229

0-315-26960-X

National Library of Canada — Bibliothèque nationale du Canada

Canadian Theses Division — Division des thèses canadiennes

Ottawa, Canada
K1A 0N4

PERMISSION TO MICROFILM — AUTORISATION DE MICROFILMER

• Please print or type — Écrire en lettres moulées ou dactylographier

Full Name of Author — Nom complet de l'auteur
Jan Elizabeth Thompson

Date of Birth — Date de naissance
Oct 6, 1956

Country of Birth — Lieu de naissance
Canada

Permanent Address — Résidence fixe
876 Charlesworth Drive
Sarnia, Ontario
N7V 2S2

Title of Thesis — Titre de la thèse
An Empirical Versus Conceptual Model of the Stanford-Binet For Interpretive Purposes

University — Université
University of Alberta

Degree for which thesis was presented — Grade pour lequel cette thèse fut présentée
Master's of Education

Year this degree conferred — Année d'obtention de ce grade
Spring 1984

Name of Supervisor — Nom du directeur de thèse
Dr. H. L. Janzen

Date
April 19/84

Signature
Jan E. Thompson

NL-91 (4/77)

THE UNIVERSITY OF ALBERTA

AN EMPIRICAL VERSUS CONCEPTUAL MODEL OF THE STANFORD-BINET FOR INTERPRETIVE PURPOSES

by

Jan Elizabeth Thompson

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF EDUCATION

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

SPRING 1984

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR JAN ELIZABETH THOMPSON

TITLE OF THESIS AN EMPIRICAL VERSUS CONCEPTUAL MODEL OF THE STANFORD-BINET FOR INTERPRETIVE PURPOSES.

DEGREE FOR WHICH THESIS WAS PRESENTED MASTER OF EDUCATION

YEAR THIS DEGREE GRANTED SPRING 1984

(Signed) Jan E. Thompson

PERMANENT ADDRESS:
876 Charlesworth Drive
Sarnia, Ontario
N7V 2S2

DATED April 10 1984

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 AN EMPIRICAL VERSUS CONCEPTUAL MODEL OF THE STANFORD-BINET FOR INTERPRETIVE PURPOSES, submitted by JAN ELIZABETH THOMPSON in partial fulfillment of the requirements for the degree of Master of Education.

....
Supervisor

V R Nyberg..............

Grace Malicky..............

Date April 10, 1984

ABSTRACT

The usual first step in any thorough diagnostic assessment for learning problems is the administration of a standardized intelligence scale such as the Stanford-Binet or one of the Wechsler Intelligence Scales. In addition to obtaining a general index of current intellectual functioning (i.e. an IQ score), clinicians usually look to these tests to aid them in developing hypotheses about specific intellectual strengths and weaknesses. It is virtually standard procedure for clinicians to engage in some form of profile analyses in the interpretation of intelligence tests. The Wechsler Scales have an empirically well defined factor structure. The diagnostic potential of the factors has been well validated and profile analysis is considered to be a legitimate practice in the interpretation. Such, however, is not the case for the Stanford-Binet. The Binet is an age scale and has not been designed to be factorially pure. Evidence exists, however, that the Binet is factorially complex and that some diagnostically useful factors can be identified at various ages. To date, however, very little effort has been directed towards clarifying the factor structure of the current revision of the scale. In the absence of a clear empirical factor structure, a number of conceptual models have been developed to aid the clinician in profiling individual strengths and weaknesses in Binet protocols. The validity of these conceptual models, however, has not been validated either in terms of the empirical factor structure of the test or in terms of their reliable use by clinicians. The Stanford-Binet has a number of advantages over the Wechsler Scales in some

instances and a better assessment of its diagnostic potential would be helpful. To make a preliminary step in this direction was the aim of the thesis.

Part A of the study involved a factor analytic investigation of the Stanford-Binet across mental ages VI through XI employing data obtained from test protocols of 220 children between the ages of three and twelve seen at the Education Clinic in the Faculty of Education at the University of Alberta. Separate analyses were conducted across test years VI through VIII and across test years VIII through XI. The analyses at years VI through VIII revealed five interpretable group factors in addition to a general factor. These were identified as Verbal, Nonverbal Reasoning; Verbal; Visualization and Visual Judgement; Control of Impulsivity; and Visual-Motor Integration factors. The analysis at years VIII through XI also revealed five interpretable group factors. These were Verbal; Verbal, Nonverbal Reasoning; Visualization; Difficulty; and Meaningful, Nonmeaningful Memory factors.

An attempt was also made to statistically conform the empirical factors to the conceptual factors outlined by Sattler (1960) in his conceptual system for profile analysis of the Stanford-Binet. This attempt, for the most part, was unsuccessful. This was due in part to the fact that the conceptual factor structure could not be adequately defined given the limited number of variables included in each analysis, and in part to the fact that the majority of the test items were revealed by the factor analyses to be multidimensional (i.e. the factor loadings were multivocal). In Sattler's system items are assigned uniquely to categories. A more informal comparison, however, which took into account the multidimensional nature of the variables revealed there to be some consistency between the kinds of dimensions defined by Sattler and revealed by the factor analytic investigations.

Part B of the study assessed the degree to which 42 graduate student clinicians agreed with Sattler in the assignment of Stanford-Binet items to the conceptual categories. Each clinician was given the category names and definitions and asked to place each Binet item into the most appropriate category. Three choices were allowed but these were to be made on a priority basis. Results indicated that if only the first choice selections were taken into account, then the proportion of clinicians agreeing with Sattler was quite variable from item to item. When all the choices were taken into account, the agreement proportions were much improved. Comparisons of students' ratings with the empirical factor structure also revealed some significant degree of consensus across test years VI through XI. In some cases where there was disagreement with Sattler, the students' views were supported by results of the factor analyses. In cases where there was agreement with Sattler, further support was often present in the form of the empirical factors.

Implications of these findings for the use of profile analysis in the interpretation of the Stanford-Binet are discussed.

ACKNOWLEDGEMENTS

It is a great pleasure to thank those people who made this thesis possible.

Dr. Henry Janzen, my thesis supervisor, was a source of continuous support and encouragement. Not only has Dr. Janzen taught me much about the assessment of children's abilities but also much about teaching through his patience, his constant positive feedback, and his model of hard work.

I would like to thank Dr. Grace Malicky and Dr. Vern Nyberg, committee members, for their patience and encouragement and for the promptness and care with which they read and commented on the manuscript. I would also like to thank Dr. Nyberg for all he taught me about measurement.

I must express my sincere appreciation to John Reddon whose helpful comments and advice greatly assisted in the preparation of the manuscript.

These acknowledgements would not be complete without an expression of my gratitude to all the faculty and staff of the Department of Educational Psychology for their help and encouragement throughout my entire graduate program, a period marked by considerable personal as well as academic stress. Special thanks go to Dr. Harvey Zingle, Dr. Robert Short, Mrs. Emma Collins, Mrs. Aggie Ganchev, and Dr. John Paterson, all of whom were particularly supportive.

Finally, this thesis truly would not have been completed without the support of my family and of John and his family. These were the people who would not let me quit.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF TABLES

CHAPTER ONE

1. OVERVIEW

1.1 Introduction

More time and effort has probably been expended in attempting to measure intelligence than on any other project in psychological measurement. For centuries men have puzzled over and are still puzzling over the enormous differences in intellectual capacity which exist among individuals. As societies developed and people's roles within them became more specialized, it was increasingly urgent to find some way to evaluate individual abilities as accurately as possible. It was under the pressure of these practical needs that psychologists began in earnest their effort to measure intelligence. During the 1900's Alfred Binet developed an examination for the Parisian school system to identify who could and who could not do the school work. The development of this test is usually cited as the major cornerstone in the history of intellectual assessment (Thorndike, 1975). There now exist a multitude of different tests and procedures used in the measurement of mental abilities - most of which are, in some respects at least, decendents of the original test devised by Binet (Thorndike, 1975). Among these is included the Stanford-Binet Form L-M which is the result of three revisions of Binet's scale undertaken in the United States in 1916, 1937 and 1960. This test, which is an age scale, is very much in use today.

1.2 The Problem

The Stanford-Binet Form L-M is one of the oldest, best validated, and most widely used individual tests of intelligence. It is rivaled only by the Wechsler Intelligence Scales which have borrowed extensively from the theoretical framework upon which the Stanford-Binet was developed. Although the Stanford-Binet was designed to provide a single index of overall intellectual potential - that is, an IQ score - attempts have been made to examine more specifically the item content in hopes of better identifying the various specific abilities being measured. This has come about not only through the more theoretical research on the 'nature of intelligence' but also out of real clinical need. Assessment of intellectual ability has served as a means of predicting performance. When a child's rate of learning, however, is less than what would be predicted, an IQ score neither explains nor accounts for this failure (Meeker, 1969). Hence, the need for much more detailed diagnostic information which would ideally allow the clinician to profile individual strengths and deficits and thus provide a more direct link between assessment and remediation. There are many diagnostic tests available which presume to measure one or more of the various abilities thought to be vital to learning and such tests are invaluable. Given, however, that the administration of an individualized intelligence scale is usually routine in any thorough assessment, that the Stanford-Binet has a history of thorough validation, that many items on the Binet are similar in content to items on other diagnostic tests and finally, that the age range for which the Binet is suitable is wide, it would be helpful if it could be validly used to obtain more specific information which could aid the clinician in directing the diagnostic process.

Since the first revision of the original scale in 1916 by Lewis Terman at Stanford University there have been a number of attempts to examine the internal structure of the Stanford-Binet using factor analytic procedures. Burt and John (1942a) conducted an extensive series of investigations on the original Binet scale. They found evidence for the influence of other abilities in addition to that of 'general intelligence' which the test was intended to measure. In a later factor analytic investigation of the 1916 Stanford revision they were able to isolate, in addition to a general factor, seven distinct group factors (Burt and John, 1942a,b). These findings led Burt and John (1942b) to conclude that more specific abilities play a larger part in individual test performance than might otherwise be inferred.

Wright (1939) also attempted to isolate primary abilities in a part of the original Stanford-Binet. She was able to isolate a common factor and six interpretable primary factors.

One of the most well known factor analyses was conducted by Quinn McNemar using the standardization data collected on the 1937 Stanford revision (McNemar, 1942). McNemar had fully expected that he would find the test to be highly saturated with one common factor to the exclusion of obvious group factors, this being consistent with the design of the test to provide a single index of intellectual ability which is comparable across individuals. He found, however, evidence for the influence of other specific abilities in addition to 'general intelligence'. His results provided every indication that the 1937 Forms L and M of the Stanford-Binet were more factorally complex than was originally predicted.

Jones (1949) also analyzed standardization scores on Forms L and M of the 1937 revision of the Stanford-Binet. Four separate analyses were conducted to include several mental age levels. In contrast to the findings of McNemar (1942) no evidence for a general factor was found. According to Jones, the IQ variance could be explained by psychologically meaningful group factors at the various mental age levels. Six factors were isolated at years XII through XIV. Three factors were interpreted across years X through XII. Four factors were found at years VIII through X. Analyses at years VII and VIII isolated four meaningful factors.

In a methodological refinement of his earlier work Jones (1954), repeated the factor analysis at years XII through XIV employing the same correlational data as used in the previous study. A total of nine factors were ultimately interpreted.

There have been relatively few attempts to date to examine the factor structure of the 1960 revision of the Stanford-Binet. Ramsey and Vane (1970) found seven factors as the result of an analysis of mental ages IV-6 through VI of the 1960 revision. No factor in this analysis was interpreted as a general intelligence factor.

Hallahan, Ball and Payne (1973) examined the factorial composition of the Short Form of the 1960 revision of the Stanford-Binet across mental ages III-6 through V. These researchers were able to isolate only three factors - a general factor and two group factors.

Stormer (1967), in an attempt to identify the components of the Stanford-Binet and the aspects of intelligence it does or does not measure, examined the 1960 revision of the scale in conjunction with a battery of other tests employing an external factor analytic procedure. His sample included over four hundred fifteen year old subjects. Stormer concluded that while memory and spatial aptitudes are minimally measured, the major portion of the variance in the upper range of the Binet is attributable to the verbal factors of fluency, reasoning and production.

In summary it can be seen that there is considerable evidence to suggest that the Stanford Binet is more factorially complex than one might predict. Even Terman and Merrill (1960), authors of the 1960 revision, in the process of arguing that "the very factors that contribute to the Stanford-Binet's success as a measure of general intellirence must interfere with its usefulness as a measure of the various separate aspects of mentality . . ." (p. 12), are forced to admit that while "a single common factor can explain the performance on the Stanford-Binet, this does not exclude the possibility that group factors might be present at some age levels" (p. 34). Despite, however, this strong evidence that the Stanford-Binet is indeed factorally complex, a quick examination of the results of the various investigations seems to yield some lack of agreement. While there has been general consistency in the types of factors found, the exact number and composition has varied from one investigation to another. Detailed comparison of results has been hampered by the fact that different revisions of the scale have been used, that researchers have analyzed different portions of the test, that subject groups have differed in size, age and IQ ranges, and that statistical methodology has varied from study to study. Hence, while one is left with a strong impression that the 1960 revision of the Stanford-Binet might have considerable potential as a diagnostic instrument, a clear and consistent definition of its factorial composition has not yet been achieved.

According to Ramsey and Vane (1970) "almost all school and clinical psychologists utilize some form of subtest analysis when evaluating individual intelligence test results". What then, are the current users of the Stanford-Binet doing? If the Wechsler Intelligence scales were used, say Ramsey and Vane (1970), the clinician could rely upon the factors outlined in factor analytic studies reported by Wechsler and others. The scarcity of factor analytic studies of the 1960 revision and the relative inconclusiveness of analyses of earlier revisions, prohibit similar possibilities with the Stanford-Binet at this time. The psychologist, thus, has been forced to rely on logical analysis (Ramsey and Vane, 1970).

Several attempts have been made to develop a classification scheme for the Stanford-Binet based on categories which will allow placement of individual test items into meaningful groups on the basis of content or face validity. Sattler (1965), for example, has developed a system which classifies items into seven major ability categories and five subcategories. Using this system the clinician can quickly profile an individual's test performance and fairly readily identify particular strengths and weaknesses. Valett (1964) has devised a similar schema which assigns items to six categories. Meeker (1969) has also developed a classification system for the Stanford-Binet based upon Guilford's Structure of Intellect Model. Guilford (1967) has proposed a three-dimensional model of intelligence and has employed factorial methods to test it. In this model 120 different hypothesized intellectual abilities are defined. Meeker's procedure for classifying items according to the Structure of Intellect model identifies 55 of Guilford's abilities in the Stanford-Binet.

At this point, it is appropriate to ask some questions about the validity of these classification systems. How much agreement exists between the different schemes in their assignment of items to ability categories? Given the categories, do trained clinicians agree on the placement of items within them? What kind of correspondence exists between these conceptual systems and the empirical factors? Finally, is subtest analysis even a legitimate activity for Stanford-Binet users? To date very little effort has been directed towards answering the first three questions and without those answers it is not possible to respond accurately to the final question. In order to provide definitive answers considerably more effort must be directed towards clarifying the factor structure of the 1960 revision of the Stanford-Binet across all age levels. A much more indepth examination of the various conceptual classification systems must also be undertaken which assesses both their correspondence to the empirical factor structure of the test and their use as a reliable clinical tool. To make a preliminary step in this direction is the aim of this thesis.

1.3 Statement of Objectives

The first objective is to provide a brief history of the Stanford-Binet and in so doing place its development within an appropriate theoretical framework.

The second objective is to undertake a detailed review of some previous factor analytic studies of the various revisions of the Stanford-Binet. Emphasis will be placed on a comparison of both the methodology and the resultant factor solutions.

The third objective is to conduct a factor analytic study of the 1960 revision of the Stanford-Binet over the mental age levels VI through VIII and VIII through XI using several differing factoring methods and employing a clinical population of children ranging in age from three to twelve.

The fourth objective is to assess the reliability of student clinicians' judgements in assigning Stanford-Binet items to Sattler's classification scheme.

The fifth objective is to combine the results of the present factor analyses and clinician judgement study with past research in an attempt to provide a better assessment of our current status with respect to evaluating the validity of using subtest analyses with the Stanford-Binet.

The final objective is to provide direction for further research.

1.4 The General Design

The predominant component of the research consisted of a factor analytic study of the 1960 revision of the Stanford-Binet across the mental age range VI through XI. Test data from two groups of children, one group ranging in chronological age from three to six years and the other from seven to twelve years were included in the analyses. Separate analyses were conducted with data from the younger group across the age range VI to VIII of the test and with data from the older group across the age range VIII to XI. Both exploratory and confirmatory methods of analysis were employed. In the confirmatory analyses an attempt was made to conform the empirical factors to Sattler's conceptual ones.

In another phase of the research student clinicians in Counselling and School Psychology programs at the University of Alberta were asked to judge each of the Stanford-Binet items with respect to the conceptual categories outlined in Sattler's classification scheme.

1.5 Plan of the Thesis

In Chapter II the history of the Stanford-Binet is outlined and its development is placed within the context of the psychometric approach to the definition and measurement of intelligence. Previous factor analytic studies of the Stanford-Binet are reviewed. The conceptual subtest classification systems of Sattler and Valett are examined.

Part A of Chapter III provides a description of the two sample groups employed in the Factor Analytic Study. Subject selection procedures are explained. Sampling adequacy is evaluated. Part B of the chapter is concerned with the sample employed in the Clinician Judgement Study. Subject selection procedures are described and sampling adequacy is evaluated.

In Parts A and B of Chapter IV the methods of analysis employed in the Factor Analytic and Clinician Judgement Studies respectively are outlined.

In Part A of Chapter V the results of the Factor Analytic Study are presented. Comparisons of factor solutions obtained within groups across the several factoring methods and across groups within the various methods of analysis will be emphasized. Part B of the chapter is devoted to the results of the Clinician Judgement Study. Comparison of the clinicians' classification of test items to both Sattler's schema and to the empirical factor solutions are made.

Finally, in Chapter VI the results of the thesis are summarized, and implications of these results and directions for further research are presented.

CHAPTER TWO

2. REVIEW OF LITERATURE

2.1 Introduction

In order to evaluate fully the extent to which further study of the internal structure of the Stanford-Binet may be useful, it is first necessary to examine the development of the test. An understanding of the Binet's place within the theoretical framework surrounding the 'nature of intelligence' is also important. Hence, before a review of earlier factor analytic studies of the Stanford-Binet is undertaken, some relevent historical background is provided. Following a comparison of results of previous factor investigations attention will be directed to the examination of the alternate classification systems of Sattler and Valett. Sattler's model, being the one of primary interest in this study, will be described in some detail. Valett's system, because it is similar to Sattler's and because it is also extensively used by clinicians, will be outlined briefly for comparison purposes. Since the use of Meeker's classification system, based on Guilford's Structure of Intellect Model, is currently limited primarily to research applications, it will not be discussed.

2.2 Historical Development of the Stanford-Binet Form L-M

Although Alfred Binet is usually given the credit for developing the first intelligence test, it was Francis Galton who was actually the creator of the attribute we now call 'intelligence' in as much as "factors which make a man eminent had not previously been assembled under one heading" (Liungman, 1975, p. 13). Much of the concept of 'intelligence' was thus pioneered by Galton in the late 1800's. He claimed that there were enormous biologically inherited differences in the quality of understanding between different individuals. After showing that various degrees of mental capacity, from genius through mediocrity to idiocy, did exist, Galton set out to demonstrate that they were biological entities just as height, eye colour and so forth. He opened a laboratory in which he set out to measure the physical features of his subjects hoping to prove that there indeed would be a relationship between such attributes as head size and intelligence. In the tradition of Galton researchers continued to systematically study sensory functions and simple psychological processes in search of individual differences.

Alfred Binet, however, was critical of what he felt to be an overly simplistic approach. He felt that the tests of that time were too heavily weighted in the direction of sensory functioning and simple psychological processes. He also argued that they failed to contain a sufficiently varied sample of measures related to diverse mental faculties:

> We are approaching here a new, difficult, and very little explored subject . . . If one looks at the series of experiments that have been made - the mental tests - one is astonished by the considerable place reserved to the sensations and simple processes, and by the little attention lent to superior processes, which some [experimenters] neglect completely.
>
> The objection will be made that the elementary processes can be determined with much more precision than the superior ones; this is true, but people differ much less in these elementary processes than in the complex ones; there is no need, therefore, for as precise a method for determining the later as for the former, a

> point which is often forgotten. Anyway, it is only by applying ourselves to this point that we can approach the study of individual differences.
>
> Let us recall once more that the objective sought is not to determine all the differences among the psychic faculties of individuals, but to determine the strongest and most important ones . . . This is a rule that has not been considered and followed by anyone. We must expend our attention on superior psychic faculties . . . (quoted in Wolf, 1973, p. 145).

Binet's own search, however, for an understanding of these 'superior psychic faculties' was less than straightforward. Throughout his career the direction of his research varied considerably. Areas of endeavor ranged from the measurement of tactile thresholds, to the intensive and systematic investigation of the thinking and behavior of his two daughters, to the study of mental illness and mental retardation. Binet did not conceive of 'intelligence' as having a relatively independent existence in personality since the weight of his writing stressed the unity of functioning in each individual (Sarason, 1976). Binet was interested in human behavior in all its forms. This drive to permeate the human mind, to understand it as a complete whole caused Binet to be over ambitious in his goals and less directed in his research efforts than he might have been. Despite this, Wolf's (1973) review of Binet's research and writing prior to 1905 indicates that he had begun to germinate ideas that were much closer to modern conceptions of intelligence than were those of many theorists who followed. Binet's statement in 1890 that "it would not appear absolutely impossible that sometime in the near future one could succeed in measuring intelligence, that is, reasoning, judgement, memory, the ability to make abstractions" is remarkable only in its similarity to current thinking on the nature of intelligence (quoted in Wolf, 1973, p. 143). Binet's developmental studies of his two daughters were, for their time, masterpieces of a marvelously insightful mind. He fastidiously recorded differences in the girls' attention span and concentration. He studied their language usage. He compared their memories and their ability to form visual images. He monitored their development of number-sense. He tested their abilities to make judgements about length and quantity. He compared their behavior in social situations. Although he could not adequately define the differences, Binet felt that the younger child differed distinctively from the older child not only in amount of knowledge but also in manner of thinking and reasoning. The astute reader will immediately be impressed by how much these ideas and methods of research were similar to those of Piaget, one of the most eminent of recent investigators in the study of children's intellectual development. In fact, "Piaget has long been aware and appreciative of Binet as a subtle analyst of thought processes" (Wolf, 1973, p. 331). According to Sarason (1976) this is far higher praise than Piaget accorded the intelligence test developers who came after Binet.

Binet's intensive study of his daughters provided impetus for the continued search for 'superior faculties'. He conducted hundreds of studies in which he attempted to examine such things as memory, attention, comprehension, word fluency, mental imagery, coordination skills and quick visual judgements, and task persistence. "His subjects included infants, children, adults - normal, retarded, mentally ill, criminal and representative of all classes of society" (Wolf, 1973, p. 333). Throughout this period Binet struggled gallantly to discover how these various 'faculties' were synthesized to form the global entity 'intelligence'. He had difficulty in coming to terms with the nature of these 'faculties'. "He argued that a memory or a volition, for example, is imaginery and does not exist. The reality lay in acts of memory, acts of volition, or little particular and distinct events" (Wolf, 1973, p. 144). Yet Binet persisted in speaking of faculties of memory, attention, and the like as though they were inherent entities or traits of the individual. Yet again, he also argued that "intelligence was not a constant, or a fixed amount. Intelligence was educable" (Sarason, 1976, p. 581). While contending that there were limits to the extent to which intelligence was susceptible to development, Binet felt that "with practice and training, and especially with appropriate methods of teaching we can augment a child's attention, his memory, his judgement - helping him to literally become more intelligent than he was before . . . right up to the moment when he arrives at his limit" (quoted in Wolf, 1973, p. 207).

Binet was not only a developmental psychologist and an experimental psychologist but also very much a clinician. He became very involved in trying to find ways to help and change human behavior. His advocation of 'mental orthopedics' which was aimed at teaching children to listen, remember, and judge more effectively is only one example of Binet's clinical involvement. He also directed a considerable amount of effort toward the study of the thinking and behavior of mentally retarded children and adults and in fighting for their humane treatment and appropriate education. "Among all those who have participated in the development of intelligence tests, it is most unlikely that anyone rivals Binet for his first hand, investigative experiences with diverse populations" (Sarason, 1976, p. 583).

Despite Binet's extensive research and clinical experience, his amibition to devise mental tests which would allow for appropriate evaluations of individual differences was continually thwarted by his own confusion over the exact nature, number and organization of the 'superior processes'. "It appears [in 1903] that Binet had the ingredients to make a measure of intelligence, but how was he to put them together in a framework of the many 'faculties' that, although representing acts rather than entities were still acts that, in the minds of Binet and other psychologists required separate and distinct testing?" (Wolf, 1973, p. 160).

By 1905 Binet was under pressure by the French Ministry of Public Education to develop a simple, convenient and precise test which would readily identify those who were unable to cope with the regular school program and hence were in need of special education. Around this time Binet integrated a number of earlier insights which allowed him to form a somewhat different approach to the problem and hence to escape from his preoccupation with the measurement of individual faculties. Binet had already concluded that only complex functions yielded significant individual differences and that single tests of any function were useless - that they must always include a number of measures of each. He had also come to realize that in order to make meaningful comparisons among individuals, norms must be established by measures of 'normal' children at different age levels. Given that there were socially-accepted age norms at which children should have completed grade levels in school, this could be used as a method for determining 'normal' achievement. What remained was for Binet to take the step from attempting to measure each separate faculty to testing responses to questions or 'tasks', whatever their psychological components might be. Binet did take this step in 1905 and he and his associate Theodore Simon began an extensive search for intellectual 'tasks' that would lie in a hierarchy of difficulty according to the ages at which a significant majority of children performed correctly. They tried out their items (many of which were identical to or modifications of tasks Binet had used throughout his previous fifteen years of research) in institutions, hospitals, schools and creches. After much testing the first Binet scale was finally developed consisting of thirty items arranged roughly in order of difficulty to test children from three to twelve or older. "Data on the performance of 'normal' children carrying out tasks that cut across these faculties proved to be the breakthrough in Binet finally being able to develop a useable scale" (Wolf, 1973, p. 176).

What exactly, then, did Binet feel he was measuring with his test? According to Wolf (1973, p. 179), Binet made it clear that the test did not analyze special aptitudes although he mentioned their probable intrusion into items. Rather, he felt that it was generally measuring a fundamental agent which he called 'judgement'. It would immediately appear that this represented a rather dramatic shift in thinking on the part of Binet from the view that intelligence was defined by a number of independent abilities, to the development of a test in which the implication is made that intelligence is one general thing. Tuddenham (1962), however, provides a more consistent interpretation: "Regarding intelligence as a product of many abilities, Binet sought in his tests to measure not an entity or single dimension - 'general intelligence' - but rather an average level - 'intelligence in general' (p. 489). Thorndike (1975) also argues that "it was not that Binet failed to recognize that functions were involved other than judgement and that each might be individually evaluated. It was rather that he felt it most important to provide a unitary overall appraisal of level of mental functioning, and believed that a judicious pooling of a variety of different tasks provided the best basis for such an appraisal . . ." (p. 4). This would appear to be a very important distinction.

In 1908 the 1905 scale was revised by Binet and Simon. This revision expanded considerably the scope of the test. It could be used to distinguish among the intellectual abilities of normal children as well as to distinguish the mentally defective child from the normal child. This required, of course, modification and ~~addition~~ of items. The concept of mental age was also clarified. Tests were grouped according to age level and the mental age of the child was determined at the highest level where he/she could pass all the items but one plus credit of one year for every five items passed at a more advanced level (Brody and Brody, 1976, p. 4).

In 1911 Binet, alone, once again revised his scale to include a number of technical changes. The age levels of some items were altered. Five items were placed at each age level. Fifteen year and adult levels were added. Binet established the basal year at the age where all tests were passed and two-tenths of a year was granted for each test passed above this basal. Although procedures for determining the mental level were made explicit, no provisi was made for the calculation of an intelligence quotient (IQ) which would permit the comparison of ability of individuals at different ages (Brody and Brody, 1976, p. 5).

Although Wilhelm Stern developed the concept of IQ in 1911 Binet did not feel that his test warranted the kind of precision that the calculation of intelligent quotients would imply. As noted by Wolf (1973, p. 201), Binet was not convinced that the differences among all levels of ability were strictly quantitative. He was concerned that there might also be some qualitative differences. This would be consistent with his view of the scale as an average of 'intelligence in general'. If the scale was to be used to distinguish among levels of ability, the similarities among them, except in degrees, should be maximal to permit the same instrument to be used to compare the total group.

Many revisions of the original Binet-Simon test were eventually made. One of the first and the most important was the one undertaken in the United States by Lewis Terman at Stanford University in 1916 (Terman, 1916). This is the revision from which the current form of the Stanford-Binet is directly decended. Terman expanded the test by adding some new items. He also rearranged the age level of some of the already existing ones. The new items were very similar to those developed by Binet. The test was standardized on a sample of approximately one thousand children between the ages of five and fourteen and four hundred adults. Effort was made to obtain a representative sample of the American population. No foreign born children were included in the sample. Terman had no doubts about the precision of his test. He indicated a person's relative position to his own age group by computing the ratio of mental age to chronological age. The IQ, then, as it was originally conceptualized was simply this ratio multiplied by one hundred.

