCE
BINET IV QUANTITATIVE REASONING IQ

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F prob.
Between Groups	3	29758.16	9919.38	46.82	0.000
Within Groups	138	29232.63	211.83		
Total	141	58990.79			

TABLE XV111
ONEWAY ANALYSIS OF VARIANCE
BINET IV SHORT-TERM MEMORY IQ

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F prob.
Between Groups	3	23203.26	7734.42	40.43	0.000
Within Groups	139	26586.80	191.27		
Total	142	49790.06			

TABLE X1X
ONEWAY ANALYSIS OF VARIANCE
BINET IV VOCABULARY

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F prob.
Between Groups	3	5901.11	1967.03	40.97	0.000
Within Groups	140	6720.04	48.00		
Total	143	12621.15			

TABLE XX
ONEWAY ANALYSIS OF VARIANCE
BINET IV COMPREHENSION

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F prob.
Between Groups	3	4238.29	1412.76	23.38	0.000
Within Groups	137	8276.91	60.41		
Total	140	12515.20			

TABLE XX1
ONEWAY ANALYSIS OF VARIANCE
BINET IV ABSURDITIES

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F prob.
Between Groups	3	2440.95	813.65	14.67	0.000
Within Groups	118	6543.34	55.45		
Total	121	8984.29			

TABLE XX11
ONEWAY ANALYSIS OF VARIANCE
BINET IV PATTERN ANALYSIS

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F prob.
Between Groups	3	3245.75	1085.58	18.70	0.000
Within Groups	139	8067.09	58.03		
Total	142	11323.84			

TABLE XX111
ONEWAY ANALYSIS OF VARIANCE
BINET IV COPYING

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F prob.
Between Groups	3	1836.14	612.04	11.31	.0000
Within Groups	105	5681.89	54.11		
Total	108	7518.03			

TABLE XX1V
ONEWAY ANAYSIS OF VARIANCE
BINET IV MATRICES

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F prob.
Between Groups	3	2158.79	719.59	4.55	.0000
Within Groups	80	3954.09	49.42		
Total	83	6112.89			

TABLE XXV
ONEWAY ANAYSIS OF VARIANCE
BINET IV QUANTITATIVE

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F prob.
Between Groups	3	6531.63	2177.21	38.37	.0000
Within Groups	137	7772.50	56.73		
Total	140	14304.14			

TABLE XXV1
ONEWAY ANALYSIS OF VARIANCE
BINET IV NUMBER SERIES

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	2642.80	880.93	26.30	.0000
Within Groups	66	2210.28	33.48		
Total	69	4853.08			

TABLE XXV11
ONEWAY ANALYSIS OF VARIANCE
BINET IV BEAD MEMORY

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	4259.43	1419.81	23.56	0.0
Within Groups	138	8315.20	60.25		
Total	141	12574.73			

TABLE XXV111
ONEWAY ANALYSIS OF VARIANCE
BINET IV MEMORY FOR SENTENCES

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	4708.50	1569.50	28.72	.0000
Within Groups	135	7375.24	54.63		
Total	138	12083.75			

TABLE XX1X
ONEWAY ANALYSIS OF VARIANCE
BINET IV MEMORY FOR DIGITS

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	1895.63	631.87	13.74	.0000
Within Groups	86	3954.01	45.97		
Total	89	5849.65			

TABLE XXX
ONEWAY ANALYSIS OF VARIANCE
BINET IV MEMORY FOR OBJECTS

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	757.55	252.51	4.9	.0032
Within Groups	82	4163.43	50.77		
Total	85	4920.98			

TABLE XXX1
ONEWAY ANALYSIS OF VARIANCE
WRAT-R ARITHMETIC

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	24619.38	8206.46	51.29	.0000
Within Groups	113	18079.85	159.99		
Total	116	42699.24			

TABLE XXX11
ONEWAY ANALYSIS OF VARIANCE
WRAT-R READING

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	26930.14	8976.71	57.29	.0000
Within Groups	113	17705.15	156.68		
Total	116	44635.29			

TABLE XXX111
ONEWAY ANALYSIS OF VARIANCE
WRAT-R SPELLING

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	19908.42	6636.14	25.77	.0000
Within Groups	113	29096.56	257.49		
Total	116	49004.99			

TABLE XXXIV
ONEWAY ANALYSIS OF VARIANCE
VMI

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	367.46	122.48	15.38	.0000
Within Groups	86	684.63	7.96		
Total	89	1052.10			

TABLE XXXV
ONEWAY ANALYSIS OF VARIANCE
PASS FULL SCALE

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	1513.65	504.55	3.05	.0344
Within Groups	68	11247.62	165.40		
Total	71	12761.27			

TABLE XXXV1
ONEWAY ANALYSIS OF VARIANCE
PASS GENERAL ABILITY

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	109.02	36.34	3.85	.0130
Within Groups	68	640.25	9.41		
Total	71	749.27			

TABLE XXXV11
ONEWAY ANALYSIS OF VARIANCE
PASS ARITHMETIC

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	21.35	7.11	.92	.4348
Within Groups	68	524.64	7.71		
Total	71	546.00			

TABLE XXXV111
ONEWAY ANALYSIS OF VARIANCE
PASS SCHOOL SATISFACTION

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	22.64	7.54	.96	.4137
Within Groups	68	523.29	7.81		
Total	71	545.94			

TABLE XXX1X
ONEWAY ANALYSIS OF VARIANCE
PASS READING/SPELLING

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	105.75	35.25	3.51	.0198
Within Groups	68	661.68	10.02		
Total	71	767.44			

TABLE XL
ONEWAY ANALYSIS OF VARIANCE
PASS PENMANSHIP/SPELLING

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	65.45	21.81	2.41	.0738
Within Groups	68	613.65	9.02		
Total	71	679.11			

TABLE XL1
ONEWAY ANALYSIS OF VARIANCE
PASS CONFIDENCE

SOURCE	df	SUM OF SQUARES	MEAN SQUARES	F RATIO	F Prob.
Between Groups	3	21.14	7.05	1.08	.3622
Within Groups	68	442.36	6.51		
Total	71	463.50			