In 1937 the Stanford-Binet underwent a second significant revision (Terman and Merrill, 1937). The single 1916 scale was replaced by two parallel forms L and M. It was intended that the new scales should test as nearly as possible the same aspects of intelligence as did the earlier revision. The aim of the 1937 revision was to expand the test to include items at both the upper and lower ends, to improve upon the standardization sample and to refine instructions for administration and scoring. Prior to the standardization hundreds of items were developed and pre-tested. Every attempt was made to include items which were similar in form to and correlated highly with those on the earlier test. Only those demonstrating rapid rise in percents passing at successive age levels were tentatively selected. Form L and M were standardized by administration of both forms to approximately three thousand Caucasion children so as to sample the main geographical areas of the United States. Approximately one hundred subjects were included at each half year interval from one and a half to five and a half years, two hundred at each age from six to fourteen, and one hundred at each age from fifteen to eighteen. Each group was evenly divided between the sexes and all subjects were within one month of their birthdays (or half-year birthdays for subjects below age six). Several modifications were then made until the final two scales were equivalent with respect to difficulty range, reliability, and validity. Each scale contained 129 items. Each was so standardized for difficulty as to yield mean IQ's which approximated as closely as possible to one hundred, that is, mean mental age of unselected subjects corresponded with mean chronological age. The entire revision was carried out over a period of ten years (McNemar, 1942, p. 1-7).

In 1960 a third and final revision of the Stanford-Binet was undertaken (Terman and Merrill, 1960). The most significant change was to incorporate back into a single scale the best items from the L and M forms of the 1937 scale. No attempt was made to restandardize the scales but the existing standards were evaluated against further data which had been collected during the early 1950's. The evaluation was based primarily upon the test performance of approximately forty five hundred subjects between the ages of two and a half and eighteen years. The subjects were not chosen to constitute a representative sampling of American school children nor was there proportional distribution of subjects across age or mental age groups. Any changes in difficulty of subtests was determined by comparing percents passing the individual tests in the 1950's with percents passing in the original standardization group. Criteria for selection of items remained the same. Only those items which demonstrated an increase in percent passing with age and which had high correlations with total test score were retained. Items which were found to have changed significantly in difficulty since the original standardization were either eliminated or relocated. A few items were dropped because cultural changes had rendered them inadequate. Provision was also made to have an alternate item at each age level.

Several changes were also introduced which were directed at correcting several structural problems apparent in the 1937 scale. Firstly, the mean IQ's that the scale yielded at some levels turned out to be somewhat above one hundred. In the 1960 revision adjustments were made to make the average mental age that the scale gives more nearly equal to the average chronological age at each age level.

A shift was also made from the calculation of IQ's by the ratio method to the calculation of deviation IQ's. The revised IQ's became standard scores with a mean of one hundred and a standard deviation of sixteen. These values approximated the values obtained on the total 1937 standardization sample. Standard score IQ's eliminated the problem of atypical variability which occurred at some age levels on the 1937 revision and allowed for better interage comparisons.

The 1960 revision of the Stanford-Binet, then, is the scale which is currently in use. It is an age scale which taps abilities ranging from the two year level to superior adult. There are a total of 138 items, seven at each half year from year II to year IV-6. The seventh item in each case is an alternate and is not included in administration unless one of the first six is not correctly administered. There are also seven items (the seventh again being an alternate) at each year level from year V to year XIV. The Average Adult subtest contains eight items and one alternate. Each of the three Superior Adult subtests contains six items and an alternate. One month of mental age credit is given for each correct item from year II to year V. Two months credit is given for each item between years VI and Average Adult. Four, five, and six months credit respectively are given for each item in the three Superior Adult categories. The basal age is designated as that level at which all tests are passed which just precedes the level where the first failure occurs. In computing the mental age the testee is credited with his/her basal age. The appropriate number of months credit are then added for each item passed above the basal until a ceiling is reached. The ceiling is that level at which all tests are failed. Mental ages are then readily converted to deviation IQ's by reference to the appropriate tables.

In 1972 a restandardization of the 1960 revision was undertaken (Terman and Merrill, 1973). The purpose was to update the norms and to develop new IQ tables to reflect the performance in the 1970's of subjects over the entire range of the test. At all age ranges the tests and the directions for adminstration and scoring remained the same. The 1972 norming sample consisted of approximately twenty one hundred subjects, one hundred at each Stanford-Binet age level. The subjects had all been part of, or were siblings of children who had been part of, a large-scale norming of the Cognitive Abilities Test undertaken across the United States in 1970. The communities sampled for the standardization of the Cognitive Abilities were stratified according to size, geographic region, and indices of economic status. In order to generate an appropriate norming sample for the Stanford-Binet subsamples, stratified according to mental age, of the Cognitive Abilities Test norming sample were identified and used. To obtain cases at the younger ages it was necessary to rely upon siblings of the original Cognitive Abilities norming group. Stratification was based upon the Verbal scale of the Cognitive Abilities Test because it was felt that scores on this scale would have the highest correlation with the Stanford-Binet IQ's. Younger children were categorized upon the basis of the performance of their siblings. Based upon the results of this standardization 1972 Revised IQ tables were published (Thorndike, 1973). Generally, the traditional relationship between chronological age, mental age and IQ does not hold for the 1972 norms. Using the same chronological age and mental age, the IQ obtained with the 1972 norms is usually slightly lower than that obtained with the 1960 norms. The largest changes are at the preschool level. For instance, a four year old child must obtain a mental age of four years and six months to obtain an IQ of one hundred (Waddell, 1980). Holbroyd et al. (1976) and Davis et al. (1975), however, feel these changes are reflective of true changes in the rate of mental development of children and hence the 1972 norms represents a definite improvement over the old norms in this respect.

To this point, the development of the Stanford-Binet has been traced from its earliest conception to its current form. "Binet's last revision was dated 1911; it is astonishing to see how negligible have been the changes in substance or scope since that time" (Wolf, 1973, p. 331). Throughout its seventy-five year history, the Binet has served as "the workhorse of the psychometrical appraisal of cognitive development, the standard against which other tests of cognitive abilities have been evaluated" (Thorndike, 1975, p. 3). In order to fully appreciate the historical roots of the Binet, however, it is important to examine the theoretical context in which it was developed. It has already been demonstrated that the tasks which Binet originally proposed "stemmed not from any elegant theory nor from any precise definition of intelligence" but were a result of painstaking observation of 'intelligence in action'. His selection of items for his test was based upon empirical tryouts with groups of children of various ages (Thorndike, 1975, p. 3). Although Binet did feel that intelligence was mediated by the interaction of a number of specific 'faculties' he abandoned any attempt to separate out individual abilities in his test. While urgent practical needs interfered with Binet developing a precise theory of intelligence, there were many others who were not so diverted.

The creation of tests which were seemingly so successful in distinguishing individual differences in intelligence provided the very framework from which the theorists could then work. It was felt that detailed analysis of these tests would reveal the 'nature of intelligence'. Advancements in mathematical and statistical techniques, specifically the development of the concept of correlation and the resulting techniques of factor analysis, both stimulated and were stimulated by the study of individual differences. These techniques "were seen as affording scientists the means for discovering the essence of intelligence" (Maloney and Ward, 1976). The critical issue which emerged was whether intelligence was comprised of a single general factor or whether there were a number of separate factors. Given separate factors, what was their nature and how were they interrelated? Spearman (1923, 1927) was the first to present a theory of intelligence based upon statistical analysis of test scores. Spearman was convinced that all intellectual activities shared a common factor which he termed 'g'. In studying the correlations between a wide variety of tests of complex mental functioning, he was impressed by the extent of positive relationships between them. This lead him to believe that the tests were all measuring the same underlying process. Since the correlation between different tests was never perfect, Spearman concluded that each test possessed a component which was strictly unique, to itself. This he termed the specific or 's'. This 's', being unique, did not contribute to the correlation between tests. Spearman advanced the method of factor analysis as a way of determining the contribution of particular tests to 'g'. Spearman's theory came to be known as the two-factor theory of intelligence.

While it was true that all tests of intelligence correlated positively, it became obvious that the intercorrelation between items tended to fall into identifiable patterns. This finding lead to the conclusion that besides the common process specific to all tests, there were also processes common to specific types of tests. That is, in addition to the common factor there also existed a number of group factors. This became known as the group-factor model. The relative importance of these group factors versus the general factor was the subject of some controversey. Spearman, himself, was forced to acknowledge the possible presence of some group factors but was unwilling to relegate them anywhere near the importance of the common factor (Spearman, 1927).

Thurstone (1935, 1947), on the other hand, in a radical departure from Spearman's two-factor approach, formulated the multiple-factor theory of intelligence. He postulated the existence of seven broad multiple factors in place of any general factor. These seven factors became known as Thurstone's Primary Mental Abilities and included: Number, Word Fluency, Verbal Meaning, Reasoning, Spatial Awareness and Perceptual Speed.

Forming a compromise between these positions were the hierarchical theories of Burt (1941) and Vernon (1950). According to these theorists, abilities were organized in a hierarchical fashion with a 'g' factor subsuming all other factors. 'G' could then be broken down into several broad group factors. These could then be subdivided into progressively smaller individual factors. Vernon (1950), for example, posited a general factor, two major group factors (verbal-educational versus practical-mechanical), several minor group factors and multiple specific factors.

Attempts have also been made to place the hierarchical model within a developmental framework. Garrett (1946) hypothesized that in young children the general factor would be the most important factor in accounting for variance in abilities. With time and experience, though, specific abilities would become differentiable. Adult intelligence, then, would be characterized by multiple specific components at various levels.

Guilford (1967) provided yet another schema for the organization of intellectual abilities. Guilford's conceptualization involved a three-way classification system designed to encompass and organize intellectual aptitude factors. The ideas basic to the theory were formulated following the factor analysis of many tests. These ideas were subsequently refined and the Structure of Intellect Model (SOI) was developed. Every intellectual ability in the structure is characterized in terms of the operation which is employed, the content involved and the product which results. 'Operations' are defined as "major kinds of intellectual activities or processes; things the organism does with the raw materials of information, information being defined as that which the organism discriminates" (Meeker, 1969, p. 13). There are five subclasses of Operations: Memory, Cognition, Evaluation, Divergent Production and Convergent Production. 'Contents' are "broad classes or types of information discriminable by the organism" (Meeker, 1969, p. 22). Contents are differentiated by four subcategories: Figural, Symbolic, Semantic and Behavioral. 'Products' define "the organization that information takes in the organism's processing of it" (Meeker, 1969, p. 23). Products can be differentiated in six ways: Units, Classes, Relations, Systems, Transformations and Implications. Hence, the complete schema yields a total of 120 cells in each of which one factor or ability is expected. Guilford has devoted considerable effort in attempting to specify and verify the 120 proposed abilities. He feels that he has adequately identified one hundred of the cells and specific tests have been developed which supposedly measure these factors.

It can be seen from this overly brief overview that the early applications of the psychometric approach in an attempt to reveal the nature and organization of intellectual abilities produced a noteable lack of concensus. Anywhere from one to one hundred and twenty distinct abilities organized both hierarchically and morphologically have been proposed.

It is well known, however, that the apparent disagreements which emerged over the factorial composition of intelligence were due in some considerable degree to differences in statistical methodology. The term factor analysis does not describe a single invariant statistical method but rather is a generic name for a variety of mathematical procedures for defining groupings of variables and for determining which variables belong to which group (Nunnally, 1978, p. 327). A number of mathematically correct solutions can be obtained from the same set of data employing differing factoring methods. Given this number of mathematical possibilities, the decision as to which approach will supply the more scientifically useful information is generally based primarily upon the particular theoretical orientation of the investigator. Hence, there is nothing in the factor analytic methods themselves that can demonstrate that one factor solution is technically more correct than another. Once this issue is clarified, then, it no longer seems as surprising that the different investigators arrived at differing conceptions about the nature and organization of intellectual abilities. "Traits which are identified through factor analysis are simply an expression of correlation among behavioral measures. They are not underlying entities or causal factors, but descriptive categories. Hence, it is conceivable that different principles of classification may be applicable to the same data" (Anastasi, 1983). Spearman and his followers, for example, employed the method of tetrad differences or centroid analysis without rotation of centroid axes. These methods invariably favoured finding a single general factor. Statistical and mathematical advancements in factor analysis allowed Thurstone to employ the rotational techniques which would effectively abolish the general factor producing only group factors thus giving results consistent with his theoretical orientation. The hierarchical model of Burt and Vernon was made possible by further developments in methodology which allowed for the uncovering of higher-order factors (Adcock, 1954, p. 36).

A further complication in comparing factor solutions is the influence of age and selectivity of samples used in the study. To a certain extent the factor solution is affected by the heterogeneity of the group studied. When the sample is young or fairly heterogeneous with respect to age, education, ability and so forth, evidence for a 'g' factor is usually found. When older or homogeneous samples are employed 'g' tends to disappear and to be replaced with numerous specific factors (Maloney and Ward, 1976, p. 190).

Given the awareness of the interrelationship between theory and methodology, what, then, are the current notions about the organization of intellectual abilities? Generally, with probablay few exceptions, it could be said that the current prevailing opinion lies somewhere between that of Spearman and Guilford. From a psychometric view intelligence is now for the most part seen as being multidimensional, that is, to be made up of a number of distinct but related abilities. The exact nature and number is still a matter of some debate but are generally felt to include a number of specific abilities related to the understanding and use of language (vocabulary, fluency, reasoning and comprehension), as well as a number of abilities which are primarily nonverbal in nature. These include such skills as spatial perception, nonverbal reasoning and eye-hand coordination. Memory skills are also usually proposed. With the acceptance of a multidimensional view, however, the focus of investigative efforts has tended to shift towards defining patterns of ability rather than level of ability (Mahoney and Ward, 1976, p. 193).

In recent years there has also been increasing emphasis placed on the study of processes that underlie observed responses. For example, to be able to answer that a baseball and an orange are different because one is edible and one is not (a question on the Stanford-Binet) would seem to require some abstraction function. Many investigators are becoming increasingly interested in studying not only the content of responses but also the methods employed to produce them. Both cognitive and developmental psychology have been influencial in producing this change of focus. Developmental psychology has taken the approach that there are not only quantitative changes in intelligence with age but there are also qualitative changes in processes which occur with development. Cognitive psychology with its emphasis on problem-solving and information-processing has likewise been concerned with the study of underlying intellectual processes. It will be recalled that Binet originally was concerned with examining not only what children knew but the way they thought. As stated earlier, his intensive analysis of his daughters led him to conclude that the younger child differed distinctively from the older child in her manner of thinking, reasoning and remembering. In a critique of a colleague's attempt at developing a test of intellectual ability Binet stated: "Since all the themes were of equal value, and each was given a single composite score, the results presented no analysis of the kind of successes and failures; for example, there were no comparisons of the relative difficulties of response to abstract versus concrete materials" (quoted in Wolf, 1973, p. 175).

Binet, however, abandoned his attempt to define specific processes in favour of finding test items which would discriminate well between the varying overall ability levels regardless of the underlying processes involved in the tests themselves. This led to the focus on the quantitative nature of intelligence in testing. For years the issue was simply "how much?". Given, then, our current conception of the nature of intelligence being made up of a number of distinct abilities and our concern with the processes which may underly these abilities where does this leave us with respect to the assessment of intelligence? Are tests like the Stanford-Binet which provide us with a single quantitative index of overall intellectual functioning obsolete? The answer to this question at this time must be "no". Firstly, there currently are few, if any adequately normed tests of overall intellectual functioning which define patterns of abilities or tap specific underlying cognitive processes. The Wechsler scales have made a small step towards separating tasks requiring nonverbal kinds of abilities from those requiring skills more related to verbal ability but, in general, are similar to the Stanford-Binet in content and were designed primarily to provide a general overall index of current intellectual functioning. There is no doubt that as our understanding of cognitive processes increases that the nature of assessment instruments will change. This, however, still appears to be some distance in the future. In the meantime, however, "it seems reasonable to expect that the Stanford-Binet will continue to be used extensively for some time" (Davis an Rowland, 1975).

Despite the lack of well defined theory guiding its development, the Stanford-Binet has proven itself generally to be a valid measure of overall scholastic ability (Davis and Rowland, 1975; Brooks, 1977, Bossard and Galusha, 1979). While, as mentioned previously, a general index of ability is often not sufficient in allowing the clinician to reach a diagnosis in cases of academic underachievement, obtaining such an index is usually a vital first step in the clinical process. With the exception of the McCarthy Scales of Children's Abilities, no scale other than the Stanford-Binet exists for children between the ages of twenty-four and forty-eight months. The Wechsler Preschool and Primary Scale of Intelligence is not suitable for children younger than four. The McCarthy Scales, developed in 1972 and suitable for children from two and one half to eight and one half have not been subjected to anywhere near the rigorous tests of validity that has the Stanford-Binet. According to Gerken et al. (1978), if the purpose of testing is to determine a child's general level of intellectual functioning, one must still proceed with caution in using the McCarthy

Scales. Furthermore, Gerken argues that there is no evidence to support the use of the McCarthy Scales for specific educational planning; therefore it does not have an advantage over the Stanford-Binet in that respect. Recent emphasis on early childhood education and early intervention with young children at risk for learning difficulties has increased the need for assessment instruments suitable for pre-schoolers (Davis and Rowland, 1975). Young children also seem to enjoy the variety and frequent shifting of tasks which an age scale like the Binet allows. Fagan et al. (1969) found that in terms of ease of administration, clinicians preferred the Stanford-Binet to the Wechsler Preschool and Primary Scale.

The inclusion of tasks at the younger age levels also makes the Binet useful in assessing the mentally retarded of any age. Also, its extension into the superior adult range makes the test ideal for assessing gifted children since it allows for adequate ceiling (Evans and Richmond, 1976). The more limited age ranges for which the Wechsler preschool and children's scales are suitable has been one of their greatest weaknesses.

In summary, it would appear that "the venerable Stanford-Binet retains its position as the most broadly applicable measure of intellectual functioning" (Levinson and Thomas, 1979, p. 291).

The question now arises as to whether the clinician can obtain more than just an IQ score from the Stanford-Binet. Can particular patterns of ability or cognitive processes be identified in Binet profiles? As described earlier, most clinicians are already using some sort of profile analysis with the Binet, usually employing one of the conceptual classification schemes proposed by Sattler or Valett. As yet these systems have not been adequately compared to factor analytic studies. Results of previous factor analytic studies, however, indicate that while obviously not factorially pure, the Stanford-Binet is indeed factorially complex. Given the history of its development this is not surprising. Binet, himself, admitted the probable intrusion of specific abilities into items. What remains to be seen, however, is whether the factor structure is sufficiently clear at the various age levels that profile analysis can be undertaken. Given Binet's view of the clinical process it is doubtful that he would object to efforts in this direction. Binet, himself, felt that there were qualitative as well as quantitative differences in the performance of individuals on the test. Binet stressed an intensive study of the individual. He saw the clinical process as much more involved than is currently thought. Binet never planned for his test to be used in the way it often has been, that is, apart from a thorough clinical assessment of the individual. Burt and John (1942a) also provide evidence that subtest analysis was, in fact, common practice with the original Binet scales: "The effects of these various factors, irrelevant as they may seem in the assessment of intelligence, are nevertheless exceedingly suggestive in clinical examinations. To the experienced tester they yield instructive side-lights on the special intellectual or temperamental disabilities of the individual child; and in the record-card drawn up for the original Binet scale we endeavoured, so far as possible, to arrange similar tests in similar columns, so that special abilities or disabilities might be recorded straight away by a glance at the child's record" (p. 118).

Now that the development of the Stanford-Binet has been placed within the appropriate historical and theoretical context, it is now possible to examine in depth previous attempts to examine its factor structure.

2.3 Previous Factor Analytic Studies of the Stanford-Binet

2.3.1 The 1916 Revision

Burt and John (1942a) conducted an extensive series of investigations over a period of ten years using data obtained from British school children administered the original Binet-Simon scale. They found strong statistical evidence for the influence of other abilities in addition to that of 'general intelligence' which the test was intended to measure. These specific abilities were thought to be verbal, numerical, manual, relational (involving reasoning) and educational (involving special knowledge) in nature.

In order to evaluate the effects of the American revision on the nature of special abilities being tapped by the test, these investigators undertook a factor analytic study of the tests alloted to ages X and XII of the 1916 Stanford revision. No items were contained at age XI in this revision (Burt and John, 1942a,b). The sample consisted of 483 children ranging in age from ten to fourteen and a half. The group was chosen to be as homogeneous as possible with respect to mental age. IQ's ranged from 80 to 100. Seven factors were extracted from the matrix of tetrachoric correlations between items by the centroid method. Factor extraction by the centroid method results in factors which usually have loadings on every test involved. Furthermore, the first factor may have all positive factor loadings but the other factors have both positive and negative loadings in about equal proportions, that is, they are bipolar. The first factor, with positive saturations throughout, was interpreted by Burt and John to be a general factor, although they do say that since eight of the twelve tests employed in the analysis were verbal, it was likely that the factor possessed a strong verbal bias.

The second factor which was bipolar in nature was interpreted to be an age level or maturity factor since results yielded, with the exception of one item at each level, negative loadings at age ten and positive loadings at age twelve. The authors argue that this effect of maturity or acquired knowledge was outstanding in this analysis in part because with the absence of tests at age XI, there was a fairly wide gap between tests selected as appropriate for ten year olds and those selected for children two years older.

The second bipolar factor was interpreted to be verbal, numerical and spatial in nature. Vocabulary, reading, abstract words, sentences and verbal absurdities items had positive saturations and are obviously verbal in nature. Repeating numbers, copying drawings items and the ball and field tests had negative saturations. The copying drawings and ball and field tests were further separable from repeating digits items on the basis of their loadings on the second bipolar factor and hence were interpreted as representing visio-spatial and arithmetical factors respectively.

The third bipolar factor was felt to encompass immediate memory, vocabulary and verbal comprehension abilities. Items involving repeating digits and copying drawings from memory were grouped together and were presumed to represent a memory factor. Three vocabulary items were also grouped and were interpreted to reflect an ability to understand isolated words. The grouping of reading, mixed sentences, comprehension, fables, verbal absurdities and ball and field items were assumed to reflect an ability to comprehend situations.

The remaining three factors did not account for a statisically significant amount of the test variance and hence were not interpreted.

Burt and John (1942b) also reinterpreted their factors after rotation of the original centroid axes to simple structure. This procedure serves to maximize the number of zero or nearly zero factor loadings and to minimize the number of negative loadings thus producing a number of positive group factors in place of the bipolar factors. A small number of high positive loadings and few negative loadings on group factors simplifies the interpretation of the factor solution. A general factor and seven group factors were interpreted. The nature of the factors were much the same as before. The common factor was interpreted as a general intelligence factor. The group factors included an age or maturity factor, a verbal factor, a vocabulary factor, a comprehension factor, a number factor, a spatial factor and a memory factor.

Burt and John (1942b) put forth the hypothesis that "the major group factors are not single and self-contained abilities, but unfold into a number of factors which are still more specialized. Thus 'verbal ability' proves to be highly complex - ability for understanding isolated words, for example, being relatively independent of facility in dealing with verbal patterns and both of comprehension. Similarly, the 'spatial factor' may be either visual or kinaesthetic; while 'comprehension' may involve either predominantly verbal or predominantly visual imagery" (p. 161).

Based upon the results of their analyses, Burt and John conclude that the presence of additional factors is not without advantages. If duly informed as to the relative weight of the different factors influencing the tests, an experienced tester might use the scale, not only to estimate a child's general intelligence, but also to gain insight into specific abilities and disabilities.

Wright (1939) also attempted to isolate primary abilities in a part of the original Stanford-Binet. Her sample consisted of 456 children between the age of ten years and ten years eleven months and ranging in mental age from six to fourteen. Only those items passed by more than ten percent and less than ninety percent of the sample were retained. Thus a total of thirty-one items from years VII to XIV were included in the analysis. Tetrachoric coefficients were used to obtain the intercorrelations for the variables. The correlation matrix was factored by the centroid method. Seven factors were extracted. The centroid axes were graphically rotated until an approximation to simple structure was achieved. A common factor and six group factors were interpreted. Only items having factor loadings of .30 or above were used in the psychological interpretation of factors.

Wright offered two explanatory hypotheses for the common factor. The first hypothesis being that the factor was of the nature of 'g' or 'general intelligence'. The second hypothesis was formulated on the assumption that mental functions, at first undifferentiated, develop from general to specific, and that there are individual differences in rate of mental development. That is to say, that the factor could best be thought of as one of maturation. Wright admitted that her data did not allow for any discrimination between those two hypotheses since if in a population of children at the same chronological age, those children of a more accelerated rate of mental growth have all their abilities more advanced than children of a slower rate of growth, there would be a correlation existing between the primary abilities that would not be present in the adult population, thus explaining the loading of all items on a common factor. Wright also acknowledged the possible influence of the test design on the common factor since the items were selected by a method which stressed their intercorrelations. Based, however, upon previous analyses of other test batteries carried out by Thurstone with adult populations which did not find evidence of a common factor, Wright concluded that the maturational hypothesis was the most tenable.

In order of their contribution to the total variance, the first five group factors were labelled: Verbal Relations, Numerical, Spatial, Reasoning, and Induction. The sixth factor could not be adequately defined.

The numerical factor included tasks such as counting backwards, repeating digits, making change, and calculating dates. The spatial factor involved tasks such as drawing designs from memory, reversing hands of a clock, paper cutting and copying a diamond. The verbal factor included tasks requiring the definition of words, the interpretation of fables, the recognition of verbal absurdities, the completion of sentences and comprehension. The reasoning factor included ball and field problem questions and interpretation of picture items. The fifth factor was labelled only tentatively as Induction since the nature of the items which loaded on it were quite variable but all seemed to be related to finding a rule or principle.

2.3.2 The 1937 Revision (Forms L and M)

McNemar (1942) conducted the most extensive factor analytic study of the Stanford-Binet to date employing the 1937 standaradization data. McNemar undertook fourteen separate analyses organized so that each item of the scale (Form L and M) would be included in at least one analysis. The number of items included in each analysis varied from nineteen to thirty-five. Tests which were identical in both forms were only included once and tests which were repeated with different passing standards within the particular age range being examined were only included once as well. Two hundred subjects at each chronological age from two to seven, with the exception of age five where there were one hundred, were included in the sample. There were also two hundred subjects each at ages nine, eleven and thirteen, and one hundred each at ages fifteen and eighteen. McNemar did not provide data on the IQ ranges within each of his analysis groups but it is assumed that the distribution approximated that of the original standardization groups. Tetrachoric correlation coefficients were obtained between the variables for each analysis. Three centroid factors were extracted in each case.

McNemar found that for a given analysis all the items were saturated with a common factor as expected. He noted, however, that at some age levels some items were low in general factor saturation. This included such items as block building, formboard, motor coordination, copying a circle, paper folding, copying a bead chain, picture absurdities, block counting, memory for stories, repeating digits and paper cutting. The low loadings on the general factor hinted to McNemar of the possibility of the existence of one or more group factors.

Items in the analysis of mental years II and II-6 plotted with reference to the second and third centroid axes provided some evidence of a separation between items involving identifying objects, items involving a motor component and items involving predominantly memory skills.

The analyses involving items from years II, II-6 and III resulted in a similar pattern of groupings. McNemar notes that the results are unlikely to be due to chance or sampling since the variance of the first factor residuals indicated that one factor was not sufficient to explain the intercorrelations of items at these age levels.

Items in the analyses involving mental years III through V, according to McNemar, did not fall into any logical groups of clusters.

Hence, no meaning was attached to the second or third centroid factors.

With the five and six year old groups across mental ages V through VII there was evidence of a separation between items which were primarily verbal in nature and those requiring the application of number concepts.

At year levels VIII through X there appeared to be distinct verbal and memory factors.

Analyses across years XII through XIV provided evidence for separate groupings of 'memory', 'verbal' and 'problem' types of tests.

The results based on age group eighteen for the analysis of the items at the three Superior Adult levels indicated that more than one factor was needed to explain the intercorrelations between the items. Groupings of memory items were quite evident but otherwise the pattern formed by the remaining items was, according to McNemar, not easily interpretable.

Although he argued that the general factor was the predominant factor in the test at all ages, McNemar admitted that the existence of some group factors at some age levels was probable but that the limited samples employed were insufficient for definitely establishing their importance or nature. He stated, however, that "they appear to be of sufficient prominence to cause small, though not inconsequential, qualitative differences between the IQ's of two individual's when the mental maturity of either or both is at any one of the levels where these factors emerge" (p. 117).

Jones (1949) applied orthogonal rotational procedures to the centroids found by McNemar at chronological age levels seven, nine, eleven and thirteen. Further centroid factors were also extracted and included in the rotations. Graphical rotation of centroid axes was guided by the desire to both approximate simple structure and maximize the possible psychological meaning of the results. In the interpretation of factors, items with loadings of .40 or greater were considered significant.

Across mental ages XII through XIV six meaningful factors were found. The greatest contribution to the total variance was made by a factor which was labelled Verbal. Thirteen of the thirty items included in the analysis loaded significantly on this factor. Some items defining the factor were vocabulary, abstract words, dissected sentences, verbal absurdities, picture absurdities and memory for stories.

The second factor was classified as a reasoning factor and included such items as plan of search and ingenuity. The presence of verbal items such as vocabulary, abstract words and memory for sentences could not be explained.

The third factor was also called a reasoning factor and was identified by ingenuity, problems of fact and memory for sentences items. No hypotheses as to the difference between these factors was offered apart from the comment that since reasoning is a complex process it would not be surprising to find more than one reasoning factor in a single analysis.

The fourth factor appeared to be a memory factor and included such tasks as repeating digits, memory for words and copying a bead chain from memory. The presence of two items which did not appear to have a memory component (paper cutting and minkus completion) could not be explained. Absent from the memory factor were memory for designs, memory for stories and memory for sentences items. It was argued that this indicated the probable existence of several distinct memory factors. Memory for form or memory for coherent vebal material did not necessarily require the same ability as memory for digits or unrelated words.

The fifth factor, defined by only two distinct items, was tentatively labelled a visualization factor since the common variable relating the items appeared to be the ability to manipulate a set of visual imagery.

Items defining the sixth factor all appeared to require the recognition of spatial relationships, it thus being called a spatial factor.

At mental ages X through XII the verbal factor again predominated. In addition memory and spatial factors were defined. Two other factors did not contribute significantly to the total test variance and were not interpreted.

At years VIII and X verbal, spatial, reasoning, and memory factors were identified. A residual factor was not interpreted.

At years VI and VII verbal, reasoning, and memory factors were clearly defined. A fourth factor accounting for 8.8 percent of the variance was difficult to interpret but was tentatively classified as a number factor. It included number of fingers, counting taps, repeating digits and days of the week items. The presence of comprehension and sentence building items on this factor, however, could not be explained.

Noteable in the results of this study is the absence of a common factor. This seemed particularly surprising given that using the same data, McNemar found the common factor to be the most significant in accounting for the test variance. Jones argues that "items on the Stanford-Binet do not measure a unitary factor, that the relations among items can be explained most efficiently and consistently in terms of group factors, each contributing to a subject's total score or IQ" (Jones, 1949, p. 317). He also notes, however, that "it is not doubted that by other factor methods the same data might be shown to produce a general factor. In light of several advantages of the multiple factor approach it was felt advisable to remain within the limits of Thurstone's centroid method and to impose arbitrarily no general factor on the solution" (Jones, 1949, p. 317).

On the basis of the increased number and clarity of factors at the older age levels, Jones further argues that his results support the hypothesis that mental abilities tend to become increasingly differentiated with increasing age.

Jones (1954) subsequently repeated the analysis of the data from the thirteen year old group. This time more centroid factors were extracted from the correlation matrix in order to assure as complete a factorial description of the items as possible. Oblique rotations of centroid axes were substituted for orthogonal rotations in an effort to obtain a more precise definition of factors. The thirty distinct items on Forms L and M at mental age levels XII, XIII and XIV were included in the analysis. A total of ten centroid factors were extracted, the first three being those extracted by McNemar and the remainder extracted by Jones from McNemar's third factor residuals. Items having rotated factor loadings of .30 or greater were considered to be significant contributors to the factor.

The first three factors were interpreted as verbal comprehension factors. All of the items which loaded significantly on the first verbal factor also appeared high on the verbal factor discovered by orthogonal rotation (Jones, 1949) and the relative loadings on items were similar. Loading on the factor were vocabulary, abstract words, opposites, dissected sentences, and minkus completion tasks. The interpretation of this factor was restricted to "the process of supplying previously learned linguistic responses, primarily word definitions" (p. 133). The second verbal factor was defined as "an ability to manipulate words in a manner such that an appropriate meaningful relationship is imposed" (p. 134). Tasks associated with this factor included minkus completion, verbal absurdities, dissected sentences, reasoning, and abstract words. The third verbal factor, according to Jones, was less amenable to interpretation but was thought to represent "the verbalization of gross ideas as contrasted with the definition or manipulation of words which serve as elements of these ideas" (p. 134). Representing this factor were response to pictures, problems of fact, memory for stories, dissected sentences and abstract words items.

The fourth and fifth factors were thought to represent two distinct memory factors roughly distinguishable along the dimension of meaningfulness. The first factor was thus defined as "the ability to reproduce immediately after presentation, a sequence of disconnected elements" (p. 135). Memory for digits items were representative of this category. The second memory factor was defined as the "ability for verbatim recall of meaningful verbal material" (p. 136). Memory for sentences and memory for words items were included in this grouping.

The sixth factor appeared to be a spatial factor and was thought to involve "the visualization of movement within a particular configuration" (p. 137). Orientation of direction, memory for stories (the development of a visual image of events was thought to aid performance here), pictorial absurdities, and induction items (also requiring vizualization of events) loaded on this factor.

The seventh factor was classified as a reasoning factor. Problems of fact, ingenuity, and induction items were representative of this class.

The characteristic common to solutions of items on the eighth factor appeared to be "the ability to fuse a perceptual field into a single percept" (p. 140) and was subsequently called a closure factor. Plan of search and picture absurdities items were representative of this domain.

The ninth factor was tentatively classified as a perceptual carefulness factor and was thought to be reflective of the "ability to carefully and precisely perceive the details of a spatial configuration" (p. 142). Paper cutting, memory for designs, copying a bead chain from memory, and picture absurdities items contributed significantly to this factor.

The tenth factor did not contribute significantly to the total variance and hence was not interpreted.

From these results Jones concluded that despite the inclusion of four additional centroid factors and despite the oblique system of rotated factors, the findings were not radically different from those achieved earlier with orthogonal rotation of only six factors. The verbal factor of the earlier analyses was represented by three verbal factors in this analysis. The memory factor became more refined. The visualization factor remained essentially unchanged. The first reasoning factor of the previous study became the closure factor in this analysis. The second reasoning factor in the original investigation and the reasoning factor in the later analysis were virtually identical, and finally, the spatial factor found previously was interpreted as the perceptual carefulness factor here.

2.3.3 <u>The 1960 Revision</u>

Ramsey and Vane (1970) found seven factors as the result of a factor analysis of years IV-6 through VI of the 1960 revision of the Stanford-Binet. Test results of 152 children ranging in age from three to seven were employed in the analysis. The test protocols were selected from a larger group of 227 that had been obtained as part of a routine intellectual evaluation carried out in nursery schools and Head Start day care centres. The 152 tests were selected on the basis of having at least four passes and four failures across the mental age levels selected for analysis. This criterion was set in order to prevent the appearance of spuriously high correlations between subtest items. The IQ's of the sample ranged from 76 to 157. The mean IQ was 109.4.

The matrix of tetrachoric correlation coefficients between the eighteen items was subjected to a principal components analysis. Factors with eigenvalues greater than one were extracted. Loadings of .30 or greater were considered significant in interpreting the factors. Twelve of the eighteen subtests loaded on the first factor. Six remaining factors were bipolar and hence difficult to interpret. To facilitate interpretation, all the factors were rotated by the varimax method. This results in a solution which is approximately equivalent to the rotations to 'simple structure' described previously.

The principal components solution is of such a nature that it extracts the maximum amount of variance for the first factor and, therefore, the chances of finding a common factor are maximized. In this case, however, Ramsey and Vane did not feel that the first factor warranted a common factor or 'g' interpretation having only twelve of the eighteen items loading on it. Hence, they rejected the hypothesis that the Stanford-Binet in this age range could be explained in terms of a general factor.

The first factor, thus, was interpreted as a verbal factor and included vocabulary, comprehension and opposite analogies items. Picture completion and maze tracing items were also surprising contributors to this factor. This was rationalized as being due to the fact that comprehension of instructions would be an important variable mediating the performance on these items.

The subtests which loaded on the second factor were thought to all involve visual or visual-motor abilities. These included maze tracing, copying a square and aesthetic comparisons tasks.

The third factor was identified by only one item and hence was not interpretable.

Items associated with the fourth factor were thought to all require visual ability and judgement. Pictorial similarities and differences, and picture completion items were representative of this factor.

The fifth factor was, according to Ramsey and Vane, somewhat more difficult to interpret. The items loading on this factor, however, were thought to all involve some aspect of acquired knowledge. The number of items having low loadings on this factor, however, argued against this being a 'g' type factor.

The two items loading on the sixth factor seemed to involve control of impulsivity. One of the tasks involves carrying out a series of orally presented commands (Three Commissions) and the other requires solving orally presented arithmetic problems (Number Concepts). Ramsey and Vane also admitted that this, alternatively, could be a memory factor.

Items defining the seventh factor were thought to require the use of visual imagery.

Hallahan, Ball and Payne (1973) examined the factorial composition of the Short Form of the 1960 revision of the Stanford-Binet. (Four items at each age level are marked by asterisks and make up the Short Form.) The sample included 363 children enrolled in Head Start Programs. The children ranged in age from three to six years. The mean age was 4.3 years. The mean IQ was 91.6. Only items passed by more than 10% but less than 90% of the subjects were included in the analysis. This left fifteen items across the mental ages III-6 to V. Raw data was converted to a matrix of tetrachoric correlations. Factors were extracted by means of principal axes analysis. Communalities were estimated by iteration. On the basis of the Scree test three factors were extracted and rotated orthogonally according to the varimax procedure. Loadings with values of .30 or greater were considered in the interpretation of factors.

Three subtests did not load on the first factor. These were copying a square, opposite analogies and picture completion tasks. Hallahan et al. argued that since the items not loading were the three most difficult subtests and were clearly separated in terms of degree of difficulty from the other subtests, the first factor could appropriately be defined as a common factor.

The second factor was linked to Ramsey and Vane's factor of visual ability and judgement. The grouping of items was identical in both studies within the overlapping mental age ranges.

The nature of the subtests loading on the third factor strongly suggested, according to Hallahan et al., a verbal factor. The presence of Picture Completion: Man on this factor puzzled these investigators. It was noted, however, that this item also loaded on Ramsey and Vane's verbal factor. In addition to this item, four out of six other overlapping items corresponded with those on Ramsey and Vane's verbal factor. Three commissions and definitions items also loaded on this factor in this study but not in the Ramsey and Vane study.

2.3.4 Summary of Findings From Previous Factor Studies

All of the studies reviewed here were consistent in their identification of some salient group factors in the Stanford-Binet (see Table 2.1). The exact number and nature, however, varied somewhat from investigation to investigation. The existence of a general factor was not confirmed by all the investigators. A number of difficulties are encountered in attempting to compare the various results. Most obvious, of course, is that analyses were conducted on three different revisions of the scale. While the scales were generally very similar in content, some new items were added and others deleted from revison to revision, and still others were altered in their placement within the scale. Two forms of the 1937 revision provided the added advantage of a larger item pool with which to work. In addition to this problem of multiple revisions, the nature of the Stanford-Binet, being an age scale, means that it is impossible to include the entire scale within any one analysis since all subjects are not administered all items. This means that investigators have examined different portions of the scales. Subject populations have also varied, not only in age range, but in IQ range as well. These differences have been combined with variations in statistical methodology. Factor extraction procedures have differed. Criteria for deciding upon the number of factors to be extracted have not been identical across studies nor has the use of rotational procedures. Despite these complications, however, if the factor structure of the Stanford-Binet is indeed significantly clear as to be useful diagnostically there most surely should be some agreement between studies within comparable mental age levels.

Table 2.1

Summary of Results of Previous Factor Analytic Studies of the Stanford-Binet

Study	Location of Items	Age of Subjects	Factors
Burt & John (1942)	X, XII	10-0 to 11-6	'G', Maturity, Verbal, Vocabulary, Number, Comprehension, Spatial, Memory
Wright (1939)	VII-XIV	10-0 to 10-11	'G' or Maturity, Verbal, Number, Spatial, Reasoning, Induction
McNemar (1942)	II-S.A.	2-0 to 18-11	'G' only-but some indication or 'motor', 'naming', 'numerical', 'verbal', 'memory', and 'problem' tests forming distinct groupings
Jones (1949)	VII-VIII	7-0 to 7-11	Verbal, Reasoning, Memory, Number
	VIII-X	9-0 to 9-11	Verbal, Spatial, Reasoning, Memory
	X-XII	11-0 to 11-11	Verbal, Spatial, Memory
	XII-XIV	13-0 to 13-11	Verbal, Reasoning I., Reasoning II, Memory, Visualization, Spatial
Jones (1954)	XII-XIV	13-0 to 13-11	Verbal I, Verbal II, Verbal III, Rote Memory, Meaningful Memory, Spatial, Reasoning, Closure, Perceptual Carefulness

Table 2.1

Continued

Study	Location of Items	Age of Subjects	Factors
Ramsey & Vane (1970)	III-6-V	3-0 to 7-11	Verbal, Visual and Visual-Motor, Visual Ability and Judgement, Acquired Knowledge, Control of Impulsivity, Visual Imagery
Hallahan et. al. (1973)	IV-6-VI (Short Form)	3-0 to 6-0	'G', Visual Ability and Judgement, Verbal

Wright (1939) and Burt and John (1942a,b) both analyzed the first Stanford revision of the Binet. Wright analyzed items across the mental age levels VII to XIV employing a large sample of ten year old subjects ranging widely in mental age. Burt and John analyzed items at years X and XII employing a large sample of subjects ranging in age from ten to fourteen and having average IQ's. Both found evidence for a general factor. Burt and John also interpreted their first bipolar factor to be a maturity factor on the basis of mainly negative loadings for the X year items and positive loadings for the XII year items. After rotation of the centroid axes to simple structure (bipolar factors replaced by more directly interpretable group factors), however, examination of the saturation coefficients reveals that this interpretation is no longer clear. Only three items loaded .30 or greater on the maturity factor, one at age X and two at age XII. These were mixed sentences, abstract words and verbal absurdities items. There does not appear to be any obvious reason why these items would be more sensitive to age than would, say, the vocabulary, items whose loadings on this factor were near zero.

Both investigators also identified numerical, spatial and verbal group factors. Examination of the items loading on these factors which were common to both studies reveals almost perfect agreement, the only exception being the comprehension item which loaded on the verbal factor for Wright but did not load significantly on the equivalent factor for Burt and John. Rather, it loaded on a separate comprehension factor. Wright did not identify a separate comprehension factor. Burt and John also identified a separate vocabulary factor which Wright did not. While the vocabulary and abstract words items loaded significantly on the verbal factor in both studies, they were not separated out as a separate grouping in Wright's study. Wright identified a reasoning factor which Burt and John did not. With one exception, however, none of Wright's reasoning items were included in the Burt and John study. Furthermore, in examining Burt and John's items, one does not encounter any other items which would seem to be tapping reasoning in any significant way. Burt and John also identified a separate memory factor. This included two repeating digits items and a memory for designs item. The repeating digits items were the two items which also determined Burt and John's numerical factor. In Wright's study two repeating digits items, counting, and making change items loaded on the numerical factor. No separate memory factor was interpreted. It could be argued, however, that three of Wright's numerical items also require auditory memory skills since the making change item requires the retention of data given orally within the context of a word problem. The making change item was not included in the Burt and John study.

From this it can be seen that, within the limits of common items, there were a significant number of similar groupings of items in both studies.

McNemar (1942) and Jones (1949, 1954) both analyzed the standardization data from Forms L and M of the 1937 revision. McNemar included every item of both forms within at least one of fourteen separate analyses. Jones repeated four of the analyses employing the data from the seven, nine, eleven and thirteen year old subjects. The four analyses combined covered all the test items from year VIII to year XIV. He later repeated the analyses at experimental age thirteen extracting more factors and employing oblique rotations. McNemar extracted three centroid factors in each analysis. The sizeable first centroids were interpreted as evidence of a common factor in each case.

In conducting his analyses McNemar's main hope was to confirm that the majority of the test variance was attributable to a single common factor and not the influence of specific group factors. This he felt was vital if one were to argue for the equivalence of IQ scores. If significant group factors did exist this would mean that equivalent scores might actually differ qualitatively. McNemar did, however, find that a number of items did have low loadings on the common factor. He admitted the possibility of isolating meaningful group factors by means of rotating centroid axes. At the younger ages he found some evidence for the 'motor', 'naming' and 'numerical' tests to form distinct groupings. At the middle age ranges 'verbal' and 'memory' tests appeared to be separable. At the upper age levels there was a tendency for the 'verbal', 'problem', and 'memory' types of tests to be separated. McNemar's assumption that "these small 'group factors' could not contribute sufficiently to IQ variance to invalidate the comparability of IQ's of the same magnitude for individuals of approximately the same life age" prevented him from further attempting to clarify their nature (McNemar, 1942, p. 123).

Jones, however, took on this task. He extracted several more centroids from McNemar's original data and then applied a series of orthogonal rotations of the centroid axes in an attempt to clarify any group factor. Jones did not interpret the original first centroids as common factors, choosing to argue that the majority of variance of the test could be explained in terms of group factors. According to Jones, none of the rotated factors had enough significant loadings to be interpreted as a common factor.

Analyzing items at the VI to VIII year levels Jones isolated verbal, memory, number, and reasoning factors. At the VIII and IX year levels he found evidence for verbal, spatial, reasoning and memory factors. At levels X to XII verbal, memory and spatial factors were defined. Finally, at the XII to XIV year levels verbal, reasoning, memory, spatial and visualization factors were isolated. The complexity of the factor structure appeared to increase with age. In general these results confirmed and elaborated upon what McNemar had only hinted at. One can also see considerable similarity between the factors identified here and those identified by Burt and Wright. The visualization factor was the only one which was not identified by the earlier researchers. If the three items which defined this factor are identified, however, it is seen that two of the items were not included in the earlier studies. In comparing other factors item-by item, a good deal of correspondence is found across studies in the nature of many items defining the factors.

In his reanalysis of the data from the thirteen year old group, Jones isolated generally the same verbal, reasoning, memory and spatial factors. The original verbal factor, however, appeared as three factors which were thought to reflect more specialized facets of overall verbal skill. These might roughly be classified as vocabulary, fluency and comprehension. The memory factor was similarly divided to include meaningful and nonmeaningful components. The original reasoning factor was also further refined.

Ramsey and Vane (1970) and Hallahan et al. (1973) analyzed parts of the 1960 revision. Ramsey and Vane examined the scale over the age levels IV-6 to VI and Hallahan over age levels III-6 to V. Hallahan included only Short Form items. Mean IQ's were within the average range in both cases. Ramsey and Vane isolated seven distinct factors. They did not find evidence for a common factor before rotation of component axes. It should be noted, however, that these investigators interpreted those items loading .30 or greater on the factor as significant contributors to that factor. Others (i.e: McNemar) did not use as strict a criterion in interpreting a common factor. Ramsey and Vane found verbal, visual-motor, visual ability and judgement, general knowledge, control of impulsivity, and visualization factors. The verbal factor was similar in nature to the verbal factor identified by others. no visual-motor factor had been identified by the earlier investigators but, with the exception of McNemar, previous studies did not include items at these mental age levels. McNemar did notice a grouping of motor items with his younger groups. The visual ability and judgement factor appears similar to Jones's perceptual organization factor but the items loading on this factor had not been included in previous studies with the exception of McNemar's. The general knowledge factor had also not been previously encountered. Ramsey and Vane insisted that it was not broad enough to be a 'general intelligence' factor. The control of impulsivity factor similarly had not been identified in the earlier studies. The two items which defined this factor had previously been identified as reflecting primarily auditory memory components. It is possible, however, that at very young ages, it is control of impulsivity which is the important variable. In any case, two items are not sufficient for clearly defining a factor. Visualization factors were identified in several of the previous investigations and the types of defining items were fairly consistent.

Hallahan identified only three factors in her analysis using the Short Form. These included a common factor (in constrast to Ramsey and Vane), a visual and visual-motor factor, and a verbal factor. The verbal factor, within the constraints of common items, was almost identical to Ramsey and Vane's. The visual and visual-motor factor paralleled very closely Ramsey and Vane's visual ability and judgement factor.

In summary, then, while previous factor analytic studies reviewed here have varied in a number of respects, there is considerable evidence that some consistent group factors which may be helpful in diagnosis can be identified at various age levels in the earlier revisions of the Stanford-Binet. The data currently available on the 1960 revision, however, is limited. Ramsey and Vane's study examined the younger age levels and found some relatively clear factors. Stormer (1967), while using a somewhat different procedure, also found evidence for verbal, reasoning, memory and spatial aptitudes at the upper levels. Based upon the studies of earlier revisions and the history of the test's development, the factor structure is not likely to be identical across the entire span of the test and what needs to be undertaken is a series of further factorial investigations which will eventually examine the entire scale.

It has been mentioned previously that in the absence of clear empirical factors having been identified so far in the 1960 revision of the Stanford-Binet, clinicians have resorted to alternate item classification schemes developed from a conceptual basis. In the following section two of the most frequently used systems will be described.

2.4 Conceptual Models of Item Classification

2.4.1 Sattler's Model

According to Sattler (1965), his classification schema is based upon somewhat arbitrary groupings according to item content. A factor analysis was not conducted nor judges employed for reliability estimates of categorizations. Sattler admits that some items could be classified in different categories but that a case can be made for each of the current placements on the basis of content or face validity.

Sattler's system includes seven major categories each representing a general area conveying the nature of the function measured. Five subcategories deliniate more specific functions.

The Language category involves items tapping maturity, extent and quality of vocabulary. Tests such as vocabulary, abstract words, word naming, rhymes, and sentence building would be included in this category.

Conceptual Thinking, while closely associated with language ability, is primarily concerned with abstract thinking. Similarities, opposite analogies, differences, and proverbs are examples of the kinds of tests placed in this category.

The Reasoning category is subdivided into verbal and nonverbal reasoning. According to Sattler, the verbal absurdities items are the prototype for this category. The pictorial absurdities and orientation problems represent a model for the nonverbal reasoning items. Reasoning includes the perception of logical relations, discrimination ability, and analysis and synthesis. A spatial reasoning factor is also included in some items.

Numerical Reasoning includes items specifically geared to numerical problems. Items included in this category usually involve a significant concentration component as well.

The Memory category is subclassified into meaningful, non-meaningful, and visual memory components. Memory for digits, memory for sentences, and memory for designs would all be included in this category.

Sattler describes the Visual-Motor category as containing items concerned with manual dexterity, eye-hand coordination, and perception of spatial relations. Constructive visual-imagery may be involved in such items as paper folding. Nonverbal reasoning ability is allied to this area.

Finally, the Social Intelligence category includes aspects of social maturity and social judgement. Comprehension and finding reason items reflect social judgement, whereas the items concerning obeying simple commands, response to pictures, and aesthetic comparison reflect social maturity in the opinion of Sattler. This category apparently overlaps with the reasoning category so that items in the later category should also be considered to reflect social comprehension.

Sattler has assigned each item of the Stanford-Binet, including alternates, to one and only one of these categories or subcategories. While indicating that some categories may overlap, he has not taken this into account in the assignment of items.

Sattler (1965) provides a breakdown by age of the proportion of items, not including alternates, in each category. In general, Language is the most highly represented category across the entire span of the test. In the II to V age level reasoning and numerical reasoning items do not occur, whereas visual-motor and nonverbal reasoning items each occupy 21% of the distribution second only to language items which occupy 26%. In the VI to X age level language and memory items each occupy 19% of the distribution. In the XI to XIV level visual-motor items are not represented. Language, reasoning, and memory items predominate. At the adult levels conceptual thinking items occupy the second rank with 27% of the distribution, whereas social intelligence, visual-motor and visual memory items do not occur.

Based upon this data the previously mentioned indicators tha the Stanford-Binet does not measure the same factors at each age level would seem to be further supported. Sattler, himself, argues that his data support a developmental model of intelligence. At the younger age levels language, visual-motor capacities, nonverbal reasoning and social awareness are skills of primary importance. The child has not acquired the capacity for functions such as abstract reasoning, numerical reasoning, and memory to any large extent. During the middle years more advanced cognitive functions are developed. Stress is given to memory functions and the development of temporal perspective. Capacity for abstract reasoning and conceptual thinking begins. The vast importance of conceptual thinking and reasoning is evident at the upper age levels. According to Sattler, the capacity for symbolic thinking coupled with language functions portrays the shorthand of higher intellectual functions.

Sattler (1974) has also outlined a method for profiling specific areas of strength and weakness on the Stanford-Binet. Even given the well-defined classification system, the clinician still has difficulty determining whether an individual's performance in one area of functioning is significantly better or worse than in other areas. The age-level format of the scale does not permit a simple way of calculating significant differences between tests which are passed and those which are failed. Differences between successive year levels do not mean the same thing throughout the scale. The difference, for example, between year level II and year level III represents approximately a 50% increase in mental development, whereas the difference between year X and year XI represents an increase in mental development in the range of 10%. Hence for a two year old passing tests at the III year level may indicate superior functioning, while for a ten year old passing tests at age XI may be within the limits of average functioning. Sattler's Standard Deviation method reduces the possiblity of over-interpretation of minor fluctuations in a child's performance and allows the clinician to more accurately evaluate the scatter in a Binet profile. The Standard Deviation method is based on the assumption that tests failed within plus or minus one standard deviation of the child's chronological age or mental age represent expected fluctuations in ability. Tests passed or failed within these limits are not considered to indicate specific strengths or weaknesses.

Kaufman and Waterstreet (1978) prepared a table which allowed for the simple determination of the band of year levels surrounding a child's chronological or mental age demarcing the range of normal variability. The chronological age reference point would be used for normative comparisons. The mental age reference point allows for intraindividual comparisons.

Chase and Sattler (1980) subsequently refined the table by reducing the size of age increments and thus allowing for more precise determination of the range of normal variability.

2.4.2 Valett's Model

Valett's (1964) classification system is also based upon conceptual groupings of items. He includes six major categories: Vocabulary and Verbal Fluency, General Comprehension, Judgement and Reasoning, Arithmetic Reasoning, Visual-Motor Ability, and Memory and Concentration. Valett does not distinguish specific subcategories of these. Unlike Sattler, Valett allows for the inclusion of items in multiple categories. There is considerable overlap, for example, between General Comprehension, and Judgement and Reasoning categories. Items included in Valett's Memory, Visual-Motor, and Arithmetic categories are almost identical to those included in the parallel categories within Sattler's model. Valett's Vocabulary and Verbal Fluency category overlaps significantly with Sattler's Language category. General Comprehension has many items in common with Sattler's Social Intelligence category. Judgement and Reasoning is most similar to Sattler's Conceptual Thinking and Reasoning categories. Silverstein (1965) calculated coefficients of overlap between the categories in the two schemes and when the categories with the greatest amount of overlap were taken as equivalent, the schemes agreed in classifying 75% of the total items. It would appear, then, that Sattler's and Valett's models have much in common despite their somewhat different category labels (see Table 2.2).

2.5 Summary

In this chapter, the historical development of the Stanford-Binet Intelligence Scale has been traced. A study of the theory and research which preceded the development of the first Binet scale provided some indication that an examination of its factor structure might be fruitful. Previous factor analytic studies of the several revisions revealed the scale to be factorially complex. A number of difficulties relating to differences in methodology were encountered in attempting to compare the results of the various studies. Despite this, evidence was present that some useful factors could be reliably isolated at the various ages in the earlier revisions. Insufficient efforts have been made to date, however, to define the factor structure across the entire age span of the 1960 revision.

Table 2.2

Conceptual Factors Proposed by Sattler (1965) and Valett (1964) for the Stanford-Binet Form L-M

Sattler	Valett
Language	
Conceptual Thinking	
Verbal Reasoning	Vocabulary and Fluency
Nonverbal Reasoning	General Comprehension
Numerical Reasoning	Judgement and Reasoning
Meaningful Memory	Arithmetic Reasoning
Nonmeaningful Memory	Visual-Motor Ability
Visual Memory	Memory and Concentration
Visual-Motor Ability	
Social Intelligence	

In the absence of information pertaining to the empirical factor structure, several alternate item classification systems have been developed out of a purely conceptual framework. The systems of Sattler and Valett were described. While the categories in these systems have at least some superficial similarity to the empirical factors, an indepth comparison has yet to be made between the factor structure of the 1960 Stanford-Binet at various ages and these conceptual models. The following chapters will describe a preliminary step in this direction.

CHAPTER THREE

3. EXPERIMENTAL DESIGN

3.1 Introduction

Part A of the chapter is devoted to the factor analytic investigation of the Stanford-Binet. The design of the study is outlined. No effort is made, however, to discuss in detail the statistical methods employed. This has been reserved for the following chapter. The sample groups are described and sampling adequacy is evaluated.

Part B describes the methods employed in the Clinician Judgement Study. The sample group is described and sampling adequacy is evaluated. The integration of the Factor Analytic Study and the Clinician Judgement Study is clarified.

PART A: THE FACTOR ANALYTIC STUDY

3.2 The Design

The factor structure of the Stanford-Binet Intelligence Scale Form L-M was examined over the mental age range VI to XI. Several reasons existed for the selection of this particular range for examination. Firstly, due to its particular suitability for younger children, the Stanford-Binet is often the test of choice for preschoolers and students in the earlier elementary grades. Somewhat older children, unless gifted or severely mentally retarded, tend to more frequently be given the Wechsler Intelligence Scale For Children (WISC-R). Hence it was desired to employ younger subjects in the current study. The total sample, thus, included children ranging in age from two years to twelve years eleven months. Sufficient data was then available from this group to allow for analysis of levels VI through XI.

A second consideration influencing the decision concerned the fact that Ramsey and Vane (1970) examined years IV-6 through VI of the scale. The current study provided a natural extension of that work.

Separate sets of analyses were conducted employing all distinct items at year levels VI through VIII, and at year levels VIII through XI. Children ranging in age from three through six were included in the analyses at the younger levels. Children ranging in age from seven through twelve participated in the analyses at the older levels.

Both sets of data were subjected to a principal components analysis followed by Varimax rotation of significant components. The resultant component solution in each case was compared to that obtained when Varimax rotational methods were replaced by Promax methods.

Both orthogonal and oblique Procrustes rotational procedures were then subsequently employed in an attempt to fit the principal components to Sattler's conceptual factors across both age ranges.

3.3 The Samples

3.3.1 The Younger Group

Sixty-eight males and forty-two females were included in the younger group. Their test protocols were chosen from a larger group of 146 on the basis of having at least three passes and three failures out of the eighteen items contained at year VI through VIII. Thus, older or extremely bright children who might have failed only one or two items in this range, and younger children or children with lower IQ's who might have passed only one or two items in this range were excluded.

The children ranged in age from 3.08 years to 6.92 years, the mean age being 6.51 ± .96 years. The IQ's of the group ranged from 84 to 158 with a mean of 112.18 ± 16.82. Mental ages ranged from 4.42 years to 8.83 years with a mean of 6.57 ± .95 years.

All of the children had been assessed at the Education Clinic at the University of Alberta in Edmonton, Alberta between 1979 and 1983. The reasons for drawing samples from this clinical population were twofold. Firstly, it was felt that the kinds of problems being experienced by the children referred to the Education Clinic would typify, to some extent at least, the kinds of problems routinely encountered by Educational Psychologists. The majority of other factor analytic studies of the Stanford-Binet have employed normal populations. A second even more pragmatic reason was the ready availability of a sufficient number of test protocols within the target age range.

Of the 110 children in this group, sixty-six had been referred to the Clinic by parents, thirty-eight by school personnel, and six by other agents such as physicians or social workers. Thirty percent were referred by parents whose interest was simply in obtaining information on the child's current level of intellectual functioning. No specific referral problems were specified. Of the remaining seventy percent, approximately twenty-five percent of the children were referred for problems in language and/or reading. Behavioral, emotional and social problems were specified by the referral agent in twenty-one percent of the cases. Questions of possible giftedness or the need for enrichment constituted thirteen percent of the referral reasons. Problems in fine or gross motor development were specified in twelve percent of the cases. Arithmetic problems were specified in seven percent of the clinical reports. Concerns about speech development were expressed by the referral agent in four percent of the cases. Finally, sixteen percent of the referral questions were specifically related to school readiness.

These percentages were calculated from information gathered by the researcher from each assessment report. Classifications were made on the basis of the referring questions as outlined by the clinicians and not on the basis of test results or final diagnoses. Many of the children presented with several problems and hence were included in multiple categories. Behaviorial problems, for example, frequently coexisted with academic problems. Concerns about language development, speech, or behavior were often behind requests for school readiness assessments.

3.3.2 The Older Group

Sixty-two males and forty-eight females were included in the older group. Their test protocols were chosen from a larger group of 155 on the basis of having at least three passes and three failures out of the twenty-four items contained at year VIII through XI.

The children ranged in age from 6.67 years to 12.42 years, the mean being 8.67 $\pm$ 1.33 years. The IQ's of the group ranged from 72 to 163 with a mean of 103.46 $\pm$ 16.38. Mental ages ranged from 6.50 years to 15.58 years with a mean of 9.25 $\pm$ 1.60 years.

All of the children had been assessed at the Education Clinic at the University of Alberta between 1979 and 1983. Of the 110 children in this group, thirty-one had been referred by parents, seventy-six by school personnel, and three by social workers. Sixteen percent of the children were referred by parents or school officials whose interest was mainly in obtaining information on the child's current level of intellectual functioning. No specific referral problems were specified in clinicians' reports. Of the remaining eighty-four percent approximately forty-seven percent of the children were referred with problems in reading, spelling and/or language. Problems of a behaviorial, emotional or social nature were specified by the referral agents in thirty percent of the cases. Arithmetic difficulties were being experienced by twenty-one percent of the children. Fourteen percent of the referrals involved questions of possible giftedness. Motor problems were of concern in ten percent of the cases. Speech difficulties were specified in six percent of the reports.

3.4 Sampling Adequacy

With the exception of the forty-five cases excluded in the older sample and the thirty-six cases excluded in the younger group, test data from virtually every Stanford-Binet administered to children aged three to twelve between September 1979 and February 1983 was included in the analyses. The eighty-one cases excluded tended to involve children at the extremes of IQ or age since these were the children who did not meet the pass/fail criteria across the target items. This is reflected in the larger standard deviation of IQ scores of the deleted groups. The mean IQ of the children deleted from the younger group was 106.56 $\pm$ 28.38. The mean IQ of the children deleted from the older group was 100.78 $\pm$ 29.83. Hence, except possibly at the extremes of IQ the sample groups were likely representative of the population of children between the ages of three and twelve routinely administered the Stanford-Binet at the University of Alberta Education Clinic.

The Education Clinic is a free clinic and exists primarily for the purpose of training graduate students in School Psychology and Counselling programs. It accepts referrals from parents, schools and other outside agencies. The majority of children, however, tend to come from middle or upper socioeconomic backgrounds. Children seen at this clinic also, on the whole, have problems of a relatively less severe nature than might be encountered in some clinical or school settings. This is reflected in the nature of the two sample groups. The mean IQ levels were well into the average range. No children with severe physical or sensory handicaps were included. The nature of the emotional or behavioral problems would not, in the majority of cases, be classified as severe. Academic difficulties tended to range from mild to moderate. Furthermore, a significant number of the referrals were made out of interest or concerns about possible giftedness. This was particularly true in the case of younger children.

PART B. THE CLINICIAN JUDGEMENT STUDY

3.5 The Design

Interrater reliability of assignment of items to Sattler's classification scheme for the Stanford-Binet Form L-M was evaluated. Graduate student clinicians enrolled in the winter or spring sessions of the introductory practicum in School Psychology at the University of Alberta were provided with the titles and definitions of each of the ten categories outlined by Sattler. They were also provided with a list of all the Stanford-Binet items, excluding alternates, from year II to Superior Adult III grouped by year. Printed instructions requested the students to rate each of the items according to which of Sattler's abilities it was most felt to be measuring at the given age level. This was to be done by placing the appropriate ability number (i.e. 1-10) beside each item. Spaces were also provided to allow the students to signify, where applicable, a maximum of two other abilities which they felt an individual item to also be significantly tapping. The instructions were to rate these other abilities on a priority basis (see Appendix I).

Ratings were completed by each class during a ninety minute lecture period. Use of the Stanford-Binet manual was encouraged. Instructions were also reviewed orally by the investigator.

3.6 The Sample

Forty-two graduate students, twelve males and thirty females, completed the ratings. All were enrolled in either the winter or spring session of the Seminar and Practicum in School Psychology in the Department of Educational Psychology at the University of Alberta. The emphasis of this practicum is on training students in the administration and interpretation of individualized intelligence tests including the Stanford-Binet. Thirty-one of the raters were in their first or second year of a Master's of Education program in Counselling and School Psychology or Special Education. Seven were in their first or second year of the Doctoral program in Counselling and School Psychology. Four were working towards post-graduate diplomas in Special Education. The majority were between twenty-six and thirty-five years of age. Ten students were older and three were younger. Twenty-four of the raters had previous teaching experience. Seven had taught for more than ten years. Half of the students had some previous clinical experience in testing or counselling outside of the University setting. Of these, eight had some experience directly in intellectual assessment. For the remaining raters, this practicum was their first exposure to individualized intelligence testing. The twenty-two students in the winter class had all administerd at least three Stanford-Binet's prior to completing the ratings. Most had administered five and several had administered more. A number of the students in the spring class had not yet administered the test and few had administered more than one or two. The students, however, had been introduced to Sattler's classification system in their lectures and readings and, hence, were familiar with it on a theoretical if not on a practical basis.

3.7 Sampling Adequacy

This sample was obviously less than ideal in a number of respects. Firstly, the number of raters was small. Secondly, the majority of the students were inexperienced in the use of the Stanford-Binet and many were inexperienced in psychoeducational assessment in general. While this study is a prototype of the kind of research which needs to be undertaken in evaluating the conceptual item classification models of Sattler and others, the extent to which the current results may be generalizable will unfortunately be somewhat limited.

3.8 Integration of the Two Studies

The Factor Analytic and the Clinician Judgement Studies combined allowed for evaluation of Sattler's conceptual classification schema for the Stanford-Binet from two different perspectives. The extent to which Sattler's conceptual factor structure approximates the empirical factor structure and, within the limits of the restricted sample, the extent to which the schema is reliably used by clinicians could both be assessed.

CHAPTER FOUR

4. DATA ANALYSES

4.1 Introduction

Part A of the chapter is devoted to description of the statistical methods employed in the factor analytic investigation of the Stanford-Binet. Statistical criteria for the selection of items included in the analyses are outlined. Scaling of the sample observations is described. The theoretical basis of principal components analysis is explained in terms of hyperspace geometry. The decision to employ principal components analysis rather than common factor analysis is defended. The eigenvalue-one criterion for the number of factors is explained and defended. Orthogonal and oblique rotations of the principal components are described. Finally, the Procrustes technique of fitting the principal components to Sattler's conceptual factor structure is outlined.

Part B describes the method employed to assess the agreement between the student clinicians and Sattler on the assignment of the Binet items to the conceptual categories.

PART A: THE FACTOR ANALYTIC STUDY

4.2 Item Selection

4.2.1 The Younger Group

As outlined in the preceding chapter the portion of the Stanford-Binet to be analyzed within the context of the current investigation was in part determined by the existence of a sufficient number of test protocols in which the pass/fail criterion of at least three passes and three failures was met. For the younger group this occurred at the test ages VI through VIII. To have included protocols in which most items were passed or in which most items were failed would have resulted in spuriously high correlations between items.

Once the target portion of the test had been selected, however, a further issue needed to be addressed. The nature of the Stanford-Binet is such that a number of the items are repeated with differing passing standards at several age levels. Within the test age range VI through VIII two items are repeated. Vocabulary occurs at year VI and at year VIII. Comprehension IV occurs at years VII and VIII. This poses some difficulty in that the performance on two identical test situations, although scored differently, cannot be regarded as yielding two experimentally independent variables and hence the correlation between the two items would be spuriously high (McNemar, 1942, p. 104). Some investigators who have previously factor analyzed portions of the Binet have chosen to ignore this problem and have included repeated items in the analysis (Wright, 1939; Burt & John, 1942a,b). Others, on the other hand, chose to include only distinct items (McNemar, 1942; Jones, 1949; 1954). In the present investigation each item was included only once. In cases where items occurred at multiple year levels, the item was included at the year where the number passing it approximated most closely the ideal of fifty percent. Variance of dichotomous variables is a maximum when the proportion passing is 0.5, and the product-moment correlation between two dichotomous test items (phi) is thus restricted by the extent to which the percentage of persons passing the first item differs from the percentage of persons passing the second item (Nunnally, 1978, p. 143). Thus, Vocabulary was passed by 79 percent of the younger children at year VI and 39 percent at year VIII. The item was eliminated at year VI. Similarily, Comprehension IV was passed by 46 percent of the younger children at year VII and 20 percent at year VIII. The item was therefore eliminated at year VIII. This left a total of 16 items remaining within the target years VI through VIII. Of these items only those which were passed by more than 10 percent and less than 90 percent of the children were to be considered for inclusion in the analyses. This criterion was employed to prevent serious restriction of the inter-item correlations. The remaining 16 items, however, all met this difficulty criterion and the elimination of further items was unnecessary. The items included in the younger group analyses and their location within the test are given in Table 4.1. The mean (i.e. the proportion passing) and the variance for each item are presented in Table 4.2.

4.2.2. The Older Group

Sufficient data meeting the pass/fail criterion was available across test ages VIII through XI from the older sample. Within this test range two items are repeated. Vocabulary occurs at age VIII and again at age X. At age VIII 95 percent of the older children passed this item. At age X 36 percent passed the item. Following the same criterion as for selection of items in the younger sample, the Vocabulary item was eliminated at year VIII. Similarily, the Memory for designs I item occurs at age IX and again at age XI. It was passed by

Table 4.1

Stanford-Binet Items Included in Analyses of Younger Group Data

Item Name	Item Location
(1) Differences	VI-2
(2) Mutilated Pictures	VI-3
(3) Number Concepts	VI-4
(4) Opposite Analogies II	VI-5
(5) Maze Tracing	VI-6
(6) Picture Absurdities I	VII-1
(7) Similarities: Two Things	VII-2
(8) Copying a Diamond	VII-3
(9) Comprehension IV	VII-4
(10) Opposite Analogies III	VII-5
(11) Repeating 5 Digits	VII-6
(12) Vocabulary	VIII-1
(13) Memory for Stories: The Wet Fall	VIII-2
(14) Verbal Absurdities I	VIII-3
(15) Similarities and Differences	VIII-4
(16) Naming the Days of the Week	VIII-6

Note: Numbers in ( ) may be substituted for item names in some tables.

Table 4.2

Mean and Variance of Items Included in the Younger Group Analyses

Item Name	Mean	Variance
(1) Differences	.755	.185
(2) Mutilated Pictures	.836	.137
(3) Number Concepts	.836	.137
(4) Opposite Analogies II	.791	.165
(5) Maze Tracing	.864	.118
(6) Picture Absurdities I	.491	.250
(7) Similarities: Two Things	.355	.229
(8) Copying a Diamond	.464	.249
(9) Comprehension IV	.455	.248
(10) Opposite Analogies III	.555	.247
(11) Repeating 5 Digits	.255	.190
(12) Vocabulary	.391	.238
(13) Memory for Stories: The Wet Fall	.336	.223
(14) Verbal Absurdities I	.118	.104
(15) Similarities and Differences	.273	.198
(16) Naming the Days of the Week	.200	.160

61 percent of the children at year IX and by 23 percent of the children at year XI. Of the remaining 22 items within the target age range all were passed by more than 10 percent and less than 90 percent of the children. It was not necessary to eliminate any further items from the analyses. The items included in the older group analyses and their location within the test are presented in Table 4.3. The mean and variance for each item are given in Table 4.4.

4.3 Scaling of the Sample Observations

Test responses from both the older and younger groups were coded dichotomously with 1.0 representing items passed and 0.0 representing failed items. Each set of raw data (i.e. the younger and the older) was subsequently converted to a correlation matrix of phi coefficients. The phi coefficient is the product-moment correlation between dichotomous variables. In the factor analytic studies of the Stanford-Binet described previously tetrachoric correlations were used in place of phi coefficients. The tetrachoric correlation coefficient is used to estimate the product-moment correlation of two continuous, normally distributed variables from dichotomized versions of these variables. Employing the tetrachoric coefficient with data which is inherently dichotomous at the time of the analysis would have the aim of estimating what the product-moment correlation would be if the two variables were continuous. Nunnally (1978, p. 136-137), however, presents several persuasive arguments against the use of tetrachoric correlations. Firstly, it is usually illogical to estimate what the product-moment coefficient would be if both dichotomous variables were

Table 4.3

Stanford-Binet Items Included in Analyses of Older Group Data

Item Name	Item Location
(1) Memory for Stories: The Wet Fall	VIII-2
(2) Verbal Absurdities I	VIII-3
(3) Similarities and Differences	VIII-4
(4) Comprehension IV	VIII-5
(5) Naming the Days of the Week	VIII-6
(6) Paper Cutting	IX-1
(7) Verbal Absurdities II	IX-2
(8) Memory for Designs I	IX-3
(9) Rhymes: New Form	IX-4
(10) Making Change	IX-5
(11) Repeating 4 digits Reversed	IX-6
(12) Vocabulary	X-1
(13) Block Counting	X-2
(14) Abstract Words I	X-3
(15) Finding Reasons I	X-4
(16) Word Naming	X-5
(17) Repeated 6 Digits	X-6
(18) Verbal Absurdities IV	XI-2
(19) Abstract Words II	XI-3
(20) Memory for Sentences II	XI-4
(21) Problem Situation II	XI-5
(22) Similarities: Three Things	XI-6

Note: Numbers in ( ) may be substituted for item names in same tables.

Table 4.4

Mean and Variance of Items Included in the Older Group Analyses

Item Name		Mean	Variance
(1)	Memory for Stories: The Wet Fall	.809	.154
(2)	Verbal Absurdities I	.564	.246
(3)	Similarities and Differences	.791	.165
(4)	Comprehension IV	.764	.180
(5)	Naming the Days of the Week	.827	.143
(6)	Paper Cutting	.509	.250
(7)	Verbal Absurdities II	.445	.247
(8)	Memory for Designs I	.609	.238
(9)	Rhymes: New Form	.809	.154
(10)	Making Change	.464	.249
(11)	Repeating 4 digits Reversed	.664	.223
(12)	Vocabulary	.355	.229
(13)	Block Counting	.400	.240
(14)	Abstract Words I	.318	.217
(15)	Finding Reasons I	.345	.226
(16)	Word Naming	.436	.246
(17)	Repeated 6 Digits	.336	.223
(18)	Verbal Absurdities IV	.191	.154
(19)	Abstract Words II	.227	.176
(20)	Memory for Sentences II	.245	.185
(21)	Problem Situation II	.181	.104
(22)	Similarities: Three Things	.464	.249

continuous. Unless direct steps are taken to convert the dichotomous variables into continuous variables, such estimates serve only to create the impression that the variables have more explanatory power than they actually do. Nunnally notes that it is often tempting to employ tetrachoric coefficients rather than phi coefficients because the former are always larger except in the rare case where the proportion of subjects 'passing' (i.e. the p values) are equivalent for both dichotomous variables.

A second reason, according to Nunnally, for not employing the tetrachoric coefficient with dichotomous data is that it is frequently a very poor estimate of the product-moment correlation. The tetrachoric coefficient depends upon a very strict assumption of normality of the underlying continuous variables. If the assumption of normality is not met the resulting estimate may be quite inaccurate.

Finally, the most obviously salient of Nunnally's arguments against the use of tetrachoric correlations for the current investigation is that it is illegitimate from a purely mathematical point of view to employ estimates of product-moment correlation coefficients in any type of multivariate analysis including factor analysis. So strongly does Nunnally feel about the importance of these arguments that he concludes his discussion of tetrachoric correlation with the statement that "there is nothing wrong with using tetrachoric correlations in mathematical models relating to psychometric theory, but they should definitely not be used to determine the correlation between sets of empirical data" (p. 137).

4.4 Principal-Components Analyses

4.4.1 Definition and Theoretical Background

The correlation matrix developed from the younger group data and the correlation matrix developed from the older group data were both subjected to principal components analysis using the Fact20 program (DERS-45-184, Division of Educational Research Services, The University of Alberta, 1980).

Principal components analysis and related models such as common factor analysis belong to the more general class of mathematical techniques which fall under the rubric of factor analytic methods. The term 'factor analysis' stands for a broad category of approaches to conceptualizing groupings of variables and for determining which variables belong to which group. Thus, factor analysis involves conceptualizing variables as relating to underlying factors and performing analyses that will test for or discover such factors (Nunnally, 1978, p. 327-328). "In very gross and simplified terms, the problem posed by factor analysis is to find a set of composites or latent variables of a single set of observed variables that will account for the variation among subjects on observed variables" (Thorndike, 1978, p. 242). Criteria are usually set in a given situation that will serve to dictate the character of the composites. For example, one is most often interested in composites which are as few in number as possible and can account for the greatest proportion of observed variance (Thorndike, 1978, p. 242).

Component-factor analysis begins with a rectangular matrix of data. The matrix contains the standardized scores of N 'persons' on n measures. The n variables can be viewed as being plotted in standardized space of N people. Given n linearly independent variables in a space of N people, there is an n-dimensional subspace of the N space that completely contains the n variable vectors. The requirement of linear independence means that n dimensions must be used to completely describe the subspace. The geometric relations among the variables are described by the correlation matrix. The correlations in the matrix are the cosines of the angles between the variable vectors. Thus, an analysis of the correlation matrix is an analysis of the variable space directly and the people space of N dimensions can be ignored (Thorndike, 1978, p. 242-243). Often, then, "factors are spoken of as dimensions and factoring is spoken of as dimensionalizing of a space of variables" (Nunnally, 1978, p. 344).

The loading of a variable with a factor is represented by the cosine of the angle between the vector for the variable and the vector for the factor. The sum of the squared loadings in each row of the resultant factor matrix will be equal to 1.00. That is to say, that all the variance of a particular variable is accounted for by the set of n factors and that the sum of cross products in any two rows of the matrix of factor loadings will equal the correlation between the two variables (Nunnally, 1978, p. 345).

If, however, n dimensions are required to completely describe the n-variable space, of what possible benefit is carrying out a complex mathematical procedure to obtain as many composites as variables? One benefit of the composites is that they are uncorrelated with one another. Describing n space in terms of n correlated variables is more complex than describing it in terms of n uncorrelated dimensions since this eliminates the necessity of accounting for the relationships between the variables.

A second reason for employing uncorrelated composites in place of the original variables is that the composites are generally of unequal size and importance. The standardized variables each have unit variance. It is impossible to confirm that any one is more imporant or pervasive in its influence than another. The composites, however, vary in size, making it possible to conclude that one is more significant as a dimension of variance than another. In fact, it is argued by factor theorists that the small composites can virtually be ignored as they account for a relatively small proportion of the variance of a set of variables and do not markedly contribute to the understanding of the relationships among the variables.

The rank of a matrix is the number of linearly independent sources of variation among the variables in the matrix. The rank of a correlation matrix is usually equal to the number of variables. When n dimensions are used to describe an n-variable space, the result is termed a full rank solution and any variable vector can be located exactly. It is most often the case, however, that fewer dimensions can quite adequately describe the relationships among the variables and a reduced rank solution is possible (Thorndike, 1978, p. 245-247).

Variables will frequently be correlated with each other. When this occurs the vectors representing the variables tend to fall into groups. Vectors that are grouped together represent variables that have much in common. Other groups of vectors that are largely independent of the first group might also exist, as might other groups which have negative correlations with the variables in the first group. If each of the n variables was uncorrelated with the others, the variable vectors would form the axes of a hypersphere of n dimensions. As some of the variables become correlated, however, the hypersphere would become an n-dimensional hyperellipsoid. The length of the hyperellipsoid in any direction from the origin is a function of the correlations among the variables. If the first axes (composite or component) is located so that it is the longest possible axis of the figure, that axis will have the highest average of squares of correlations of the variables with the axis. In other words, the sum of the squared factor loadings will be a maximum. By locating the second axis so that it is orthogonal to the first and follows the second longest dimension of the hyperellipsoid the same criterion is applied to the remaining covariation. When each succeeding axis is placed at right angles to those preceding it and through the longest remaining dimension of the figure, a set of principal components is obtained. In this way the first principal component will account for the highest proportion of the variance in the set of variables. The second principal component will account for the next highest proportion of the variance and so forth. The later components account for such a small amount of the variance that they are usually discarded (Thorndike, 1978, p. 252-255).

In summary, then, the principal components are the minimal orthogonal dimensions required to linearly reproduce the original data (Rummel, 1970, p. 338). The principal components are mathematically unique for a data matrix in which the variables are not all uncorrelated. The process of developing composites whether under these restrictions or some others is generally referred to as the extraction of factors.

To this point the terms 'factor' and 'component' have been used interchangeably in reference to the composites. This was done in order to facilitate the exposition. In reality, however, the term 'component' refers to a composite of a set of variables developed from an unreduced correlation matrix, that is, one in which 1's are contained in the principal diagonal. The term 'factor' is more accurately employed in association with common-factor analytic models.

4.4.2 Comparison of Common-Factor and Component-Factor Models

In the general factor analysis model the total variance of a variable is assumed to result from three sources. The common variance is that portion of the total variance that a variable shares with one or more of the other variables in a set. The specific variance of a variable is that portion of the total variance which is unique to that variable. Finally, the error variance is that portion of the total variance that is due to error of measurement. Each of these sources of variance is uncorrelated with the other two. In most actual analyses no effort is made to separate specific variance from error variance. These are usually lumped together and called unique variance. The aim of factor analysis, then, is to attempt to separate that variance of variables that can be explained by common factors from the unique variance. In principal components analysis (or any other type of components-factor analysis where 1's are placed in the principal diagonal of the correlation matrix) this aim is only partially achieved. The principal components are linear combinations of variables and hence the two sources of variation become somewhat mixed because the variables themselves, including the uniquenesses, determine the factors (Nunnally, 1978, p. 405-406).

In light of this problem, a number of techniques have been developed which seek mathematical solutions that define factors entirely in terms of this common variance among the variables. These techniques usually involve substituting some measure or estimate of the common variance - communalities - in place of unities in the principal diagonal of the correlation matrix. This limits the factor analysis to delineating common factors. For example, one approach to common-factor analysis is to employ communality estimates in the principal diagonal of the correlation matrix and then to submit it to the same type of analysis as described above for principal components. This is known as a principal axes solution (Nunnally, 1978, p. 406).

While substituting communalities in the principal diagonal may appear to be a simple solution to the problem of unique variance, some difficulties do emerge. Firstly, the correlation of any variable with itself, of course, is 1.0. Hence, the diagonal elements of any 'real' correlation matrix are unities. Secondly, if a factor loading is defined as the correlation of a standardized variable with a linear combination of a set of variables, then to compute that loading, the formulas require that unities be placed in the principal diagonal of the correlation matrix. If anything other than unities are employed, one is not correlating an actual variable with a linear combination of actual variables. Thus, the resultant factors are, in reality, hypothetical. They are hypothetical in that they can only be estimated from the actual variables (Nunnally, 1978, p. 405-406).

A second difficulty involves the determination of the communalities themselves. The communalities cannot be known until the common factors are defined. The delination of these factors, however, depends upon the correlations and the communalities in the principal diagonal of the correlation matrix. This inability to determine precisely the communalities creates a basic indeterminancy in the common factor model. It then becomes necessary to insert some estimates of the communalities in the correlation matrix. The resulting common factors, then, can themselves only be estimates to some degree of the true common factors of the data (Rummel, 1970, p. 312).

The squared multiple correlation of a variable with all the others has been shown to be the lower bound of the communality (Guttman, 1957), and is frequently felt to be the best estimate (Rummel, 1970, p. 167). A number of other approaches, however, do exist but will not be discussed here.

While arguments exist for the use of both component-factor analyses and common-factor analyses, component analyses have the distinct advantage of being conceptually simple and straightforward mathematically. Considerable evidence exists which indicates that in most instances common factor analysis and principal components analysis produce similar results if there are truly any clear groupings of variables in the correlation matrix (Velicer, 1976; Jackson & Chan, 1980; Velicer, Peacock & Jackson, 1982). It was for these reasons that the decision to employ principal components analysis in the current investigation was made.

4.4.3 Eigenvalue-one Criterion for Extraction of Principal Components

The extraction of principal components is usually terminated at the point where no further meaningful variance remains. The definition of meaningful variance, however, in component (and common) factor models is a matter of some debate. The best number of factors in any case depends to a considerable extent upon the researcher's judgement and the nature of the questions being asked. It is a problem for which no definitive solution yet exists. There are, however, a number of criteria both of an inferential and mathematical nature which can be used to guide the decision, all having some advantages and some disadvantages. Some of the most commonly used criteria include the size of the residual correlations (i.e. the difference between the reproduced correlations calculated from the factors extracted and the actual correlation), the distribution of factor loadings, the distributions of eigenvalues, and factor interpretability (Rummel, 1970, p. 169).

In the present investigation the eigenvalue-one criterion was employed in the determination of the number of principal components to be extracted from both the younger group and the older group data. In other words, only those components having eigenvalues of 1.0 or greater were retained for possible interpretation.

The eigenvalues equal the sum of the column of squared factor loadings for each component. They measure the total amount of variance accounted for by a component for the variables as a group. The eigenvalue-one criterion was first proposed by Kaiser (1960) as the 'best' solution to the number-of-factors problem. The eigenvalue-one criterion evolved in part as an application of Guttman's (1954) work which demonstrated that the number of eigenvalues of a factor matrix that are greater than 1.0 constitutes a "weak lower bound" on the number of factors - that is, the smallest number of factors that can adquately explain the data variance. Kaiser (1960) further argued that the eigenvalue-one criterion was reasonable in that it excluded the factors not accounting for at least the total variance of a single standardized variable. Kaiser (1960) also demonstrated that for a factor to have adequate positive Kuder-Richardson reliability it is necessary and sufficient that its eigenvalue be greater than unity. Finally, Kaiser (1960) observed that the eigenvalue-one criterion produces the most psychologically meaningful factors. While the eigenvalue-one criterion has since been questioned on a number of grounds it is still one of the most widely applied criteria (Rummel, 1970, p. 362).

4.5 Rotational Procedures

The original principal components, while they may be adequate to define the minimum dimensionality of the data, they are frequently difficult to interpret. As a result of the stipulation that each successive component account for the maximum available variance, the components are frequently located between independent clusters of interrelated variables and, thus, the clusters cannot be easily defined in terms of their loadings on these components. The first factor, for instance, will frequently have high loadings for most of the variables and, conversely, most of the variables may have high loadings on a number of factors (Rummel, 1970, p. 373).

A second important characteristic of the unrotated factor solution is that the initial location of the factor (component) axes in the configuration of vectors cannot be controlled. This has important implications for the case where the researcher has some preconceived notion of the factor structure of a set of variables and wishes to employ factor analysis to confirm or reject his hypothesis. The existence of a hypothesized cluster of variables may be obscured by the variance associated with other salient characteristics of variables (Rummel, 1970, 375).

In light of these important characteristics of the unrotated principal components, it is common practice for factor theorists to rotate the original set of components to achieve (according to some predefined criteria) a more useful result.

In the current investigation the unrotated principal components in both the younger and older group analyses were subjected to orthogonal Varimax, oblique Promax, and orthogonal and oblique Procrustes rotational procedures. The aim of the Varimax and Promax rotations was to aid in the deliniation of separate groupings of intercorrelated variables. The Procrustes methods of rotation were conducted in order to test the hypothesis that the factor structure of the Stanford-Binet is consistent with Sattler's conceptual factors.

4.5.1 The Varimax Rotations

Varimax is generally considered to be the best analytic orthogonal rotational technique (Nunnally, 1978, p. 305; Rummel, 1970, p. 392; Harman, 1967, p. 311; Harris, 1964). The Varimax rotational procedure involves maximizing the variance of the columns of squared factor loadings, hence, the name Varimax. As there are more high and low loadings on a factor (component), the variance of the squared factor loadings is larger. The varimax criterion for orthogonal rotation comes close to achieving the goal of simple structure. Simple structure attempts to maximize the number of components with high loadings for each variable. This tends to make each component define a distinct cluster of intercorrelated variables assuming, of course, that such variables exist. The advantage to such a solution is that there is little doubt about which variables are contributing variance to which factors. The task of interpretation is eased if the solution approximates simple structure.

This orthogonal rotational procedure does not alter the position of the variables since it is the reference axes which are moved. The length of the variable vectors also remain unchanged by the rotation. Geometrically, the square root of the communalities equals the length of the variable vectors. The communalities, therefore, also remain constant. Finally, the variance accounted for by the major unrotated components is spread across the rotated components. The rotated components tend to account for approximately equal amounts of variance.

4.5.2 The Promax Rotations

While the Varimax rotational technique maintains the orthogonality of the unrotated principal components, the Promax technique relaxes the orthogonality requirement to allow the factors to become correlated. In allowing the factors to become correlated, it is assumed that an even better approximation to simple structure can be achieved. Furthermore, it is frequently argued that given a set of highly intercorrelated variables (such as the Stanford-Binet items), it is reasonable to expect that the dimensions themselves would also be intercorrelated.

The Promax technique involves constructing an ideal matrix from the Varimax solution in which high loadings of Varimax are increased and the low loadings are decreased. This is accomplished by normalizing the orthogonal matrix by rows and columns and taking the fourth power of each loading. The oblique solution is the least squares fit to this ideal matrix (Rummel, 1970, p. 419-420).

In oblique rotation two factor matrices are deliniated, namely, the pattern matrix and the structure matrix. The matrix of pattern loadings results when the loadings of a variable on the axes are determined by lines parallel to the axes. The pattern loadings "may be interpreted as measures of the unique contribution each factor makes to the variance of the variables. They measure the dependence of the variables on the different factors, and in this sense they are regression coefficients of the variables on the factors" (Rummel, 1970, p. 399). Since the pattern loadings cannot be interpreted as correlations, the loadings squared do not precisely give the percentage of variation of the variables accounted for by a component. The pattern loadings, however, best show what clusters of variables are associated with the oblique components, and, hence, the pattern matrix is vital in interpreting the components.

The matrix of structure loadings results when loadings of a variable on the axes are determined by lines perpendicular to the axes. "The structure loadings are the product-moment correlations of the variables with the oblique factors [components]" (Rummel, 1970, p. 399).

In orthogonal rotation the pattern and structure loadings are identical and only one factor matrix is required. The loadings can be interpreted either as pattern loadings or correlations.

Instead of interpreting the oblique component axes (i.e. the primary axes) directly, a new coordinate system is frequently defined by placing alternate axes through the origin perpendicular to each of the primary axes. These are known as the reference axes and are felt by some to give a slightly better deliniation of simple structure. In the current investigation, however, the pattern loadings on the primary axes only were used in the interpretation of the component-factors.

4.5.3 The Orthogonal and Oblique Procrustes Rotations

To this point the concern has been with identification and clarification of dimensions inherent in the data. In other words, the approach has been exploratory in nature. Once the dimensions were identified the next step was to assess how consistent the statistical dimensions (i.e. the principal components) revealed in the exploratory investigation of both the older and younger group data were with the conceptual factors outlined by Sattler.

The Procrustes technique of confirmatory analysis was employed. This approach involves the forced rotation of the principal components so as to approximate a hypothesized factor structure. A target factor matrix is estimated. The hypotheses to be confirmed are stipulated by the nature of the target matrix (Nunnally, 1978, p. 400).

Sattler's conceptual factors formed the basis of the target matrices employed. To reduce the size of the matrices, however, the categories of Meaningful Memory, Nonmeaningful Memory, and Visual Memory were combined to form a single Memory dimension. Similarly, Verbal Reasoning, Nonverbal Reasoning, and Numerical Reasoning were combined to form a single Reasoning Category. This reduced the total number of Sattler's conceptual factors to six (see Table 4.5). This reduction was necessary because the relatively small number of items which were included in both the younger and older group analyses were insufficient to adequately define the larger (i.e. ten factor) target matrices.

A six factor target matrix was developed from the 16 younger group items. Items assigned to each factor by Sattler were allotted factor loadings of 1.00. All other items were given loadings of 0.00 (see Table 4.6). The Procrustes solutions were the least squares fit of the component-factors to this target matrix. In the case of the orthogonal Procrustes solution a limiting condition that the components remain orthogonal was placed on the solution. The oblique Procrustes method removed this restriction to allow the factors (components) to become correlated.

A similar six factor matrix was developed from the 22 older group items (see Table 4.7). Both oblique and orthogonal Procrustes rotations assessed the fit of the component-factors to the target factors.

Table 4.5

Target Factors For Procrustes Rotations of Principal Components
(As Derived from Sattler's Conceptual Model)

I	LANGUAGE	
II	MEMORY	Meaningful Memory Nonmeaningful Memory Visual Memory
III	CONCEPTUAL THINKING	
IV	REASONING	Verbal Reasoning Nonverbal Reasoning Numerical Reasoning
V	VISUAL-MOTOR INTEGRATION	
VI	SOCIAL INTELLIGENCE	

Table 4.6

Target Matrix For Procrustes Rotations
of Principal Components - Younger Group
(Derived From Sattler's Assignment of Items to Conceptual Categories)

	Component-Factor					
Item	I	II	III	IV	V	VI
(1)	0.00	0.00	1.00	0.00	0.00	0.00
(2)	0.00	0.00	0.00	1.00	0.00	0.00
(3)	0.00	0.00	0.00	1.00	0.00	0.00
(4)	0.00	0.00	1.00	0.00	0.00	0.00
(5)	0.00	0.00	0.00	0.00	1.00	0.00
(6)	0.00	0.00	0.00	0.00	0.00	1.00
(7)	0.00	0.00	1.00	0.00	0.00	0.00
(8)	0.00	0.00	0.00	0.00	1.00	0.00
(9)	0.00	0.00	0.00	0.00	0.00	1.00
(10)	0.00	0.00	1.00	0.00	0.00	0.00
(11)	0.00	1.00	0.00	0.00	0.00	0.00
(12)	1.00	0.00	0.00	0.00	0.00	0.00
(13)	0.00	1.00	0.00	0.00	0.00	0.00
(14)	0.00	0.00	0.00	1.00	0.00	0.00
(15)	0.00	0.00	1.00	0.00	0.00	0.00
(16)	0.00	0.00	0.00	0.00	0.00	1.00

Table 4.7

Target Matrix For Procrustes Rotations of Principal Components - Older Group (Derived from Sattler's Assignment of Items to Conceptual Categories)

Item	Component-Factor I	II	III	IV	V	VI
(1)	0.00	1.00	0.00	0.00	0.00	0.00
(2)	0.00	0.00	0.00	1.00	0.00	0.00
(3)	0.00	0.00	1.00	0.00	0.00	0.00
(4)	0.00	0.00	0.00	0.00	0.00	1.00
(5)	0.00	0.00	0.00	0.00	0.00	1.00
(6)	0.00	0.00	0.00	0.00	1.00	0.00
(7)	0.00	0.00	0.00	1.00	0.00	0.00
(8)	0.00	1.00	0.00	0.00	0.00	0.00
(9)	1.00	0.00	0.00	0.00	0.00	0.00
(10)	0.00	0.00	0.00	1.00	0.00	0.00
(11)	0.00	1.00	0.00	0.00	0.00	0.00
(12)	-1.00	0.00	0.00	0.00	0.00	0.00
(13)	0.00	0.00	0.00	1.00	0.00	0.00
(14)	1.00	0.00	0.00	0.00	0.00	0.00
(15)	0.00	0.00	0.00	1.00	0.00	0.00
(16)	1.00	0.00	0.00	0.00	0.00	0.00
(17)	0.00	1.00	0.00	0.00	0.00	0.00
(18)	0.00	0.00	0.00	1.00	0.00	0.00
(19)	1.00	0.00	0.00	0.00	0.00	0.00
(20)	0.00	1.00	0.00	0.00	0.00	0.00
(21)	0.00	0.00	0.00	0.00	0.00	1.00
(22)	0.00	0.00	1.00	0.00	0.00	0.00

PART B: THE CLINICIAN JUDGEMENT STUDY

4.6 Proportion of Clinicians Agreeing With Sattler's Assignment of Stanford-Binet Items to Categories

The proportion of student clinicians in agreement with Sattler's assignment of each Stanford-Binet item (excluding alternates) to the ten conceptual categories was calculated. The proportions were calculated on the basis of the students' first rating only. The degree of consensus was then examined within each conceputal category. Where applicable, the students' assignment of items was also compared to the loadings on the empirical factors.

CHAPTER FIVE

5. RESULTS

5.1 Introduction

Part A of the Chapter is devoted to the results of the factor analytic investigation. The item intercorrelations are presented for the variables included in both the younger and older group analyses. The unrotated factor matrices are discussed. The factors emerging from both analyses are defined after both Varimax and Promax rotation. The orthogonal factors are compared to the oblique factors in each case. Results of the exploratory analyses are summarized and the factor structure is compared across the two groups.

Results of the confirmatory rotations of the empirical factors to Sattler's conceptual dimensions are discussed. The implications of these results for Sattler's model are outlined.

Part B of the chapter is devoted to the results of the Clinician Judgement Study. The proportion of clinicians agreeing with Sattler's assignment of items to the ten conceptual categories on the first choice is presented for each Stanford-Binet item. An evaluation of the overall degree of consensus between Sattler, the student clinicians and, where applicable, the results of the factor analyses is made. Finally, an integration of the results of the Clinician Judgement Study and the Factory Analytic Study is undertaken.

PART A: THE FACTOR ANALYTIC STUDY

5.2 Item Intercorrelations

5.2.1 The Younger Group

The intercorrelations for the 16 items included in the younger group analyses are shown in Table 5.1. The lower off-diagonal elements only are presented since the matrix is symmetric. As all the values along the principal diagonal are unity, these have been replaced by the squared multiple correlations. The squared multiple correlation coefficient of a variable with all the other variables multiplied by 100 measures the percentage of linear variation that can be explained for that variable by the others. The squared multiple correlations were generally moderate and ranged from 0.18 to 0.41.

As expected, the majority of the intercorrelations were positive. Four of the correlations were negative but three of these did not meet criteria for statistical significance. In order for a correlation to achieve statistical significance (i.e. to be significantly different from zero) at the conventional 0.05 level its absolute value must be greater or equal to 0.19 (degrees of freedom = 108). The correlation of -0.22 between variables (13) and (14) (i.e. between Memory for Stories: The Wet Fall and Verbal Absurdities I items) is statistically significant but the reason for the negative relationship is not readily explained. It may be attributable to measurement error. The same two items had a small positive, but not statistically significant, correlation in the analysis employing the data from the older sample.

It is noteable that the positive correlations between the items were not as high as would be anticipated given the nature of the test

Table 5.1

Item Inter-Correlations - Younger Group

Item	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	0.29							
(2)	0.03	0.15						
(3)	0.09	0.07	0.23					
(4)	0.28	0.08	0.32	0.25				
(5)	-0.04	0.11	0.04	0.12	0.23			
(6)	0.18	0.14	0.24	0.24	0.13	0.24		
(7)	0.33	0.17	0.23	0.15	0.02	0.30	0.33	
(8)	0.06	0.12	0.26	0.25	0.32	0.18	0.19	0.33
(9)	0.31	0.21	0.21	0.16	0.26	0.27	0.20	0.29
(10)	0.21	0.15	0.15	0.12	0.07	0.64	0.32	0.14
(11)	0.04	-0.08	0.20	0.15	-0.01	0.01	0.13	0.25
(12)	0.24	0.25	0.20	0.23	0.26	0.11	0.30	0.19
(13)	0.18	0.16	0.11	0.18	0.17	0.03	0.20	0.19
(14)	0.14	0.16	0.01	0.12	0.15	0.26	0.26	0.22
(15)	0.21	0.16	0.16	0.11	0.18	0.22	0.40	0.17
(16)	0.18	0.16	0.16	0.03	0.13	0.19	0.25	0.31

Note: Squared Multiple Correlations have been placed in the principal diagonal. Only those elements below the principal diagonal are shown. Variables are listed in Table 4.1

Table 5.1

Continued

Item	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1)								
(2)								
(3)								
(4)								
(5)								
(6)								
(7)								
(8)								
(9)	0.32							
(10)	0.16	0.27						
(11)	0.10	0.15	0.18					
(12)	0.28	0.38	0.09	0.34				
(13)	0.16	0.29	0.16	0.34	0.21			
(14)	0.23	0.10	-0.02	0.17	-0.22	0.34		
(15)	0.34	0.30	0.16	0.30	0.17	0.47	0.41	
(16)	0.09	0.31	0.23	0.25	0.13	0.03	0.26	0.30

and the method by which it was designed (i.e. each item having a significant positive correlation with the total test score). A number of the positive correlations were also nonsignificant. This holds true even if the assumption is made that all the intercorrelations should be positive (i.e. negative correlations are interpreted as being equivalent to zero). This is a reasonable assumption as one would not expect significant negative relationships amongst items all designed to correlate highly with the total test score. In this case, for the one-tailed test, a correlation must be equal to or greater than 0.16 in order for it to be considered statistically significant. The fact that phi coefficients have been employed rather than tetrachoric correlation coefficients would account to some extent for the generally lower intercorrelations obtained in this study than in the previous factor analytic investigations of the Stanford-Binet (e.g. Wright, 1939; Burt & John, 1942a,b).

In general, the positive correlations do not suggest any apparent singularities in the data. The highest positive correlation was 0.64 and, for the most part, the correlations were moderate suggesting that the factor structure will be multidimensional. The magnitude of the correlations would suggest that a one factor solution would be unlikely to account for the item intercorrelations.

5.3 The Unrotated Principal Components

5.3.1 The Younger Group

The matrix of factor loadings for the unrotated principal components extracted from the younger group data is shown in Table 5.2.

Table 5.2

Unrotated Principal Components - Younger Group

	Component-Factor					
Item	I	II	III	IV	V	h^2
(1)	0.46	0.02	-0.22	-0.46	-0.32	0.58
(2)	0.36	0.22	-0.20	0.31	-0.16	0.34
(3)	0.44	-0.27	0.41	0.25	-0.09	0.50
(4)	0.44	-0.14	0.38	0.23	-0.46	0.63
(5)	0.35	0.14	0.28	0.67	-0.06	0.67
(6)	0.45	0.34	0.32	0.28	-0.01	0.50
(7)	0.61	0.06	-0.25	0.34	0.13	0.57
(8)	0.53	-0.13	0.48	0.26	0.20	0.63
(9)	0.57	0.25	0.15	0.07	-0.17	0.44
(10)	0.52	-0.29	-0.46	0.04	0.07	0.58
(11)	0.29	-0.54	0.22	-0.13	0.40	0.60
(12)	0.62	-0.11	-0.26	0.24	-0.24	0.58
(13)	0.44	-0.37	-0.22	0.26	-0.33	0.55
(14)	0.44	0.62	0.01	-0.04	0.24	0.64
(15)	0.63	0.28	-0.19	-0.04	0.32	0.62
(16)	0.48	-0.29	-0.10	0.07	0.45	0.53
Eigenvalue	3.78	1.43	1.32	1.27	1.13	

Five component-factors having eigenvalues greater than 1.00 were retained.

The loadings on the first component-factor were all positive. The remaining four component-factors were characterized by both positive and negative loadings, that is, they were bi-polar. The first factor accounted for 23.64 percent of the total variance in the data and 42.36 percent of the common variance. The second component-factor accounted for 8.95 percent of the total variance and 16.04 percent of the common variance. The third factor explained 8.22 percent of the total variance and 14.73 percent of the common variance. The fourth factor accounted for 7.95 percent of the total variance and 14.25 percent of the common variance. Finally, the fifth factor accounted for 7.04 percent of the total variance and 12.62 percent of the common variance in the data. The five factors combined explained 55.81 percent of the total variance in the set of variables.

Variables having loadings of 0.30 or greater on a factor were employed in the psychological interpretation of that factor. Loadings of a smaller size are of doubtful significance since they indicate that less than 10 percent of the variance of a variable is explained by the factor (Nunnally, 1978, p. 423).

Loadings of the first factor were all significant with the exception of variable (11) which was the Repeating 5 Digits item. This item had a loading of 0.29. The presence of this large first factor was predicted and was consistent with the majority of previous findings. This indicates that all the items, with the possible exception of the Repeating 5 Digits item, share a common dimension. It is reasonable to expect that a general or 'g' factor would emerge when all the items were included in the test on the basis of their correlation with the total test score. If this 'g' factor is interpreted as a general intelligence factor, the fact that the Repeating 5 Digits item had a relatively low loading on this factor might be explained. The Repeating 5 Digits item would appear to involve primarily rote memory and would be less likely to require more complex manipulation of information.

No attempt was made to interpret the remaining bi-polar components prior to rotation.

5.4 Varimax Rotations of the Principal Components

5.4.1 The Younger Group

The Fact 20 program allows the user to define the minimum number of factors to rotate. In the case of the younger group analyses this parameter was set at two. Hence, the Varimax rotations were made for two, three, four and five factors. While statistical criteria are important in the determination of the correct number of factors, psychological interpretability of the solution is also a major consideration. "Those who are expert in the substance and literature of their domain and familiar with the nuances in their data may reject factors unless the factors have a certain subjective probability of significance or generalizability" (Rummel, 1970, p. 356). In light of the considerable data currently existing on the nature of abilities, the kinds of factors which would be expected in an investigation such as the one undertaken here were, to some degree at least, known.

Considerable theoretical guidelines existed to aid in assessing the psychological importance of the factors. The two, three, four, and five rotated solutions were evaluated for psychological interpretability. Items loading 0.30 or greater were considered significant in defining a factor.

Rotation of all five of the original components proved to be the most satisfactory solution in that the resultant dimensions were the most readily interpreted and the most consistent with those revealed in previous investigations. The final factor was a doublet factor (i.e. having only two significant loadings) indicating that there were not likely to be more than five signficant factors. Rotation of fewer components failed to clarify the dimensions sufficiently.

The matrix of rotated factor loadings for the five component-factors is shown in Table 5.3. The variables defining each factor are given in Table 5.4.

The first rotated component was defined by 5 of the 16 variables. There were no significant negative loadings. On the basis of the content of the defining items, this component was classified as a Verbal, Nonverbal Reasoning factor.

According to Sattler (1965), verbal reasoning includes the perception of logical relations in verbal material. The Verbal Absurdities I item requires the child to point out intellectually irreconcilable elements in statements such as "Walter now has to write with his left hand because two years ago he lost both his arms in an accident". In this case the child must comprehend the lack of logical relations in the information and verbal reasoning skill would be expected to play a key role in mediating success on this task.

Table 5.3

Varimax Rotation of Principal Components - Younger Group

	Component-Factor				
Item	I	II	III	IV	V
(1)	0.29	0.38	0.43	-0.07	-0.40
(2)	0.26	0.41	-0.04	-0.18	0.26
(3)	0.04	0.37	0.62	0.34	0.04
(4)	0.01	0.16	0.77	-0.01	0.06
(5)	0.11	0.21	0.05	-0.01	0.78
(6)	0.53	-0.11	0.46	0.02	0.07
(7)	0.54	0.34	0.18	0.25	-0.27
(8)	0.17	0.04	0.32	0.46	0.54
(9)	0.41	0.24	0.35	-0.05	0.28
(10)	0.16	0.65	-0.05	0.33	-0.14
(11)	-0.08	0.00	0.18	0.75	-0.02
(12)	0.18	0.70	0.16	0.07	0.17
(13)	-0.15	0.69	0.18	0.09	0.13
(14)	0.78	-0.04	0.00	-0.06	0.15
(15)	0.70	0.26	-0.02	0.24	0.04
(16)	0.22	0.27	-0.07	0.63	0.07
% Tot. Var.	13.50	13.08	10.79	9.79	8.65
% Com. Var.	24.19	23.43	19.33	17.55	15.50

Table 5.4

Items Having Significant Loadings After Varimax Rotation of Principal Components - Younger Group

I Verbal, Nonverbal Reasoning		II Verbal	
Item Name	Loading	Item Name	Loading
(14) Verbal Absurdities I	0.78	(12) Vocabulary	0.70
(15) Similarities and Differences	0.70	(13) Memory for Stories:	0.69
(7) Similarities: Two Things	0.54	(10) Opposite Analogies III	0.65
(6) Picture Absurdities I	0.53	(2) Mutilated Pictures	0.41
(9) Comprehension IV	0.41	(1) Differences	0.38
		(7) Similarities: Two Things	0.34

III Visualization & Visual Judgement		IV Control of Impulsivity	
(4) Opposite Anaologies II	0.77	(11) Repeating 5 Digits	0.75
(3) Number Concepts	0.62	(16) Naming Days Week	0.63
(6) Picture Absurdities I	0.46	(8) Copying a Diamond	0.46
(1) Differences	0.43	(3) Number Concepts	0.34
(9) Comprehension IV	0.35	(10) Opposite Analogies III	0.33
(8) Copying a Diamond	0.32		

V Visual-Motor Integration	
(5) Maze Tracing	0.78
(8) Copying a Diamond	0.54

The Picture Absurdities I task requires the examinee to point out the intellectually irreconcilable elements in a series of pictures. If verbal reasoning involves the perception of logical relations inherent in verbal information then, by extension, perception of illogical relations in nonverbal information would also be expected to involve nonverbal reasoning skills.

The Comprehension IV item requires the child to respond to questions of the nature "What should you do if you found on the streets of a city a three-year-old baby that was lost from its parents?". In order to succeed on this task the child must be able to anticipate cause and effect. For many older children such questions may simply require automatic or over-learned responses but for young children the task would be expected to involve some significant degree of reasoning ability.

The Similarities: Two Things test requires the child to relate two items such as wood and coal on the basis of some common property. This task demands that the child be able to form concepts and to classify objects and ideas on the basis of these concepts. Conceptualization and classification skills are just being developed between the ages of two and six. While these skills are strongly associated with overall language ability, reasoning ability is also likely to be important in mediating performance, especially at the younger age levels.

The Similarities and Differences item requires the child to point out both similarities and differences between objects such as an airplane and a kite. This task would, in general, be expected to involve the same skills as the Similarities: Two Things item.

The absence of the Differences task on this factor may be somewhat unexpected given the presence of the previous two items. This test requires the examinee to identify differences between objects such as wood and glass. This test, however, is obviously much easier than the Similarities: Two Things or the Similarities and Differences tests. This is reflected both by its occurrence at a younger age level and by the fact that it was passed by over 75 percent of the children in the sample. Hence, the need for exceptional reasoning skills may be less important for success on this task. General language ability is likely to be the most significant dimension here.

The second rotated component-factor was defined by 6 of the 16 items. There were no significant negative loadings. This dimension was identified as a general Verbal factor.

The Vocabulary test which requires the child to define a number of words would seem to be a straightforward measure of word knowledge. Word knowledge is one of the best indicators of overall language functioning. It quite obviously influences verbal comprehension and verbal expressive abilities.

The Opposite Analogies III item which requires the child to respond to questions of the type "The rabbit's ears are long; the rat's ears are . . .?" would appear to be tapping primarily conceptual thinking ability. The ability to form verbal concepts is strongly related to overall language functioning and probably to intellectual functioning in general. Conceptualization skills allow the child to relate objects and ideas and thereby to organize information and to make generalizations. The absence of the Opposite Analogies II item on this factor was, however, difficult to explain.

The Differences and the Similarities: Two Things items also tap conceptual thinking. The absence of the Similarities and Differences item on the Verbal factor seemed surprising. Finding both Similarities and Differences, however, appears to be considerably more difficult for very young children than does just finding similarities or just finding differences. It is possible, then, that for the young children, at least, reasoning ability is the vital dimension mediating performance on the Similarities and Differences test.

The Memory for Stories: The Wet Fall item requires the child to answer several questions about a story read by the examiner and the child if sufficient reading skills are present. While memory skills are undoubtedly tapped to some extent by this item, it is also very likely that verbal comprehension also is an important variable, particularly in the case of very young children. Somewhat older children would likely find the story to be quite simple and the role of verbal comprehension ability would likely be much less significant to the performance of the task. For the young children, however, it is likely to be very important.

The presence of the Mutilated Pictures item on this task may at first seem somewhat difficult to explain. This task requires the child to identify the parts which are missing in a set of pictures. The missing part, however, must be named or described verbally. Pointing is not sufficient. Word knowledge and verbal expressive skills are vital in the performance of the task. The pictures are sufficiently simple that many of the children are likely to recognize the missing part immediately. Young children may, however, encounter difficulty in producing the correct labels for the missing part.

The absence of the Comprehension IV and the Verbal Absurdities I items on this factor might be questioned. These tasks would appear to also tax the verbal comprehension ability of the child. Their presence, however, on factors I and III would indicate that verbal comprehension is considerably less important than some other skills in mediating performance on this task.

Factor III was defined by 6 of the 16 variables. There were no significant negative loadings. This factor proved to be somewhat more difficult to interpret but was tentatively identified as a Visualization and Visual Judgement Factor. The common processes appeared to be the formation of a mental image of an object or event in its absence and/or the making of a visual comparison or judgement.

While the Opposite Analogies II Item would appear to be very similar to the Opposite Analogies III item and would also be expected to tap verbal conceptual skills, it is possible that at the younger age level a different dimension becomes more influential in mediating performance on this kind of task. It may be that very young children, with their weaker language skills, depend more upon their ability to visualize the items to be compared. The presence of the Differences item on this factor would appear to provide some further support for this hypothesis.

The Number Concepts item requires the child to put a prescribed number of blocks in front of the examiner. Here, the child must judge when the correct number of blocks are in place. Very young children are given to making judgements on the basis of the immediate, perceived appearances of things. They are very much bound by their visual perception. It is not surprising that children having good visual judgement skills would be at an advantage in the performance of this task. They are likely to organize the blocks better visually and hence to make less errors in counting.

The Picture Absurdities I task quite obviously demands visual judgements on the part of the child. The child must decide if the situation depicted in the pictures is logical and, if not, decide what is wrong.

The Copying a Diamond item requires the child to reproduce a diamond-shaped figure from a model provided by the examiner. While visual-motor integration ability is likely to be the most important factor mediating performance on this task, the child must also make a visual comparison between his reproduction and the model. It is not uncommon for young children to exclaim confidently that they have reproduced the diamond accurately when they have not. Children having good visual judgement skills might be expected to notice the discrepencies and to take action to correct them.

The presence of the Comprehension IV item on this factor is more difficult to explain. It is possible that younger children who are able to visualize the events depicted in the questions are better able to relate cause and effect and are thus better able to find a satisfactory solution to the problem posed.

The fourth rotated component-factor was defined by 5 of the 16 variables. There were no significant negative loadings. Control of Impulsivity appeared to be the common element linking the items loading on this factor.

The Repeating 5 Digits item requires the child to recite from memory a series of five numbers presented orally by the examiner. While rote memory skills are likely important in determining success on this task, attention and control of impulsivity are also undoubtedly involved. It is not uncommon to encounter young children who are so anxious to respond that they barely listen or wait for the examiner to complete the sequence. These children are very likely to make errors which they may not have made had they taken the time to be more careful.

In responding to the Naming the Days of the Week item a similar tendency towards impulsivity is often observed. The child either leaves out days or instead of giving the day before Thursday, for example, will give the day after.

The Copying a Diamond task also seems to be sensitive to any tendency on the part of the child to be impulsive. Some children are in such a hurry to complete this task that their reproduction is inferior to what it might have been had they taken more time.

Similarily, impulsive children are apt to make careless errors on the Number Concepts item especially if they view it to be very simple. In their haste they make unnecessary errors.

The reason for the Opposite Analogies III item on this factor is less obvious. The impulsive child, however, might respond without stopping to remember that he must provide an opposite to the target concept. For example, one might receive the answer "wings" to the question "The dog has hair; the bird has . . .?" On the surface these questions might appear so simple that the impulsive child cannot stop himself from rushing into a careless response. It would seem, however, to be inconsistent that the Opposite Analogies II item is not present on this factor. The Opposite Analogies item was, in fact, eliminated from the factor upon oblique rotation indicating that, if present, the role of impulsivity is not major.

Finally, factor V was characterized by 2 of the 16 variables. It is usually considered unwise to interpret a factor having less than three significant loadings. This factor, however, appeared to be a clear Visual-Motor Integration factor. Apart from the two items which did load on this factor, there were no other variables within the analysis which would be expected to have a significant visual-motor integration component associated with them.

The Maze Tracing item requires the child to draw a line along the shortest path through a maze. The child's line cannot deviate outside the lines delineating the paths. For younger children visual-motor skill would be expected to play a very influential role in mediating the performance on this test.

The Copying a Diamond task would appear to be a fairly direct measure of visual-motor integration skills.

In summary, all five of the original principal components were interpreted after rotation to the Varimax criterion. Factor I was labelled a Verbal, Nonverbal Reasoning factor. Factor II was defined as a general Verbal factor. Factor III proved to be somewhat more difficult to interpret but was tentatively labelled a Visualization and Visual Judgement factor. Factor IV was classified as a Control of Impulsivity factor and factor V was defined as a Visual-Motor

Integration factor. In general, the factor structure was quite clearly defined.

Removal of the orthogonality restriction on the factor solution did not affect the nature of the factors significantly although it did allow for a slightly better deliniation of several of the dimensions. A summary of the Promax rotation of the factors follows.

5.5 Promax Rotations of the Principal Components

5.5.1 The Younger Group

The matrix of pattern loadings on the primary axes is presented in Table 5.5. Since the pattern loadings are not precisely interpretable as correlation coefficients and their magnitude can be somewhat greater, some investigators prefer to employ a higher cutoff in determining the significant variables defining each factor. The less correlated the oblique factors, however, the more their loading are like correlations of variables with factors (Rummel, 1970, p. 148). The factors emerging from the younger group data proved to have low to moderate intercorrelations and, in general, the magnitude of the loadings were similar to those of the orthogonal factors. For this reason, items having loadings of 0.30 or greater were retained on the factors. The variables defining each factor are given in Table 5.6.

The factor intercorrelations are presented in Table 5.7. The highest correlation was 0.44 and occurred between factors I and II (i.e. between Verbal, Nonverbal Reasoning and Verbal factors). Factor V (i.e. the Visual-Motor Integration factor) was virtually uncorrelated with the remaining factors. The magnitude of the factor

Table 5.5

Promax Rotation of Principal Components - Younger Group

Pattern on the Primaries					
Item	I	II	III	IV	V
(1)	0.18	0.31	0.40	-0.16	-0.42
(2)	0.21	0.42	-0.11	-0.24	0.27
(3)	-0.06	-0.04	0.62	0.30	0.01
(4)	-0.14	0.12	0.80	-0.07	0.03
(5)	0.07	0.23	-0.01	-0.05	0.79
(6)	0.55	-0.28	0.41	-0.04	0.02
(7)	0.50	0.20	0.08	0.17	-0.29
(8)	0.13	-0.03	0.26	0.43	0.52
(9)	0.36	0.16	0.29	-0.13	0.26
(10)	0.04	0.65	-0.15	0.26	-0.12
(11)	-0.13	-0.06	0.15	0.77	-0.03
(12)	0.02	0.72	0.07	-0.03	0.19
(13)	-0.35	0.78	0.14	0.02	0.16
(14)	0.89	-0.22	-0.11	-0.11	0.12
(15)	0.72	0.11	-0.17	0.17	0.03
(16)	0.18	0.20	-0.17	0.61	0.07

Table 5.6

Items Having Significant Loadings After Promax Rotation of Principal Components - Younger Group

I Verbal, Nonverbal Reasoning

Item Name	Loading
(14) Verbal Absurdities I	0.89
(15) Similarities and Differences	0.72
(6) Picture Absurdities I	0.55
(7) Similarities: Two Things	0.50
(9) Comprehension IV	0.36
(13) Memory for Stories:	-0.35

II Verbal

Item Name	Loading
(13) Memory for Stories:	0.78
(12) Vocabulary	0.72
(10) Opposite Analogies III	0.65
(2) Mutilated Pictures	0.42
(1) Differences	0.31

III Visualization & Visual Judgement

Item Name	Loading
(4) Opposite Analogies II	0.80
(3) Number Concepts	0.62
(6) Picture Absurdities I	0.41
(1) Differences	0.40

IV Control of Impulsivity

Item Name	Loading
(11) Repeating 5 Digits	0.77
(16) Naming Days Week	0.61
(8) Copying a Diamond	0.43
(3) Number Concepts	0.30

V Visual-Motor Integration

Item Name	Loading
(5) Maze Tracing	0.79
(8) Copying a Diamond	0.52
(1) Differences	-0.42

Table 5.7

Correlations Among Obliquely Rotated Principal Components - Younger Group

	I	II	III	IV	V
I	1.00				
II	0.44	1.00			
III	0.32	0.25	1.00		
IV	0.18	0.23	0.17	1.00	
V	0.04	-0.06	0.08	0.01	1.00

<u>Note</u>: Only those elements below the principal diagonal are shown.

intercorrelations provided further evidence of the existence of relatively independent dimensions in the data. The relationship between the various factors was considerably weaker than might have been expected. The oblique rotation did not alter the factors in any really significant way. This was not surprising given the relatively low factor intercorrelations.

Factor I (i.e. the Verbal, Nonverbal Reasoning factor) remained virtually intact upon oblique rotation. The only significant change was the negative loading of the Memory for Stories: The Wet Fall item on the oblique factor. In general, negative loadings are difficult to interpret and any attempt at interpretation should be considered tentative. The negative loading of this item on the Reasoning factor might possibly reflect its relatively rote nature. The examinee either recalls the details of the story or does not. No reasoning skills are presumably required.

The Verbal factor also remained relatively unchanged upon oblique rotation. The Similarities: Two Things item was eliminated. This would seem to provide further support for the hypothesis that in very young children verbal reasoning is the most influential component mediating performance on this task.

The Comprehension IV item and the Copying a Diamond item did not load significantly on the oblique Visualization and Visual Judgement factor. The loss of the Comprehension IV item was not surprising. Visualization might be employed by some children as an aid in performing this task but is unlikely to be the primary factor mediating performance. Reasoning skills are obviously more significant.

Visual-Motor Integration is the most pervasive factor involved in the Copying a Diamond Task. The role of Visual Judgement appears to be minimal in comparison.

As mentioned previously, the Opposite Analogies III item did not load significantly on the oblique Control of Impulsivity factor. This also was not surprising. While impulsivity might be expected to affect performance on this test to some degree, it is not likely to be the predominant influence. These tasks would appear to be considerably less susceptible to the effect of impulsivity than would the Repeating 5 Digits item, for example.

The Visual-Motor Integration factor remained unchanged after oblique rotation with the exception that the Differences item had a significant negative loading on the oblique factor. This negative loading could not be explained.

In summary, due to the small magnitude of the factor intercorrelations, the difference between the correlated and the uncorrelated factor solutions was quite negligible. Removing the orthogonality restriction did allow several of the factors to become slightly better defined. Verbal, Nonverbal Reasoning; Verbal; Visualization and Visual Judgement; Control of Impulsivity; and Visual-Motor Integration factors were clearly deliniated.

5.6 Item Intercorrelations

5.6.1 The Older Group

The intercorrelations for the 22 items in the older group analyses are shown in Table 5.8. Again, the lower off-diagonal items

Table 5.8

Item Inter-Correlations - Older Group

Item	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)	0.25						
(2)	0.13	0.47					
(3)	0.92	0.04	0.33				
(4)	0.17	0.33	0.24	0.28			
(5)	0.08	0.23	0.24	0.09	0.29		
(6)	0.08	0.38	0.26	0.10	0.23	0.32	
(7)	0.16	0.31	0.33	0.15	0.26	0.15	0.38
(8)	-0.11	0.12	0.05	0.04	0.08	0.18	0.04
(9)	0.12	0.32	0.26	0.33	0.33	0.03	0.16
(10)	0.36	0.45	0.39	0.31	0.38	0.22	0.45
(11)	0.29	0.27	0.06	0.06	0.24	0.23	0.06
(12)	0.26	0.50	0.33	0.28	0.24	0.35	0.45
(13)	0.11	0.23	0.19	0.19	0.18	0.25	0.20
(14)	0.08	0.33	0.35	0.29	0.26	0.09	0.45
(15)	0.11	0.22	0.23	0.27	0.03	0.10	0.16
(16)	0.15	0.29	0.32	0.23	0.11	0.17	0.24
(17)	0.10	0.28	0.18	0.08	0.17	0.20	0.21
(18)	0.06	0.29	0.19	0.22	0.16	0.20	0.40
(19)	0.15	0.43	0.28	0.20	0.25	0.27	0.30
(20)	0.12	0.33	0.24	0.22	0.21	0.22	0.30
(21)	0.11	0.27	0.19	0.07	0.17	0.25	0.13
(22)	0.22	0.49	0.39	0.22	0.28	0.22	0.41

Note: Squared Multiple Correlations have been placed in the principal diagonal. Variables are listed in Table 4.3

Table 5.8

Continued

Item	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1)							
(2)							
(3)							
(4)							
(5)							
(6)							
(7)							
(8)	0.19						
(9)	0.13	0.29					
(10)	0.11	0.31	0.54				
(11)	0.06	0.14	0.20	0.25			
(12)	0.13	0.17	0.49	0.25	0.58		
(13)	0.27	0.07	0.32	0.15	0.13	0.29	
(14)	0.07	0.23	0.38	0.12	0.51	0.00	0.55
(15)	0.03	0.11	0.25	0.07	0.38	0.07	0.24
(16)	-0.01	0.15	0.29	0.12	0.27	0.11	0.30
(17)	0.18	0.20	0.26	0.22	0.24	0.09	0.22
(18)	0.25	0.12	0.43	0.10	0.41	0.26	0.46
(19)	0.12	0.21	0.41	0.20	0.55	0.18	0.51
(20)	0.15	0.17	0.40	0.05	0.28	0.22	0.38
(21)	0.01	0.11	0.22	0.14	0.32	0.10	0.05
(22)	0.11	0.22	0.53	0.28	0.53	0.21	0.38

Table 5.8

Continued

Item	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
(1)								
(2)								
(3)								
(4)								
(5)								
(6)								
(7)								
(8)								
(9)								
(10)								
(11)								
(12)								
(13)								
(14)								
(15)	0.31							
(16)	0.17	0.34						
(17)	0.05	0.15	0.26					
(18)	0.28	0.27	0.24	0.49				
(19)	0.25	0.22	0.21	0.29	0.50			
(20)	0.07	0.31	0.40	0.42	0.35	0.39		
(21)	0.27	0.02	-0.02	-0.11	0.34	-0.01	0.33	
(22)	0.40	0.51	0.26	0.43	0.50	0.32	0.22	0.59

only are presented and the squared multiple correlations have been placed in the principal diagonal. The squared multiple correlations ranged from 0.19 to 0.59.

Four of the item intercorrelations were negative but these all failed to reach statistical significance for a two-tailed test. A number of the positive correlations also failed to reach statistical significance for either the one- or the two-tailed tests. Generally, the magnitude of the intercorrelations amongst the older group items approximated those of the younger group items. Again, the positive correlations do not suggest any singularities in the data. Although the highest positive correlation was 0.92, for the most part, the correlations were moderate suggesting the factor structure will be multidimensional. A one factor solution would be unlikely to account for the item intercorrelations.

5.7 The Unrotated Principal Components

5.7.1 The Older Group

The matrix of factor loadings for the unrotated principal components extracted from the older group data is shown in Table 5.9. Six component-factors having eigenvalues greater than 1.00 were retained.

The loadings on the first component-factor were all positive. The remaining five components were bi-polar. The first component accounted for 27.95 of the total variance in the set of variables and 48.37 percent of the common variance. The second component-factor explained 7.12 percent of the total variance and 12.32 percent of the common variance. Factor III accounted for 6.66 percent of the total

Table 5.9

Unrotated Principal Components - Older Group

	Component - Factor						
Item	I	II	III	IV	V	VI	h^2
(1)	0.32	0.31	-0.04	0.46	-0.36	0.39	0.70
(2)	0.69	0.14	0.11	-0.03	0.13	-0.02	0.53
(3)	0.56	0.06	-0.11	-0.01	0.28	-0.06	0.42
(4)	0.44	0.00	-0.22	0.13	0.45	0.42	0.64
(5)	0.45	0.08	0.27	0.30	0.18	-0.37	0.54
(6)	0.43	0.18	0.46	-0.28	-0.10	-0.04	0.52
(7)	0.59	-0.17	-0.17	-0.01	-0.11	-0.14	0.44
(8)	0.22	-0.30	0.52	-0.34	0.21	0.08	0.57
(9)	0.41	0.03	0.04	0.45	0.58	-0.09	0.73
(10)	0.74	0.01	0.02	0.16	0.00	0.11	0.59
(11)	0.34	0.33	0.36	0.32	-0.35	0.06	0.58
(12)	0.74	0.17	-0.13	-0.20	-0.17	-0.09	0.68
(13)	0.36	-0.07	0.49	-0.16	0.12	0.48	0.64
(14)	0.64	-0.24	-0.36	-0.01	-0.02	-0.30	0.68
(15)	0.42	0.24	-0.36	-0.35	0.07	0.25	0.55
(16)	0.49	-0.13	-0.25	0.10	-0.12	0.19	0.38
(17)	0.42	-0.29	0.28	0.23	-0.19	-0.20	0.47
(18)	0.59	-0.48	-0.05	-0.20	-0.16	0.15	0.67
(19)	0.67	0.15	-0.04	-0.18	-0.07	-0.26	0.57
(20)	0.56	-0.45	0.09	0.07	-0.08	-0.03	0.53
(21)	0.31	0.69	0.07	-0.27	0.13	-0.17	0.70
(22)	0.75	0.07	-0.13	-0.05	-0.16	0.06	0.61
Eigenvalue	6.15	1.57	1.47	1.25	1.17	1.12	

variance and 11.53 of the common variance in the data. The fourth component accounted for 5.69 percent and 9.85 percent of the total and common variance respectively. Factor IV explained 5.29 percent of the total variance in the set of variables and 9.16 percent of the common variance. Finally, the sixth factor accounted for 5.07 percent of the total variance and 8.78 percent of the common variance. The six factors combined explained 57.79 percent of the total variance in the data.

Variables having loadings of 0.30 or greater on a factor were again considered significant in the psychological interpretation of the factor. Loadings on the first unrotated factor were all significant with the exception of variable (8) which was the Memory for Designs I item. This item loaded 0.22 on the first factor. Again, the emergence of a general factor was anticipated. The comparatively low loading of the Memory for Designs I item on this component-factor would seem to be consistent with the failure of the Repeating 5 Digits item to load significantly on the equivalent factor in the younger group analysis. The Memory for Designs I item requires the examinee to reproduce from memory two figures to which he has been briefly exposed. The item would appear to require primarily visualization or visual memory ability rather than any complex manipulation of information.

No attempt was made to interpret the remaining factors prior to rotation.

5.8 Varimax Rotations of the Principal Components

5.8.1 The Older Group

Varimax rotations were made for two, three, four, five and six component-factors. The five solutions were evaluated for psychological interpretability. Items loading 0.30 or greater were considered significant in defining a factor.

Rotation of five of the original six principal components proved to be the most satisfactory solution in that the resultant dimensions were the most readily interpreted and the most consistent with those revealed in previous investigations. Rotation of six components produced three factors which were clearly interpretable. The three remaining factors were triplet factors (i.e. having only three significant loadings) which were impossible to interpret. Although it is generally acceptable practice to interpret factors having more than two significant loadings, the items loading on these triplet factors were such that a common underlying dimension could not be determined. It was assumed that, despite meeting the eigenvalue-one criterion, the sixth principal component was reflecting primarily error variance. The matrix of rotated factor loadings is shown in Table 5.10. The variables defining each factor are given in Table 5.11.

The first rotated factor was defined by 13 of the 22 variables. There were no significant negative loadings. This factor was identified as a general Verbal factor. The elements common to these items appeared to be word knowledge, comprehension of verbal information, and verbal fluency.

The Vocabulary and definition of Abstract Words items would seem to be straightforward measures of word knowledge.

The Word Naming task requires the examinee to produce orally as many words as possible in one minute. This test would seem to be a fairly direct measure of word knowledge and verbal fluency.

Table 5.10

Varimax Rotation of Principal Components - Older Group

	Component-Factor				
Item	I	II	III	IV	V
(1)	0.17	0.08	0.08	-0.22	0.67
(2)	0.36	0.36	0.36	0.30	0.22
(3)	0.35	0.30	0.44	0.11	0.00
(4)	0.26	0.17	0.59	-0.04	-0.10
(5)	0.12	0.03	0.46	0.30	0.34
(6)	0.11	0.35	-0.05	0.56	0.25
(7)	0.61	0.12	0.13	0.07	0.09
(8)	0.07	-0.04	0.05	0.72	-0.19
(9)	0.07	-0.03	0.83	0.06	0.11
(10)	0.52	0.20	0.36	0.18	0.32
(11)	0.04	0.10	0.02	0.17	0.73
(12)	0.58	0.51	0.07	0.13	0.22
(13)	0.10	0.08	0.14	0.60	0.09
(14)	0.72	0.13	0.23	-0.06	-0.03
(15)	0.33	0.59	0.07	-0.09	-0.14
(16)	0.55	0.07	0.14	-0.09	0.13
(17)	0.38	-0.25	0.10	0.32	0.33
(18)	0.74	-0.03	-0.02	0.29	-0.08
(19)	0.47	0.45	0.13	0.19	0.18
(20)	0.61	-0.18	0.15	0.31	0.09
(21)	-0.14	0.77	0.13	0.12	0.18
(22)	0.62	0.36	0.14	0.10	0.26
% Tot. Var.	18.15	9.53	8.70	8.47	7.87
% Com. Var.	34.43	18.08	16.50	16.06	14.93

Table 5.11

Items Having Significant Loadings After Varimax Rotation of Principal Components - Older Group

I Verbal		II Verbal, Nonverbal Reasoning	
Item Name	Loading	Item Name	Loading
(18) Verbal Absurdities IV	0.74	(21) Problem Situation II	0.77
(14) Abstract Words I	0.72	(15) Finding Reasons I	0.59
(22) Similarities: Three Things	0.62	(12) Vocabulary	0.51
(20) Memory for Sentences II	0.61	(19) Abstract Words II	0.45
(7) Verbal Absurdities II	0.61	(22) Similarities: Three Things	0.36
(12) Vocabulary	0.58	(6) Paper Cutting	0.35
(16) Word Naming	0.55		
(10) Making Change	0.52		
(19) Abstract Words II	0.47		
(17) Repeating 6 Digits	0.38		
(2) Verbal Absurdities I	0.36		
(3) Similarities & Differences	0.35		
(15) Finding Reasons I	0.33		

III Difficulty		IV Visualization	
(9) Rhymes: New Form	0.83	(8) Memory for Designs I	0.72
(4) Comprehension IV	0.59	(13) Block Counting	0.60
(3) Similarities & Differences	0.44	(6) Paper Cutting	0.56
(5) Naming Days Week	0.46	(17) Repeating 6 Digits	0.32
(10) Making Change	0.36	(20) Memory for Sentences II	0.31
(2) Verbal Absurdities I	0.36		

V Meaningful, Nonmeaningful Auditory Memory	
(11) Repeating 4 Digits Reversed	0.73
(1) Memory for Stories:	0.67
(5) Naming Days Week	0.34
(17) Repeating 6 Digits	0.33
(10) Making Change	0.32

The Verbal Absurdities and Finding Reasons I items require not only that the child comprehend complex verbal information but also that he be able to relate the ideas in a reasonably sophisticated manner. Unlike many of the items on the Stanford-Binet which require either very structured or very brief responses, the Verbal Absurdities and Finding Reasons I items tax the child's ability to express himself orally.

The Similarities: Three Things item requires the child to relate three items on the basis of some common characteristic. The task demands that the child be able to form concepts and to classify objects on the basis of these concepts. This type of conceptual thinking, as mentioned previously, is strongly associated with overall language functioning. The Similarities and Differences item is similar to the Similarities: Three Things item in that it also demands conceptual thinking.

The Memory for Sentences item requires the child to repeat verbatim sentences which have been dictated by the examiner. No omissions, substitutions or changes in word order are permitted. While there is certainly a strong memory component involved in this task, it would also be reasonable to expect that a child with good knowledge of word meanings and syntax would be at an advantage in the performance of this task.

The Making Change item requires the child to provide the solution to arithmetical word problems presented orally by the examiner. While this item loads significantly on several other factors, its presence on the Verbal factor was not surprising. Before the child can make the arithmetical computation required, she must understand and remember the relevant information. Children having good language skills would likely have some significant advantage in the performance of this task.

The one item which was difficult to explain in terms of its loading on the Verbal factor was the Repeating 6 Digits test. This task would appear to involve primarily rote memory skill and did, in fact, load on a clearly discernible Memory factor as well. The Repeating 5 Digits item, an almost identical task, did not load on the Verbal factor for the younger group.

The absence of the Rhymes: New Form item on the Verbal factor was somewhat surprising. This task requires the child to answer questions of the type: "Tell me a number which rhymes with three". One might expect that the child's word knowledge would be the most important factor in determining success on this task. It is possible, however, that since this task was easy for the majority of the children in the sample (i.e. 81 percent of the children passed it) exceptional verbal ability was not required for success by this group.

The Memory for Stories: The Wet Fall might also have been expected on this factor. This same item did load significantly on the Verbal factor for the younger group. As mentioned previously, however, the content of the story is quite simple and it is unlikely that many of the older children would fail to comprehend it. Also, the answers required are limited to a few words. Hence, the factor most likely to be mediating performance in this task is the child's ability to remember the relevant details of the story. For older children, then, memory would appear to be the most important dimension involved in the task. The item did, in fact, load on a clearly defined Memory factor.

The second rotated component was defined by 6 of the 22 items. There were no significant negative loadings on the factor. On the basis of the content of the defining items, this factor was classified as a Verbal, Nonverbal Reasoning factor.

The Problem Situation II item requires the child to answer a question pertaining to a brief story read by the examiner. The required information is not directly given in the story and must be deduced by the child. This task obviously requires some significant degree of reasoning ability.

The Finding Reasons I item requires the child to see the connection between cause and effect in familiar situations. The examinee is required to respond to questions of the nature "Give two reasons why children should not be noisy in school". Children having good verbal reasoning skills should be at an advantage in the performance of this task.

The reason for the presence of Vocabulary and Abstract Words II items on the Reasoning factor is perhaps less obvious. It would seem reasonable that children having well developed vocabularies, good verbal expressive ability and the capacity to think in abstract terms (i.e. those children having good overall verbal ability) would also tend to have good verbal reasoning skills. Given this argument, however, the absence of the Abstract Words I item on this factor cannot be explained.

The Similarities: Three Things item is comparable with the Similarities: Two Things item which was included in the younger group analyses. It would also be expected to load on the Reasoning factor as it does. The Absence of the Similarities and Differences item on this factor may thus seem difficult to explain. Finding both similarities and differences between two items would not, however, pose nearly as much difficulty for the older children as for the younger. It was, in fact, quite easy for this group (i.e. 79 percent of the older children passed the item). Thus, it is likely that the Similarities and Differences item becomes sufficiently easy for older children that exceptional reasoning skill may no longer be required for success on this task.

For the Paper Cutting test the examiner cuts a notch in a square of paper which has been folded once. The child must draw what the paper would look like if it were unfolded. This task is repeated with the paper folded twice. Nonverbal reasoning skill would appear to be important here. According to Sattler (1965) nonverbal reasoning involves the perception of logical relations inherent in nonverbal information, discrimination ability and analysis and synthesis. The understanding of spatial relations is also frequently involved. The Paper Cutting test would appear to require all these skills.

The absence of the Verbal Absurdities items on the Reasoning factor was somewhat surprising. Verbal Absurdities I did load significantly on the Reasoning factor in the younger group analyses. After Promax rotation the Verbal Absurdities I item did also load on this Reasoning factor as well. The continued absence of Verbal Absurdities II and IV even after oblique rotation could not be explained.

The absence of the Comprehension IV item on this Reasoning factor might also seem inconsistent given its presence on the Reasoning factor in the younger group. This item, however, appeared to be quite easy for the majority of the older children. It was passed by 76 percent of the older sample. The Comprehension IV task, then, is likely sufficiently easy for older children that exceptional reasoning skill may not be an advantage in the performance of the task.

The third rotated component was defined by 6 of the 22 items. There were no significant negative loadings. This factor proved to be somewhat difficult to interpret. What items loading on this factor had in common was not immediately obvious. The Rhymes: New Form item had a high loading on this factor. Comprehension IV, Similarities and Differences and Naming Days of the Week items had moderate loadings and Making Change and Verbal Absurdities I items had relatively low loadings on the factor. Apart from the fact that all of these items appeared to involve some verbal component, a more specific dimension underlying these items could not be immediately identified. Upon further examination, however, it was noteable that the first four items were all passed by more than 76 percent of the children in the sample. That is, these items all had quite low variances associated with them (see Table 4.4). With the exception of Memory for Stories: The Wet Fall, no other items had such a high percentage of ~~children~~ passing them. This factor was thus tentatively labelled a Difficulty factor. (It would more appropriately be called an Easiness factor). This interpretation did pose some problem in that the Making Change and Verbal Absurdities I items on this factor could not be explained since these items had p values close to 0.50. After Promax rotation of the components these two items were eliminated from the factor, thus providing some further support for the hypothesis that the remaining items were grouped together on the basis of their difficulty level.

The fourth component was defined by 6 items. There were no significant negative loadings on this factor. This factor was identified as a Visualization factor. That is, the tasks loading on this factor appeared to require that the child be able to retain or form a mental image of information which is not immediately available to perception.

In the Memory for Designs I test the examinee must reproduce from memory two designs which have been simultaneously exposed for a very brief time. While visual-motor coordination may be involved in this task to some degree, a critical factor mediating performance would be the child's capacity to form and retain a visual image of the designs long enough to reproduce them.

In the Block Counting task the child is shown a picture of piles of cubes arranged in two rows and is required to count the number in each pile. All of the blocks are not visible and the child must realize that there are more blocks than can be seen. It would seem obvious that good visualization skills would aid the child in appreciating the multidimensional nature of the drawing.

To succeed on the Paper Cutting task the examinee must be able to visualize what the shape of the cut in the paper will be once the paper is unfolded. It was not surprising that the item would load substantially on both the Reasoning factor and the Visualization factor.

The ability to comprehend cause-effect relationships and the ability to form mental images would both appear to be important skills in mediating performance on this task.

The presence of the Repeating 6 Digits and the Memory for Sentences variables on this factor was somewhat less predictable. It is possible that some children may approach the Repeating Digits test by producing a mental image of the numbers to be recalled. A visual image of the events depicted in the sentences of the Memory for Sentences test might possibly aid in the recall for some children. Relative to the first three items, however, these two variables had considerably lower loadings on the factor and, in fact, were eliminated from the factor when the orthogonality restriction was removed from the factor solution (i.e. after Promax rotation). This would indicate that the influence of visualization skills on the performance of these tasks was, if present at all, not pervasive.

The fifth and final component was defined by 5 of the 22 variables. There were no significant negative loadings on the factor. The factor was classified as a Meaningful, Nonmeaningful Auditory Memory factor.

The memory component in the Repeating Digits, Memory for Stories and Making Change items is obvious.

The Naming the Days of the Week item requires the child to produce information to which he has presumably been exposed prior to testing. Long-term memory skills might be expected to have some influence in mediating performance on this task.

The absence of the Memory for Sentences II item on factor V could not be explained since this task would also appear to demand the immediate recall of meaningful verbal information.

The Memory for Designs I item requires the recall of visual information so it was not surprising that it did not appear on the fifth factor. This factor appeared to be exclusively concerned with information input through the auditory channel. It is well known that the abi[illegible] deal with visual information is not necessarily consis[illegible]he ability to process auditory input.

[illegible]y, the first five principal components were interpreted [illegible] rotation to the Varimax criterion. Factor I was labelled as a Verbal factor. Factor II was defined as a Verbal, Nonverbal Reasoning factor. Factor III proved to be more difficult to interpret but was tentatively defined as a Difficulty factor. Factor IV was defined as a Visualization factor and factor V was classified as a Meaningful, Nonmeaningful Auditory Memory factor. In general, the factor structure was relatively clearly defined. There were, however, several instances where the presence or absence of particular variables on the various factors was difficult to explain. Several of the most serious inconsistencies were resolved upon Promax rotation of the components. An evaluation of these results follows.

5.9 Promax Rotations of the Principal Components

5.9.1 The Older Group

The matrix of pattern loadings on the primary axes is presented in Table 5.12. The factors emerging from the older group data were moderately correlated and, in general, the magnitude of the loadings was similar to those of the orthogonal factors. For this reason, items having loadings of 0.30 or greater were retained on the factors. The variables defining each factor are given in Table 5.13.

Table 5.12

Promax Rotation of Principal Components - Older Group

Pattern on the Primaries

Item	I	II	III	IV	V
(1)	0.12	-0.03	-0.01	-0.40	0.77
(2)	0.21	0.30	0.24	0.22	0.10
(3)	0.24	0.25	0.38	0.05	-0.13
(4)	0.17	0.12	0.60	-0.08	-0.22
(5)	-0.06	-0.05	0.43	0.18	0.29
(6)	-0.05	0.34	-0.18	0.56	0.16
(7)	0.65	0.02	-0.01	-0.04	0.02
(8)	-0.01	-0.03	0.02	0.81	-0.33
(9)	-0.12	-0.10	0.91	0.01	0.02
(10)	0.43	0.07	0.24	0.04	0.24
(11)	-0.10	0.02	-0.08	0.05	0.79
(12)	0.52	0.44	-0.12	0.02	0.12
(13)	-0.04	0.06	0.08	0.63	-0.02
(14)	0.79	0.02	0.10	-0.18	-0.12
(15)	0.31	0.61	-0.03	-0.12	-0.25
(16)	0.60	-0.04	0.04	-0.21	0.09
(17)	0.38	-0.38	0.01	0.23	0.32
(18)	0.85	-0.15	-0.17	0.22	-0.19
(19)	0.39	0.39	-0.03	0.11	0.07
(20)	0.66	-0.32	0.04	0.22	0.01
(21)	-0.38	0.84	0.06	0.12	0.10
(22)	0.58	0.26	-0.03	-0.04	0.17

Table 5.13

Items Having Significant Loadings After Promax Rotation of Principal Components - Older Group

I Verbal		II Verbal, Nonverbal Reasoning	
Item Name	Loading	Item Name	Loading
(18) Verbal Absurdities IV	0.85	(21) Problem Situation II	0.84
(14) Abstract Words I	0.79	(15) Finding Reasons I	0.61
(20) Memory for Sentences II	0.66	(12) Vocabulary	0.44
(7) Verbal Absurdities II	0.65	(19) Abstract Words II	0.39
(16) Word Naming	0.60	(6) Paper Cutting	0.34
(22) Similarities: Three Things	0.58	(2) Verbal Absurdities I	0.30
(12) Vocabulary	0.52	(17) Repeating 6 Digits	-0.38
(10) Making Change	0.43	(20) Memory for Sentences II	-0.32
(19) Abstract Words II	0.39		
(17) Repeating 6 Digits	0.38		
(15) Finding Reasons I	0.31		
(21) Problem Situation II	-0.38		

III Difficulty		IV Visualization	
(9) Rhymes: New Form	0.91	(8) Memory for Designs I	0.81
(4) Comprehension IV	0.60	(13) Block Counting	0.63
(5) Naming Days Week	0.43	(6) Paper Cutting	0.56
(3) Similarities and Differences	0.38	(1) Memory for Stories:	-0.40

V Meaningful, Nonmeaningful Auditory Memory	
(11) Repeating 4 Digits Rev.	0.79
(1) Memory for Stories:	0.77
(17) Repeating 6 Digits	0.32
(8) Memory for Designs I	-0.33

The factor intercorrelations are presented in Table 5.14. The highest correlation occurred between factors I and III (i.e. between Verbal and Difficulty factors) and between factors I and II (i.e. between Verbal and Reasoning factors). These were 0.45 and 0.40 respectively. The lowest correlation was 0.15 and occurred between factors II and IV (i.e. between Reasoning and Visualization factors). The magnitude of the factor intercorrelations provided further evidence of the existence of relatively independent dimensions in the data. The relationship between the various factors was, in general, not as strong as might have been anticipated. Factors II and IV, for instance, were quite unrelated.

The oblique rotation did not alter the results significantly. Allowing the moderately correlated dimensions to become correlated did allow for a somewhat better deliniation of several of the factors.

The Verbal factor remained quite intact upon oblique rotation. The only changes were that the Verbal Absurdities I and the Similarities and Differences items were eliminated and the Problem Situation II item had a significant negative loading on the factor. The Verbal Absurdities I item reappeared on the Reasoning factor. This result was not particularly surprising since this item would appear to have a very strong reasoning component associated with it. It would appear to be difficult to separate the verbal comprehension from the reasoning components in these tasks. It is likely that both are highly influential. The Similarities and Differences item loaded exclusively on the Difficulty factor after Promax rotation. Although the Problem Situation item had a very high loading on factor II (i.e. the Reasoning factor),

Table 5.14

Correlations Among Obliquely Rotated Principal Components - Older Group

	I	II	III	IV	V
I	1.00				
II	0.40	1.00			
III	0.45	0.29	1.00		
IV	0.37	0.15	0.24	1.00	
V	0.37	0.32	0.33	0.37	1.00

Note: Only those elements below the principal diagonal are shown.

its negative loading on the Verbal factor was surprising considering that the task necessarily involves some degree of verbal comprehension.

The Verbal, Nonverbal Reasoning factor became slightly better defined upon oblique rotation. The Similarities: Three Things item was eliminated from the factor. This undoubtedly reflects the fact that while reasoning may be an important component in mediating performance on this type of task for very young children, it generally would be expected to become less essential as children become more adept at conceptual thinking. Older children not passing this item might be expected to be weak in language skills in general.

The Verbal Absurdities I item had a low but significant loading on this factor after oblique rotation. As mentioned previously, this task would be expected to involve some reasoning component. The absence of the Verbal Absurdities II and IV items on the factor, however, could not be explained.

The Repeating 6 Digits and the Memory for Sentences II items no longer loaded on the Visualization factor after Promax rotation.

The Memory for Stories: The Wet Fall had a significant negative loading on this factor after oblique rotation. No explanation for this was attempted.

The Naming the Days of the Week and the Making Change items were absent from the oblique Memory factor. While a memory component may well be involved, it is quite likely that memory is not the predominant factor involved in the performance of these tasks. Naming the Days of the Week, for example, would likely be an overlearned task for older children. This is reflected in the fact that 83 percent of the children in the sample passed this task and in its presence on the Difficulty factor.

The absence of the Memory for Sentences II item on the Memory factor continued to be difficult to explain.

The Memory for Designs I item had a negative loading on factor V after oblique rotation. This may be further evidence for there being clear differences in the processing of visual versus auditory information.

In summary, the oblique rotation of the moderately correlated component factors did allow for a slightly clearer deliniation of the underlying dimensions in the data. Verbal; Verbal, Nonverbal Reasoning; Difficulty; Visualization; and Meaningful, Nonmeaningful Memory factors were defined.

5.10 Summary of the Results of the Exploratory Analyses

The results of the factor analyses of both the younger group and the older group data confirmed that the Standford-Binet is factorially complex within the mental age range VI through XI. While the presence of a common dimension or general factor linking all of the items was evident, the magnitude of the interitem correlations and the communality estimates (i.e. the squared multiple correlations) indicated that, in both groups, a single factor was unlikely to be sufficient to explain the data variance. Analyses of years VI through VIII revealed five distinct group factors. These were identified as Verbal, Nonverbal Reasoning; Verbal; Visualization and Visual Judgement; Control of Impulsivity; and Visual-Motor Integration factors.

Analyses of years VIII through XI revealed five interpretable group factors. These were identified as Verbal; Verbal, Nonverbal Reasoning; Difficulty; Visualization; and Meaningful, Nonmeaningful Auditory Memory factors. The Difficulty factor, however, was likely a result of the particular sample employed and may not be a reliable dimension. In general, the oblique factor solutions did not differ from the orthogonal solutions. In some cases the oblique rotations did allow for a slightly better deliniation of the dimensions. The finding of somewhat different factors in the two analyses was consistent with previous findings that the factor structure is not uniform across the entire test.

In general, the factors revealed in this study were consistent with those found in previous investigations (see Table 2.1). The fact that there have been no other published investigations of the 1960 revision of the Stanford-Binet across the age range VI through XI prevents an item by item comparison of the factors. Ramsey and Vane (1970), however, examined the range IV-6 through VI and found several factors which were very similar to the factors revealed in the present investigation of years VI through VIII. Their Verbal factor was very comparable to the one revealed here with a number of overlapping items. The Visual-Motor Integration factor found in their study was also very consistent with the Visual-Motor Integration factor revealed in the current analyses. Both studies also revealed similar Control of Impulsivity factors. The Visual Judgement and Visualization factor revealed in this study appeared to be somewhat of a composite of Ramsey and Vane's Visual Ability and Judgement, and Visual Imagery factors. Ramsey and Vane did not identify any Reasoning factor but rather, found a General Knowledge factor.

Jones (1949) analyzed years VIII through X of an earlier revision of the Binet. The composition of his factors was comparable to those revealed in the current analyses of years VIII through XI. Jones also found Verbal, Reasoning, and Memory factors. What Jones interpreted to be a Spatial factor was actually similar to what has been defined as a Visualization factor here. The items loading on the Visualization factor (i.e. Memory for Designs I, Block Counting, and Paper Cutting) likely do, in fact, also tap comprehension of spatial relations to a considerable degree.

Within the limits of the current investigation, some comparisons can be made between the two groups. Both analyses included the items from year VIII as well as a number of other items which were not identical but were similar (e.g. Similarities: Two Things, Similarities: Three Things, Repeating 5 Digits, Repeating 6 Digits). These common variables were useful in guiding the comparison.

In the younger group there was no distinct Memory factor. Perusal of the variables involved in the analyses, however, reveals that only the Memory for Stories, The Wet Fall and the Repeating 5 Digits items would be expected to have a strong underlying memory component. The Memory for Stories item loaded on the Verbal factor in the younger group analyses. This was not particularly surprising. Although memory skills are likely involved (results from the older group analyses would appear to confirm this), verbal comprehension ability is also apt to be an extremely potent factor in determining success on this task in very young children. The Repeating 5 Digits item loaded on the Control of Impulsivity factor for the younger children. This was consistent with other findings that for very young children, the ability to refrain from responding impulsively is often not well developed. So again, while rote memory skills are undoubtedly involved, attention and self-control are also vital. The fact that the Control of Impulsivity factor did emerge in the younger group analyses was not at all unexpected.

The analyses of the older group variables did not reveal a Visual-Motor Integration factor. With the exception of the Memory for Designs I item, however, there were no other items at years VIII through XI where Visual-motor skill would be expected to be involved to any great extent. Sattler (1965) included the Paper Cutting task in his Visual-Motor Integration category. The child does have to draw the unfolded paper but for the majority of the children the actual mechanics of the reproduction should be quite easy. The real difficulty would appear to be in the visualization of what the paper should look like.

While the Reasoning and Verbal factors were quite comparable across the two groups, some interesting trends did emerge from the comparison. Vocabulary items were consistent in loading on the Verbal factor in both groups as expected. In the older group, however, Vocabulary and Abstract Words items also loaded on the Reasoning factor possibly demonstrating the increasing association of language and reasoning abilities with age. The similarities type items (i.e. Differences, Similarities: Two Things, Similarities and Differences, Similarities: Three Things) are thought to tap conceptual thinking skills and were generally consistent in their loadings on the Verbal factor in both groups. In the younger group the Similarities and Differences and the Similarities: Two Things items also loaded strongly on the Reasoning factor. The Differences item, however, did not. In the older group the Similarities: Three Things item loaded on the orthogonal Reasoning factor but not on the oblique Reasoning factor. The Similarities and Differences item did not load on either the orthogonal or oblique Reasoning factor (Differences, and Similarities: Two things items were not included in the older group analyses). The important variable, here, appeared to be the difficulty of the task. For the younger children the Differences task appeared to be quite easy. The degree of conceptualization skill required for the task appeared to be well developed in the majority of the children in this age group. Exceptional reasoning skills did not appear to be needed. Finding similarities between two items or finding both similarities and differences, however, proved to be more difficult. These tasks did load on the Reasoning factor. For the older children the Similarities and Differences item did not load on the Reasoning factor. The Similarities: Three Things item was obviously more difficult and did load weakly on the Reasoning factor.

The Comprehension IV item also loaded significantly on the Reasoning factor in the younger group but on the Difficulty factor in the older group. These differences between the groups are interesting in that they might provide some weak evidence that the factor structure is somewhat different at different age levels not only as a result of differing items but also as a result of developmental trends in the establishment of intellectual skills.

The Visualization and the Visualization and Visual Judgement factors had no items in common so content could not be compared at this level. In general, however, the two factors differed somewhat in that several items on the Visualization and Visual Judgement factor require some judgement to be made about visual information which is available to perception. The items loading on the Visualization factor are limited to those requiring the formation of a mental image of objects or events not currently able to be perceived.

In summary, exploratory factor analytic procedures revealed five interpretable factors across mental ages VI through VIII of the Binet. Five factors were also interpreted across mental ages VIII through XI. Reasoning, Verbal and Visualization factors were comparable within the two analyses. Control of Impulsivity and Visual-Motor Integration factors were unique to the younger age levels. Memory and Difficulty factors were unique to the older age levels. The factors were, within the limits of possible comparison, quite consistent with those found in previous investigations of the various revisions of the test. Due to the differing number and nature of variables employed in each of the two analyses, it was not possible to confirm any developmental trends in the changing factor structure across the age levels of the test. Examination of common and similar items included in both analyses did, however, provide some suggestion of age-related changes in the processes mediating performance on some tasks.

Upon definition of the empirical factor structure, the fit of the empirical factors to the conceptual factor structure outlined by Sattler was evaluated. A summary of these results follows.

5.11 Oblique and Orthogonal Procrustes Rotations of the Principal Components

5.11.1 The Younger Group

The matrix of factor loadings for the orthogonal Procrustes rotation of the principal components to the target matrix developed from the younger group items is presented in Table 5.15. Although only the first five principal components had eigenvalues greater than one and have been interpreted to this point, the target factor structure defined six factors. This necessitated the extraction and rotation of six principal components. Items having significant loadings (i.e. loadings of 1.00) in the target matrix are underlined. Those items not underlined had zero loadings in the target matrix (see Table 4.6). It is evident from a quick perusal of these results that the fit was far from perfect. On factors I, II, IV, and VI, for example, items that should have loaded positively on the factor actually had zero or negative loadings. Factor III was the only factor on which all the target items loaded significantly.

The matrix of factor loadings for the oblique Procrustes rotation is shown in Table 5.16. Removal of the orthogonality restriction allowed for a slightly better fit. All of the target items did have

Table 5.15

Orthogonal Procrustes Rotation of Principal Components - Younger Group

	Component - Factor					
Item	I	II	III	IV	V	VI
(1)	0.02	0.40	0.62	0.13	-0.25	-0.02
(2)	-0.38	0.15	0.19	-0.14	0.58	0.34
(3)	0.04	-0.06	0.38	0.50	0.37	-0.22
(4)	-0.14	0.22	0.34	0.67	-0.01	-0.12
(5)	-0.16	-0.09	-0.21	0.43	0.09	0.66
(6)	0.23	0.40	0.22	0.29	0.52	0.00
(7)	0.25	0.18	0.65	-0.05	0.18	0.16
(8)	0.19	-0.27	0.09	0.58	0.27	0.34
(9)	0.05	0.30	0.22	0.40	0.03	0.41
(10)	-0.07	-0.21	0.64	-0.14	-0.04	0.31
(11)	0.40	-0.54	0.30	0.33	-0.14	-0.04
(12)	-0.29	-0.01	0.51	0.14	0.03	0.46
(13)	-0.41	-0.19	0.44	0.21	-0.21	0.31
(14)	0.48	0.44	0.54	0.06	0.06	0.52
(15)	0.43	0.16	0.39	0.01	0.01	0.55
(16)	0.20	-0.43	0.44	-0.05	0.44	0.20

Note: Items having positive loadings in target matrix are underlined.

Table 5.16

Oblique Procrustes Rotation of Principal Components - Younger Group

	Pattern on the Primaries					
Item	I	II	III	IV	V	VI
(1)	1.09	0.78	1.23	-0.31	-0.57	-0.16
(2)	1.86	-2.26	-0.13	8.49	0.10	4.54
(3)	-0.18	0.75	0.23	4.32	0.14	4.89
(4)	0.67	1.15	0.54	1.58	-0.18	1.61
(5)	1.42	0.85	-0.45	1.95	1.21	0.98
(6)	-0.72	-2.27	0.26	8.49	0.03	6.92
(7)	0.59	0.05	0.95	3.78	-0.41	3.94
(8)	0.11	1.76	-0.17	2.83	[illegible]	4.11
(9)	1.01	0.20	0.45	3.08	[illegible]	2.30
(10)	2.09	2.19	0.74	-0.35	[illegible]	0.39
(11)	-0.64	3.93	0.28	-4.02	0.32	0.52
(12)	2.78	1.52	0.56	1.84	0.23	0.86
(13)	2.85	2.96	0.48	-2.16	0.24	-2.01
(14)	-0.40	-1.47	0.44	3.71	0.26	3.14
(15)	0.50	0.45	0.75	2.12	0.16	2.80
(16)	0.44	1.20	0.12	3.72	0.15	4.85

Note: Items having positive loadings in target matrix are underlined.

significant loadings on the appropriate factors. The factors, themselves, however, were not well defined. In addition to the targeted items a number of the other items also loaded significantly on the various factors. In some cases these items had higher loadings than did the target items themselves. Upon examination of the loadings the factors were found to be difficult to interpret. Factor II, for example, which was the Memory factor had the Differences, Opposite Analogies II and III, Vocabulary, Maze Tracing, and Copying a Diamond items having loadings of 0.78 or greater. Memory would not be expected to be a predominant component in the performance of these tasks. Similarly, factor VI, the Social Intelligence factor, had items such as Number Concepts and Copying a Diamond loading higher than one of the target items. It is difficult to associate these items with social intelligence. The only factor which appeared to fit well was factor V, the Visual-Motor Integration factor. Here, only the target variables loaded highly on the factor.

In general, then, the attempt to rotate the empirical factors obained from the younger group analyses to Sattler's conceptual factors was not successful. This result, however, must be viewed with considerable caution. According to Helmes and Jackson (1977) four hypothesized loadings should ideally be used to define a factor in a targeted rotation. In the current investigation this was not possible for most of the factors because of the restricted number of variables included in the analyses and the assignment of each variable exclusively to a single category. Although Sattler's ten original categories were collapsed into six, this ideal criterion could not be met. The target factor I, for example, was defined by a single variable. Target factors II, IV, V, and VI were defined by two variables. Only factor III was defined by five variables. This means that the target factor structure was poorly defined and that the correspondence between the empirical and conceptual factors was really impossible to evaluate empirically.

If, however, items loading on the obliquely rotated empirical factors are classified according to their assignment to the six target categories it would also appear that, with the exception of the Visual-Motor Integration factor, the correspondence between the empirical factors really is not particuarly good (see Table 5.17). Items from Sattler's Reasoning, Conceptual Thinking, and Social Intelligence categories defined the oblique Verbal, Nonverbal Reasoning factor. The oblique Verbal factor was represented by items from Sattler's Memory, Conceptual Thinking and Reasoning categories. The oblique Visualization and Visual Judgement factor was defined by items from Sattler's Conceptual Thinking, Reasoning, and Social Intelligence categories. The oblique Control of Impulsivity factor was defined by items from Memory, Reasoning, Visual-Motor Integration, and Social Intelligence categories. Finally, both the items loading on the Visual-Motor Integration factor were also assigned to the Visual-Motor category by Sattler.

These findings would appear to further indicate that, with the exception of the Visual-Motor Integration dimension, there is little correspondence between Sattler's conceptual model and the empirical factor structure between the years VI through VIII. In fact, however,

Table 5.17

Sattler's Assignment of Younger Group Items to Conceptual Categories

Item Name	Category
(1) Differences	Conceptual Thinking
(2) Mutilated Pictures	Nonverbal Reasoning
(3) Number Concepts	Numerical Reasoning
(4) Opposite Analogies II	Conceptual Thinking
(5) Maze Tracing	Visual-Motor Integration
(6) Picture Absurdities I	Social Intelligence
(7) Similarities: Two Things	Conceptual Thinking
(8) Copying a Diamond	Visual-Motor Integration
(9) Comprehension IV	Social Intelligence
(10) Opposite Analogies III	Conceptual Thinking
(11) Repeating 5 Digits	Nonmeaningful Memory
(12) Vocabulary	Language
(13) Memory for Stories: The Wet Fall	Meaningful Memory
(14) Verbal Absurdities	Verbal Reasoning
(15) Similarities and Differences	Conceptual Thinking
(16) Naming Days of the Week	Social Intelligence

the discrepancy may not be as serious as it first appears. Part of the problem emerges from the multivocal nature of many of the factor loadings. Sattler assigns items uniquely to categories but in reality many items load on multiple factors. The empirical Verbal and Reasoning factors, for example, tend to overlap to some degree. Items tapping memory for meaningful information also tend to depend upon verbal comprehension skills. Furthermore, Sattler admits that within the conceptual framework the Social Intelligence category probably, in reality, overlaps with the Verbal Reasoning cateogry. The Comprehension IV item, for example, which requires the child to respond to questions such as "What is the thing for you to do when you h[illegible] broken something that belongs to someone else?" might be expected to tap both reasoning skills and social intelligence says Sattler. The fact that this item loaded on the Reasoning factor in the empirical analysis and is assigned to the Social Intelligence category by Sattler, then, may not be truly contradictory. The empirical analyses did not reveal a Social Intelligence factor although three items from this category were included. This might indicate that while these items may tap this dimension, it is not the most significant factor mediating performance.

Sattler also admits that Conceptual Thinking is highly related to overall language functioning. In fact, all the items which Sattler classified as tapping conceptual thinking loaded on the empirical Verbal factor. These two dimensions were not distinct in the empirical analyses. A number of the items from Sattler's Conceptual Thinking category also loaded on the empirical Reasoning factor further illustrating the overlap between verbal and reasoning processes.

Where Sattler's model and the empirical factor structure appear to really differ is in terms of the Visualization and Visual Judgement factor and the control of Impulsivity factor. Sattler does not identify these dimensions in his model. All of the items loading on the Visualization and Visual Judgement factor also, however, loaded significantly on one or more other factors. Sattler's assignment of these items to conceptual categories is often fairly consistent with these other loadings indicating some agreement on other relevant dimensions involved in these tasks. The agreement is less clear in the case of the Control of Impulsivity factor.

In summary, the attempt to rotate the empirical factors to fit Sattler's conceptual dimensions failed to confirm a significant correspondence between the empirical and conceptual factor structure. These results were difficult to evaluate because the target factors were inadequately defined. An examination of the variables defining each of the oblique empirical factors indicated that, with the exception of the Visual-Motor Integration factor, multiple conceputal categories were represented. This would tend to provide further evidence of a lack of correspondence. A more careful comparison, revealed that if the multivocal nature of the empirical factor loadings is taken into account, the lack of correspondence is less startling. The major problem appears to occur not so much in the definition of the major dimensions mediating performance but in the identification of the 'most important' dimensions. This will become even more apparent upon presentation of the results of the Clinician Judgement Study.

Sattler did not identify any dimensions comparable to the Visualization and Visual Judgement factor or the Control of Impulsivity factor.

5.11.2 The Older Group

The matrix of factor loadings for the orthogonal Procrustes rotation of the principal components to the target matrix developed from the older group items is presented in Table 5.18. Items having significant loadings in the target matrix are underlined. The greater number of items included in the older group analyses allowed the target factor structure to be somewhat better defined. Only target factors III and V were defined by less than three variables. It is immediately obvious from these results that once again the fit was poor. Many of the items which should have loaded positively on the various factors actually had zero or negative loadings. There was no factor on which all the target items loaded significantly.

The matrix of factor loadings for the oblique Procrustes solution is shown in Table 5.19. Removal of the orthogonality restriction allowed for a somewhat better fit of the empirical factors to the target factors. With the exception of the Memory for Designs I item (i.e. variable (8)) on the Memory factor (i.e. factor II), all of the target items did have significant loadings on the appropriate factors. There were, however, other items having significant loadings on the factors as well. In several cases these were larger than the target loadings. Factor I, the Language factor, also tended to include items from Sattler's Reasoning, Conceptual Thinking, and Social Intelligence

Table 5.18

Orthogonal Procrustes Rotation of Principal Components - Older Group

	Component - Factor					
Item	I	II	III	IV	V	VI
(1)	0.16	-0.20	0.71	0.18	0.30	-0.09
(2)	0.42	0.27	0.14	0.51	-0.06	-0.04
(3)	0.25	0.40	-0.02	0.43	0.11	-0.05
(4)	0.43	0.38	0.03	0.12	0.51	-0.17
(5)	0.13	-0.09	-0.1[illegible]	0.69	-0.05	-0.17
(6)	0.45	0.03	0.14	0.26	-0.48	0.01
(7)	0.13	0.26	0.11	0.46	0.[illegible]7	0.35
(8)	0.59	-0.04	-0.37	0.01	-0.27	0.14
(9)	0.23	0.09	-0.25	0.55	0.40	-0.38
(10)	0.42	0.17	0.26	0.52	0.18	0.10
(11)	0.22	-0.34	0.49	0.37	-0.13	-0.12
(12)	0.23	0.43	0.36	0.49	-0.16	0.21
(13)	0.79	-0.06	0.42	-0.04	-0.05	-0.01
(14)	-0.02	0.42	0.00	0.57	0.15	0.40
(15)	0.21	0.64	0.31	-0.01	0.02	0.02
(16)	0.20	0.21	0.27	0.24	0.30	0.28
(17)	0.21	-0.27	-0.03	0.51	-0.01	0.30
(18)	0.42	0.21	0.03	0.22	0.11	0.62
(19)	0.16	0.37	0.18	0.54	-0.23	0.14
(20)	0.34	0.00	-0.07	0.42	0.14	0.46
(21)	0.08	0.40	0.27	0.25	-0.45	-0.44
(22)	0.31	0.34	0.37	0.45	0.03	0.24

Note: Items having positive loadings in target matrix are underlined.

Table 5.19

Oblique Procrustes Rotation of Principal Components - Older Group

	Pattern on the Primaries					
Item	I	II	III	IV	V	VI
(1)	-0.10	0.65	0.49	0.42	-0.11	-0.01
(2)	1.05	0.17	2.14	0.88	0.27	0.56
(3)	1.41	-0.16	2.40	0.67	-0.02	0.61
(4)	0.73	-0.16	2.36	1.26	-0.34	0.63
(5)	0.17	0.32	0.83	-0.43	0.16	0.66
(6)	-0.22	0.38	0.48	0.75	0.75	0.22
(7)	1.46	0.12	1.79	0.62	-0.03	-0.08
(8)	-0.72	0.28	-0.34	1.02	0.56	0.04
(9)	1.36	0.01	1.67	-0.08	-0.29	1.04
(10)	1.10	0.38	2.06	1.02	0.05	[illegible]
(11)	-0.23	0.80	-0.19	0.01	0.3[illegible]	0.13
(12)	1.43	0.05	2.67	0.95	0.26	0.17
(13)	-1.07	0.52	0.19	1.62	0.49	0.16
(14)	2.23	-0.13	2.44	0.42	-0.23	0.00
(15)	0.76	-0.45	2.68	1.30	0.02	0.27
(16)	0.92	0.18	1.64	0.94	-0.22	-0.14
(17)	0.56	0.69	-0.15	0.10	0.16	-0.21
(18)	0.66	0.30	1.27	1.53	0.06	-0.51
(19)	1.50	0.00	2.25	0.54	0.30	0.30
(20)	0.84	0.48	0.80	0.85	0.02	-0.28
(21)	0.61	-0.34	1.89	0.09	0.50	0.90
(22)	1.26	0.20	2.40	1.13	0.12	0.10

Note: Items having positive loadings in target matrix are underlined.

categories. Factor II, the Memory factor, was, however, well defined. With the exception of the Block Counting item, no other items besides the target items had loadings of greater than 0.38. Factor III, the Conceptual Thinking factor, tended to include items from Sattler's Reasoning, Social Intelligence and Language categories in addition to the target variables. The Reasoning factor, factor IV, included items from the Social Intelligence, Language, and Conceptual Thinking categories in addition to the target variables. In addition to the target item (i.e. Paper Cutting), the Memory for Designs I, the Block Counting and the Problem Situation I items also had loadings of greater than 0.40 on factor V, the Visual-Motor Integration factor. It is difficult to identify any obvious visual-motor component in these tasks. Factor VI, the Social Intelligence factor, also included items from Sattler's Reasoning, Conceptual Thinking and language categories in addition to the target items. With the exception of the Memory factor, then, the correspondence between the empirical and target factors was certainly not perfect despite a better definition of the target factor structure.

When the items loading on the oblique empirical factors were classified according to their assignment to the six conceptual categories it was again found that multiple categories were represented on each category (See Table 5.20). Only the oblique Meaningful, Auditory Memory factor was consistent in being represented by items from Sattler's Memory category exclusively.

In summary, the situation at years VIII through XI was much the same as a years VI through VIII. Statistically Sattler's conceptual factor did not conform well to the empirical structure. No distinct

Table 5.20

Sattler's Assignment of Older Group Items to Conceptual Categories

Item Name	Category
(1) Memory for Stories: The Wet Fall	Meaningful Memory
(2) Verbal Absurdities I	Verbal Reasoning
(3) Similarities and Differences	Conceptual Thinking
(4) Comprehension IV	Social Intelligence
(5) Naming the Days of the Week	Social Intelligence
(6) Paper Cutting	Visual-Motor Integration
(7) Verbal Absurdities II	Verbal Reasoning
(8) Memory for Designs I	Visual Memory
(9) Rhymes: New Form	Language
(10) Making Change	Numerical Reasoning
(11) Repeating 4 Digits Reversed	Nonmeaningful Memory
(12) Vocabulary	Language
(13) Block Counting	Numerical Reasoning
(14) Abstract Words I	Language
(15) Finding Reasons I	Verbal Reasoning
(16) Word Naming	Language
(17) Repeating 6 Digits	Nonmeaningful Memory
(18) Verbal Absurdities IV	Verbal Reasoning
(19) Abstract Words II	Language
(20) Memory for Sentences II	Meaningful Memory
(21) Problem Situation II	Social Intelligence
(22) Similarities: Three Things	Conceptual Thinking

Conceptual Thinking factor could be distinguished from Verbal and Reasoning factors in the data. Independent Social Intelligence or Visual-Motor Integration factors could also not be derived from the data. With the exception of the Memory dimension, the attempt to fit the data to the conceptual dimensions resulted in complex factors which were difficult to interpret.

Again, however, a more careful comparison revealed that if the multidimensional nature of the variables is taken into account, some agreement does exist in the assignment of many items to relevent dimensions.

5.12 Summary of the Results of the Confirmatory Analyses

In general, the attempt to statistically conform the data to Sattler's conceptual factor structure was unsuccessful. This, however, does not mean that the model is of no value conceptually and diagnostically. Sufficient consistency does exist in the identification of common dimensions across the mental age range VI through XI to be of some significant value diagnostically. Sattler's Language and Conceptual Thinking categories combined collate reasonably well with the Verbal factor in both groups. A number of these variables also loaded on the Reasoning factor as well, however. The majority of the items that have been identified by Sattler as tapping Meaningful and Nonmeaningful Memory also emerged on the memory factor in the older group analyses. In the younger group, however, another important dimension mediating performance on several of these 'memory' tasks was Control of Impulsivity. Sattler did not identify this dimension in his model. No distinct Visual Memory factor emerged in the empirical analyses. There was only one item included in the analyses (i.e. Memory for Designs I) that might be expected to tap this dimension. This item loaded on the Visualization factor. This would not appear to be a serious contradiction since Visual Memory and Visualization would appear to be difficult to distinguish, anyway. The identification of items tapping Visual-Motor Integration did not pose any problem. There was perfect agreement between Sattler and the empirical factor with the exception that Sattler identified the Paper Cutting item as tapping primarily Visual-Motor Integration whereas it loaded exclusively on the empirical reasoning factor. No factor which appeared to be tapping exclusively Social Intelligence emerged in the empirical analyses. Sattler's Reasoning category was probably the most inconsistent with the empirical equivalents. This, in some considerable part, reflects the difficulty in separating specific reasoning skills from general verbal ability. Sattler did not identify Visualization, Visual Judgement, or Control of Impulsivity dimensions.

It would appear, then, that while Sattler's model does have some consistency with the empirical factor structure across the mental years VI through XI, it also has some inconsistencies as well. While possibly useful in guiding the diagnostic process, these limitations must be kept in mind. The assignment of items exclusively to categories is a major drawback as considerable diagnostic information may be lost if the model were to be applied rigorously. The most important dimension diagnostically for a particular task might depend not only on the age of the child but on his overall pattern of intellectual strengths and weaknesses.

Before the value of Sattler's model can be fully evaluated, however, its application by clinicians must be examined. The results of the student clinicians ratings of the Binet items according to Sattler's conceptual categories follows.

PART B: THE CLINICIAN JUDGEMENT STUDY

5.13 Proportion Agreement of Clinicians With Sattler's Assignment of Binet Items to Conceptual Categories

The proportion of student clinicians in agreement with Sattler on the assignment of Stanford-Binet items to the ten conceputal categories is presented in Table 5.21. These proportions reflect the students' first choice selection only. Every clinician placed some or all of the items in multiple categories. The items have been grouped according to category to facilitate comparisons within and across dimensions. Repeated items provide some indication of the reliability of the clinicians' assignment of items to categories although it is possible that the clinicians felt the items to be tapping different skills at different ages in some cases. Horizontal lines border the variables in each category which were included in the factor analytic study. These items have been singled out for special consideration here. Mean proportions are presented for each cateogry. Repeated items have been included in the calculation of the means. These values should, therefore, be treated with caution. The repeated inclusion, for example, of the Vocabulary item which had a very high proportion of agreement associated with it will have inflated the mean somewhat for the Verbal factor. Over the entire test the percentage of students

Table 5.21

Proportion of Student Clinicians In Agreement With Sattler's Assignment of Stanford-Binet Items to Conceptual Categories

I. Language

Test Age	Item Name	Proportion
II	Identifying Parts of the Body	0.45
II	Picture Vocabulary	0.85
II	Word Combinations	0.90
II-6	Identifying Objects by Use	0.24
II-6	Identifying Parts of the Body	0.38
II-6	Naming Objects	0.74
II-6	Picture Vocabulary	0.85
III	Picture Vocabulary	0.81
IV	Picture Vocabulary	0.81
IV	Picture Identification	0.21
V	Definitions	0.91
VI	Vocabulary	0.93
VIII	Vocabulary	0.98
IX	Rhymes: New Form	0.57
X	Vocabulary	1.00
X	Abstract Words I	0.38
X	Word Naming	0.81
XI	Abstract Words II	0.69
XII	Vocabulary	1.00
XII	Abstract Words I	0.69
XII	Minkus Completion I	0.55
XII	Abstract Words II	0.67
XIII	Dissected Sentences	0.33
XIV	Vocabulary	1.00
AA	Vocabulary	1.00
AA	Differences Between Abstract Words	0.43
AA	Abstract Words III	0.64
SAI	Vocabulary	1.00
SAI	Minkus Completion II	0.55
SAI	Sentence Building	0.60
SAII	Vocabulary	1.00
SAIII	Vocabulary	1.00
		Mean = 0.72

Note: Horizontal lines border variables in each category included in the factor analytic study.

Table 5.21

Continued

II. Meaningful Memory

Test Age	Item Name	Proportion
IV	Naming Objects from Memory	0.36
IV-6	Three Commissions	0.67
VIII	Memory for Stories: The Wet Fall	0.93
XI	Memory for Sentences II	0.79
XIII	Memory for Sentences III	0.81
SAII	Repeating Thought of Passage I: Value of Life	0.64
SAIII	Repeating Thought of Passage II: Tests	0.67
	Mean =	0.70

III. Nonmeaningful Memory

Test Age	Item Name	Proportion
II-6	Repeating 2 Digits	0.91
VII	Repeating 5 Digits	0.93
IX	Repeating 4 Digits Reversed	0.88
X	Repeating 6 Digits	0.91
XII	Repeating 5 Digits Reversed	0.83
SAI	Repeating 6 Digits Reversed	0.88
	Mean =	0.89

IV. Visual Memory

Test Age	Item Name	Proportion
III	Picture Memories	0.74
IX	Memory for Designs I	0.64
XI	Memory for Designs I	0.64
XIII	Copying a Bead Chain from Memory	0.64
	Mean =	0.67

Table 5.21

Continued

V. Conceptual Thinking		
Test Age	Item Name	Proportion
IV	Opposite Analogies I	0.36
IV-6	Opposite Analogies I	0.36
VI	Differences	0.50
VI	Opposite Analogies II	0.36
VII	Similarities: Two Things	0.71
VII	Opposite Analogies III	0.33
VIII	Similarities and Differences	0.62
XI	Similarities: Three Things	0.74
XIV	Reconcilation of Opposites	0.60
AA	Proverbs I	0.38
AA	Essential Differences	0.62
SAI	Essential Similarities	0.57
SAII	Proverbs II	0.45
SAII	Essential Differences	0.67
SAIII	Proverbs III	0.38
SAIII	Opposite Analogies IV	0.48
		Mean = 0.50

VI. Verbal Reasoning		
VII	Verbal Absurdities I	0.60
IX	Verbal Absurdities II	0.62
X	Finding Reasons I	0.21
XI	Verbal Absurdities IV	0.57
XII	Verbal Absurdities II	0.52
XIII	Problems of Fact	0.52
XIV	Reasoning I	0.74
SAII	Finding Reasons III	0.33
SAIII	Reasoning II	0.29
		Mean = 0.54

Table 5.21

Continued

VII. Nonverbal Reasoning

Test Age	Item Name	Proportion
II	Delayed Response	0.10
III-6	Comparison of Balls	0.69
III-6	Patience: Pictures	0.64
III-6	Discrimination of Animal Pictures	0.52
III-6	Sorting Buttons	0.17
IV	Discrimination of Forms	0.71
IV-6	Pictorial Similarities and Differences I	0.60
V	Pictorial Similarities and Differences I	0.52
V	Patience: Rectangles	0.57
VI	Mutilated Pictures	0.43
XIII	Plan of Search	0.71
XIV	Orientation: Direction I	0.26
AA	Orientation: Direction II	0.26
SAIII	Orientation: Direction III	0.26
	Mean =	0.42

VIII. Numerical Reasoning

Test Age	Item Name	Proportion
VI	Number Concepts	0.88
IX	Making Change	0.98
X	Block Counting	0.55
XIV	Induction	0.17
XIV	Ingenuity I	0.57
AA	Ingenuity I	0.55
AA	Arithmetical Reasoning	0.99
SAI	Enclosed Box Problem	0.48
SAII	Ingenuity I	0.62
	Mean =	0.64

Table 5.21

Continued

IX. Visual-Motor Integration

Test Age	Item Name	Proportion
II	Three-Hole Form Board	0.74
II	Block Building	0.90
III	Stringing Beads	0.93
III	Block Building: Bridge	0.76
III	Copying a Circle	1.00
III	Drawing a Vertical Line	0.95
V	Picture Completion: Man	0.45
V	Paper Folding: Triangle	0.52
V	Copying a Square	0.93
VI	Maze Tracing	0.62
VII	Copying a Diamond	0.95
IX	Paper Cutting	0.29
		Mean = 0.75

X. Social Intelligence

Test Age	Item Name	Proportion
II-6	Obeying Simple Commands	0.17
III-6	Response to Pictures: Level I	0.29
III-6	Comprehension I	0.64
IV	Comprehension II	0.60
IV-6	Asthetic Comparison	0.81
IV-6	Materials	0.21
IV-6	Comprehension III	0.52
VII	Picture Absurdities I	0.50
VII	Comprehension IV	0.69
VIII	Comprehension IV	0.67
VIII	Naming Days of the Week	0.07
XI	Problem Situation II	0.38
XII	Picture Absurdities II	0.46
		Mean = 0.51

agreeing with Sattler's assignment of items to categories on the first choice ranged from 7 to 100 percent.

Within the Language category agreement proportions ranged from 0.33 to 1.00. The Vocabulary item was assigned to this category on the first choice by all of the clinicians except at years VI and VIII. Several of the clinicians felt this item to be tapping primarily conceptual thinking at the earlier age levels. This item loaded on the Verbal factor in the factor analyses of both the younger and older group variables. The majority of the clinicians placed the Abstract Words items in the Language category but not always on the first choice. Often, Conceptual Thinking was felt to be the primary dimension being tapped. A number of students also felt those items to be tapping Verbal Reasoning skills as well. This is interesting in light of the findings that the Verbal and Conceptual Thinking dimensions were not distinguishable in the factor analyses and in light of the fact that the Abstract Words II item loaded on the Reasoning factor in the factor analyses of the older group variables. The Rhymes: New Form item was felt by the majority of the clinicians to tap primarily Language ability. Conceptual Thinking, Meaningful Memory, and Verbal Reasoning categories, however, were selected on the first choice by some of the clinicians. The item loaded exclusively on the Difficulty factor in the factor analyses of the older group items. Wright (1939) also included a Rhymes item in her analysis of the Binet factor structure which, although not identical, was similar to this item, and did not find it to load significantly on any of her factors, including the Verbal factor. This inconsistency may imply that the underlying processes involved in this task are difficult to identify. In general, the majority of the items assigned by Sattler to the Language category were also placed in the Language category by the student clinicians on one of the three choices. The Verbal Reasoning and Conceptual Thinking categories, however, competed with the Language category for the first choice in many instances. The overlap between these categories is quite consistent with the overlap revealed in the factor analytic investigation.

Within the Meaningful Memory category the agreement proportions ranged from 0.36 to 0.93. The Naming Objects from Memory task which had the lowest proportion requires the child to identify which of a number of common objects the examiner has removed. A considerable number of the clinicians felt this item to reflect Visual Memory rather than Meaningful Memory. It is interesting to note the difference between the Memory for Stories: The Wet Fall and the Memory for Sentences II items in the proportion agreement. A considerable number of the students felt the Memory for Sentences item to be tapping primarily Language. In the factor analysis of the older group items the Memory for Sentences II item loaded on the Verbal factor but not on the Memory factor. The Memory for Stories item, on the other hand, loaded exclusively on the Memory factor.

Within the Nonmeaningful Memory category the agreement proportions ranged from 0.83 to 0.91. There appeared to be very little disagreement in the assignment of items to this category. The Repeating 4 Digits Reversed item loaded on the Memory factor in the factor analytic study of the older group items. The Repeating 6 Digits item loaded on both the Verbal and Memory factors in the same analyses. The Repeating 5 Digits item loaded on the control of Impulsivity factor in the younger group analyses. No distinct Memory factor was identified in the younger group analyses.

The proportions of students agreeing with Sattler's assignment of items to the Visual Memory category ranged from 0.64 to 0.74. Several of the clinicians placed the Memory for Designs I item in the Visual-Motor Integration category and a number in the Nonmeaningful Memory category on the first choice. This item loaded on the Visualization factor in the factor analyses.

Within the Conceptual Thinking category the agreement proportions ranged from 0.33 to 0.74. A number of the student clinicians felt the Differences item to be primarily tapping Verbal Reasoning. A number of others felt it to be primarily tapping Language ability. The vast majority placed the item in the Conceptual Thinking category on one of the three choices. An identical situation existed for the Similarities: Two Things, Similarities and Differences, and Similarities: Three Things items. The comparability of these results to the factor analytic results is startling. These items tended to load on both the Verbal and Reasoning factors. While a considerable number of student clinicians felt the Opposite Analogies II item to be tapping primarily Conceptual Thinking, the majority felt it to be tapping Verbal Reasoning. In the factor analytic investigation of the younger group items the Opposite Analogies II item loaded exclusively on the Visualization and Visual Judgement factor. The Opposite Analogies II item, however, loaded exclusively on the Verbal factor. In general, the vast majority of the Conceptual Thinking items were placed in the Conceptual Thinking category by the student clinicians on one of their three choices.

The proportions of students agreeing on the first choice with Sattler's assignment of items to the Verbal Reasoning category ranged from 0.21 to 0.74. The Verbal Absurdities items were placed in the Verbal Reasoning category on the first choice far more frequently than in any other category. A number of clinicians, however, felt these items to be tapping primarily Social Intelligence. Several felt these items to be tapping Language ability primarily and several felt Conceptual Thinking to be the most important domain. In the factor analyses of the younger group variables the Verbal Absurdities I item loaded exclusively on the Reasoning factor. In the factor analyses of the older group items it loaded on both the Verbal and Reasoning factors. Verbal Absurdities II and IV loaded exclusively on the Verbal factor. The Finding Reasons item were felt by the majority of the student clinicians to reflect primarily Social Intelligence. Verbal Reasoning was a frequent second choice. The Finding Reasons I item loaded very strongly on the Reasoning factor in the factor analyses of the older group variables.

Within the Nonverbal Reasoning category the agreement proportions ranged from 0.10 to 0.71. The Delayed Response item which had a very low agreement proportion associated with it requires the child to retain in memory the position of a hidden toy for a period of ten seconds. By far the majority of the students felt this item to be tapping Visual Memory rather than Nonverbal Reasoning. The Sorting Buttons task which also had a very low proportion of agreement requires the child to group black and white buttons by colour. The majority of the clinicians felt this task to be tapping primarily Visual-Motor Integration. A considerable number of others felt it to be tapping primarily Conceptual Thinking. The most popular category for the Mutilated Pictures item was Nonverbal Reasoning. A considerable number of the clinicians, however, felt the item to be tapping primarily Visual Memory. One clinician felt it to be tapping primarily Language ability. In the factor analytic study of the younger group items, this item loaded exclusively on the Verbal factor. The Orientation: Direction items which require the testee to answer questions or solve word problems pertaining to directionality were felt to tap primarily Verbal Reasoning skills by the majority of the clinicians. In general, the Nonverbal Reasoning category appeared to be the most problematic of all the categories.

Within the Numerical Reasoning category the agreement proportions ranged from 0.17 to 0.99. The Induction item which had a very low proportion of agreement associated with it requires the testee to figure out, for example, how to measure exactly two pints of water using five and three pint cans. The majority of the clinicians felt this task to be tapping primarily Nonverbal Reasoning rather than Numerical Reasoning. By far the majority of the clinicians felt the Number Concepts and Making Change items to be tapping primarily Numerical Reasoning. The Making Change item, however, loaded on the Verbal and Difficulty factors in the Factor analytic investigation of the older group variables. The Number Concepts item loaded exclusively on the Visualization factor. A significant number of the student clinicians felt the Block Counting item to be tapping Nonverbal Reasoning rather than Numerical Reasoning. In the factor analyses, however, this item loaded on the Visualization factor not on the Reasoning factor.

Proportion agreements within the Visual-Motor Integration category ranged from 0.29 to 1.00. The majority of the clinicians felt the Maze Tracing item to be reflecting primarily Visual-Motor Integration skill. A number of clinicians, however, felt it to be tapping primarily Nonverbal Reasoning. Virtually all of the clinicians placed this item in the Visual-Motor Integration category on one of the three choices. This item loaded exclusively on the Visual-Motor Integration factor in the factor analytic investigation of the younger group items. The Copying a Diamond item was identified as tapping primarily Visual-Motor Integration by Sattler, the student clinicians and by the factor analytic procedures. The majority of the student clinicians identified the Paper Cutting item as tapping primarily Nonverbal Reasoning rather than Visual-Motor Integration. A number, however, felt it to be tapping primarily Visual Memory. Interestingly, this item loaded on the Reasoning and Visualization factors in the factor analytic investigation.

Finally, within the Social Intelligence category, the agreement proportions ranged from 0.07 to 0.81. Half of the student clinicians felt the Picture Absurdities I item to tap primarily Social Intelligence. The remaining first choice ratings were spread quite evenly between Verbal Reasoning, Nonverbal Reasoning and Conceptual Thinking categories. This item loaded exclusively on the Reasoning factor in the factor analytic investigation of the younger group variables. A similar situation existed for the Comprehension IV item. The majority of the clinicians identified it as tapping primarily Social Intelligence. The remaining first choice ratings, however, were split quite evenly between Conceptual Thinking and Verbal Reasoning categories. This item Loaded on the Reasoning factor in the younger group analyses and on the Difficulty factor in the older group factor analyses. An equal number of clinicians identified the Problem Situation II item as tapping primarily Verbal Reasoning and Social Intelligence. Virtually all of the remaining students placed this item first in the Conceptual Thinking category. This item loaded exclusively on the Reasoning factor in the factor analytic investigation. Sattler's statement that the Social Intelligence and Verbal Reasoning categories overlap would seem to be supported here. By far the majority of the student clinicians placed the Naming the Days of the Week item in the Meaningful Memory category. This item had the lowest proportion of agreement associated with it of all the items on the test. In the factor analytic investigation of the older group variables this item loaded on the Memory factor. It loaded on the Control of Impulsivity factor in the younger group analyses.

In summary, the percentage of student clinicians agreeing with Sattler on first choice assignments of Stanford-Binet items was quite variable from item to item. The majority of the items were placed in multiple categories by the student clinicians. This was consistent with the factor analytic findings that many of the variables did indeed load on multiple factors. If all of the clinicians choices were taken into account the agreement with Sattler was, with a few noteable exceptions such as Naming the Days of the Week, Delayed Response, and Sorting Buttons, much improved. The overlap between the Language, Conceptual Thinking, Verbal Reasoning, and Social Intelligence categories was quite well reflected by the clinician judgements. Verbal Reasoning, Nonverbal Reasoning and Numerical Reasoning categories did not appear to be as distinct in the minds of the clinicians as might be expected. A number of items which Sattler classified as reflecting Nonverbal Reasoning were thought to tap Verbal Reasoning skills by many of the clinicians. The Memory category, in general, appeared to be quite well defined. When all three of the clinician judgements were taken into account there was excellent agreement in the assignment of items to this dimension. The Meaningful Memory, Nonmeaningful Memory, and Visual Memory categories, however, did not appear to be as distinct from each other as might be expected. The Visual-Motor category also appeared to be well defined. When all the choices were considered there was nearly perfect consensus between the student clinicians and Sattler on the assignment of items to this category.

Within the limits of possible comparison, the clinicians' judgements also showed some startling agreement with the factor analytic findings. Often when the selections of the clinicians were in disagreement with Sattler on the first choice they were supported by the factor analytic findings. In cases where the students choices did agree with Sattler, there was often further confirmation to be found in the factor analytic study. Noteable exceptions to this were the Number Concepts, Mutilated Pictures, Making Change and Block Counting items which were felt by the majority of the clinicians to reflect reasoning skills. This was not confirmed by the factor analytic investigation. At this point the reader must be reminded that the restricted nature of the sample severely limits the extent to which the results of the Clinician Judgement Study can be generalized.

5.14 Integration of the Results of the Factor Analytic and Clinician Judgement Studies

The exploratory factor analytic investigation revealed the Stanford-Binet to be factorially complex across mental ages VI through XI. An attempt to conform statistically the empirical factor structure to Sattler's conceptual model was, for the most part, unsuccessful. Item by item comparisons, however, revealed that if the multidimensional nature of the majority of the items was acknowledged, some consistency did exist in the type of dimensions being identified empirically. Verbal, Reasoning, Memory, and Visual-Motor Integration dimensions were identifiable in both the empirical analyses and in the conceptual model. Distinct Conceptual Thinking and Social Intelligence dimensions, however, were not confirmed empirically at these age levels. Visualization, Visual Judgement, and Control of Impulsivity factors were unique to the empirical analyses.

The Clinician ratings indicated that the students generally believed many of the Stanford-Binet items to be tapping multiple intellectual processes. The agreement with Sattler on the 'most important' processes was in general only moderate. If, however, the multiple dimensions were taken into account some consistency did exist between

Sattler and the student clinicians for many of the tasks. Within the mental age range VI through XI there was also some consistency with the factor analytic results in the identification of important dimensions being tapped by the majority of the items. The implications of these findings for the diagnostic use of the Stanford-Binet are discussed in the following chapter.

CHAPTER SIX

6. DISCUSSION

6.1 Restatement of Objectives and Review of Major Results

Objective one of the thesis was to examine the historical development of the Stanford-Binet and in so doing, to determine its potential to provide diagnostic information in addition to a single index of overall intellectual potential. The detailed review undertaken provided evidence that Alfred Binet had an understanding of processes mediating intellectual behavior that was far in advance of his time. Many of Binet's theories were remarkable in their similarity to currently accepted views. The original Binet test was a composite of tasks which Binet developed through extensive clinical testing to assess dimensions such as memory capacity, word knowledge, number sense and reasoning. The breadth of Binet's clinical experience has been rivalled by few to date. The subsequent revisions of the Binet have altered the fundamental structure of the test very little and there was every reason to suspect that it might be possible to isolate specific dimensions within the test which might be useful diagnostically.

The second objective of the thesis was to review previous attempts to isolate specific dimensions in the various revisions of the test. The review confirmed that the 1960 Stanford-Binet was indeed likely to be complex factorially although not factorially pure. The varying revisions of the test used, the varying segments of the test analyzed, the varying nature of the samples included, and the varying statistical methods employed within each study made specific comparisons between the investigations somewhat difficult. In general terms, however, there was consistency in the kinds of dimensions that were identified. Very little effort has been directed towards attempting to identify diagnostically useful dimensions in the 1960 revision of the scale and it was towards this end that the major objective of the thesis was directed.

Objective three was to conduct a factor analytic study of the 1960 revision of the Stanford-Binet over the mental age levels VI through VIII and VIII through XI employing two clinical samples. The fact that the Binet is an age scale prevents the analysis of more than a small portion of the scale within the context of a single analysis. Converging evidence from several sources suggested that the 1960 Stanford-Binet was indeed factorially complex over this age range. The evidence arose from an examination of the bivariate correlations, the squared multiple correlations and an interpretation of the multidimensional factor structures. In addition to a general factor, five interpretable group factors were revealed across mental ages VI through VIII. These were defined as Verbal, Nonverbal Reasoning; Verbal; Visualization and Visual Judgement; Control of Impulsivity; and Visual-Motor Integration dimensions. With the exception of the Reasoning factor, these dimensions appeared to be similar to those found in Ramsey and Vane's (1970) analysis of the 1960 revision across years IV-6 through VI.

Across mental ages VIII through XI five group factors were also interpreted. These were Verbal; Verbal Nonverbal Reasoning; Visualization; Meaningful, Nonmeaningful Auditory Memory; and Difficulty. The definition of the Difficulty factor was considered tentative. The items loading on this factor did not lend themselves to any other obvious interpretation. If these items were linked on the basis of their difficulty (really easiness) level this factor may not be expected to be a reliable dimension across different samples. Only a replication of the analyses employing a different sample will clarify this issue. The nature of the remaining dimensions, in general, appeared to be fairly consistent with those defined by other investigators over a similar age range in earlier revisions.

Results of the two analyses indicated that the factor structure of the Binet is not entirely consistent across the age levels. This was not unexpected and supports what has been found in previous factor analytic investigations and what has been observed by Sattler (1965). It is not possible to say whether the differences do, in fact, reflect true developmental changes in abilities or merely reflect the differences in the kinds of items included at each age level. The presence of the Control of Impulsivity factor at the younger age levels, however, is interesting in that one would expect such a factor to be a more potent force in mediating performance in very young children than in somewhat older children. A number of common items included in both the current analyses also provided some weak evidence that the most important dimension mediating performance on some tasks might vary somewhat with age.

In general, the factors revealed in both analyses were reasonably well defined although there were a few instances where the presence or absence of particular variables on particular factors could not be explained. A number of variables in each analysis proved to be multidimensional (i.e. they loaded on more than one factor). This was not surprising. It is to be expected that many of the complex tasks on the Binet would tap more than one important dimension.

The fourth objective was to evaluate Sattler's conceputal model for the identification of important dimensions in the Stanford-Binet both in terms of its consistency with the empirical factor structure and in terms of its reliable use by clinicians. An attempt to statistically conform the empirical factors to Sattler's conceptual factor structure was, in general, unsuccessful. Visual-Motor Integration and Memory were the only dimensions which appeared to be at all consistent. The classification of the items on each empirical factor in terms of Sattler's conceptual categories further indicated quite poor agreement. With the exception of the Visual-Motor Integration and Memory factors, the empirical factors were all represented by multiple conceptual categories. A more informal comparison, however, revealed that when the multidimensional nature of the variables was taken into account, Sattler's system was somewhat more promising than it first appeared. Sattler's Conceptual Thinking and Language categories combined, for example, overlapped considerably the empirical Verbal factors. Sattler's Social Intelligence category overlapped to a significant degree the empirical Reasoning factors. Sattler, himself, states that this should be expected. The empirical Verbal and Reasoning categories overlapped themselves to some extent and Sattler's assignment of items to the Language versus the Verbal Reasoning category reflects this interdependence.

The student clinicians assignment of the Binet items to conceptual categories clarified the situation even further. The students' first choice assignments of items to conceptual categories were in many cases discrepant with Sattler's. In other words, in the purest sense, the model was not very reliably applied by this sample of student clinicians. When, however, all of the students' choices were considered, the agreement with Sattler was much improved. The pattern of clinician judgements also showed some agreement with the empirical factor structure at years VI through XI. In some cases where there was disagreement with Sattler, the students' impressions were supported by the results of the factor analyses.

The fifth objective of the thesis was to combine the results of the present factor analytic and clinician judgement studies with past research to provide an assessment of the validity of employing subtest analysis to aid in the diagnostic use of the Stanford-Binet. The final objective of the thesis was to provide direction for further research. It is to these tasks that attention is now directed.

6.2 Implications and Directions for Further Research

The evidence that the Stanford-Binet is factorially complex would appear to be overwhelming. Its design, however, as an age scale means that its factor structure is difficult to define across the entire scale. This can only be accomplished through multiple analyses, each encompassing a relatively small portion to the total scale. The current investigation examined years VI through XI. Verbal, Reasoning, Visualization, Visual Judgement, Memory, Visual-Motor Integration, and

Control of Impulsivity dimensions were revealed. These dimensions were comparable to those found in previous investigations. This would imply that the Stanford-Binet has considerable potential as a diagnostic instrument over and above the provision of an index of general intellectual potential. With its excellent range and variety of tasks (a much greater range and variety than is present on the Wechsler Scales), the Stanford-Binet should have the potential to provide the clinician with an abundance of important clinical information. The current investigation provided only a very small step towards tapping this potential. A repetition of the current analyses employing identical statistical methods needs to be undertaken with a different sample of children. The analyses need to be expanded to eventually encompass the entire test.

Given the way the Binet has been developed, however, it is doubtful that even a very extensive factor analytic investigation of the Stanford-Binet will reveal a factor structure that is sufficiently defined that the test can ever validly be used to make definitive diagnoses of specific intellectual strengths and weaknesses. Some of the factors are likely to be defined by a relatively small number of items. Even fewer are likely to be administered in any single case. The best that could be expected from the test as it currently exists is that a better deliniation of the factor structure would aid the clinician in developing hypotheses about specific strengths and weaknesses which could subsequently be verified or disconfirmed through further testing. This in itself, however, is a very important function. The test could be vital in guiding the diagnostic process. Areas in which no weaknesses appear would likely not require further assessment. Attention could thus be focused on specific areas of concern.

The examination of Sattler's classification scheme indicated that, while some consistency does exist between the conceptual model and the empirical analyses in the kinds of dimensions being identified, rigid application of the model is not warranted either in terms of its success in determining the most important dimensions in the tasks nor in terms of its reliable use by clinicians. A new system needs to be developed which takes into account the empirical factor structure of the test and better reflects the multidimensionality of many of the items. It is possible that Valett's model which does attempt to take into account the complex nature of the items might be a somewhat better model. In the meantime, it would seem reasonable for clinicians to continue to use these models and the available factor analytic information to guide them in making diagnostic hypotheses. The fallibility of these guides must, however, be acknowledged and every attempt made to confirm any hypotheses with additional information.

From a more long-term perspective, the ultimate goal should be to revise the current Stanford-Binet so that its factor structure could be more clearly defined. The Wechsler Scales have an advantage over the Stanford-Binet in that they have an empirically well defined factor structure. The diagnostic potential of these factors has been well validated. The Stanford-Binet, on the other hand, has a much broader variety of tasks and would appear to tap a greater number of important dimensions than do the Wechsler Scales. This would appear to be particularly true at the younger age levels where many tasks directly parallel the kinds of problems encountered by young children as they experience their environment. If the variety and developmental insight demonstrated on the Binet could be combined with the factorial purity of tests like the Wechsler Scales, a substantially more diagnostically useful test would result. According to Dr. H.L. Janzen (personal communication, 1984, the University of Alberta) efforts to achieve this goal are already underway. Janzen has recently received correspondence from Dr. J.M. Sattler of San Diego State University indicating that he is currently involved in revising and re-norming the Stanford-Binet. Although the exact structure of the new form is not known, Sattler has apparently indicated that the age scale format will be abandoned in favour of a factor scale format. The writer of this thesis wishes Dr. Sattler and his colleagues the very best in this worthy endeavour.

REFERENCES

Adcock, C.J. (1954). Factorial analysis for non-mathematicians. London: Cambridge University Press.

Anastasi, A. (1983). Evolving trait concepts. American Psychologist, 38(2), 175-184.

Bossard, M. & Galusha, R. (1979). The utility of the Stanford-Binet in predicting WRAT performance. Psychology in the Schools, 16(4), 488-490.

Brody, E.B. & Brody, N. (1976). Intelligence: Nature, determinants and consequences. New York: Academic Press.

Brooks, C. (1977). WISC, WISC-R, S-B L & M, WRAT: Relationships and trends among children ages six to ten referred for psychological evaluation. Psychology in the Schools, 14(1), 30-33.

Burt, C. (1941). The factors of the mind: An introduction to factor analysis in psychology. New York: Macmillan.

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

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

Chase, C. & Sattler, J. (1980). Determining areas of strengths and weaknesses on the Stanford-Binet. School Psychology Review, 9(2), 174-177.

Davis, E. & Rowland, R. (1975). A replacement for the venerable Stanford-Binet? Journal of Clinical Psychology, 30(4), 517-521.

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

Fagan, J., Broughton, M., Clark, B. & Emerson, P. (1969). Comparison of the Binet and WPPSI with lower class five-year-olds. Journal of Consulting and Clinical Psychology, 33(5), 607-609.

Garrett, H.E. (1946). A developmental theory of intelligence. American Psychologist, 1, 372-378.

Gerken, K., Hancock, K. & Wade, T. (1978). A comparison of the Stanford-Binet Intelligence Scale and the McCarthy Scales of Children's Abilities with preschool children. Psychology in the Schools, 15, 468-472.

Guilford, J.P. (1967). The nature of human intelligence. New York: McGraw-Hill Book Company.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. Psychometrica, 19, 149-161.

Guttman, L. (1957). Simple proofs of relations between the communality problem and multiple correlation. Psychometrica, 22, 147-157.

Hallahan, D., Ball, D. & Payne, J. (1973). Factorial composition of the short form of the Stanford-Binet with culturally disadvantaged Head Start children. Psychological Reports, 32(3), 1048-1050.

Harman, H.H. (1967). Modern factor analysis, rev. ed. Chicago: University of Chicago Press.

Harris, C.W. (1964). Some recent developments in factor analysis. Educational and Psychological Measurement, 24, 193-206.

Helmes, E. & Jackson, D. (1977). The item factor structure of the Personality Research Form. Applied Psychological Measurement, 1(2), 185-194.

Hobroyd, R. & Bickley, J. (1976). Comparison of the 1960 and 1972 revisions of the Stanford-Binet L-M. Journal of Youth and Adolescence, 5(1), 101-104.

Jackson, D. & Chan, D. (1980). Maximum-likelihood estimation in common-factor analysis: A cautionary note. Psychological Bulletin, 88(2), 502-508.

Jones, L. (1949). A factor analysis of the Stanford-Binet at four age levels. Psychometrica, 14(4), 299-331.

Jones, L. (1954). Primary abilities in the Stanford-Binet, age 13. Journal of Genetic Psychology, 84, 125-147.

Kaiser, H.F, (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Kaufman, A.S. & Waterstreet, M.A. (1978). Determining a child's strong and weak areas of functioning on the Stanford-Binet: A simplification of Sattler's S.D. method. Journal of School Psychology, 16(1), 72-78.

Levenson, R. Jr. & Zino, Thomas II (1979). Assessment of cognitive deficiency with the McCarthy Scales and Stanford-Binet: A correlational analysis. Perceptual and Motor Skills, 48, 291-295.

Liungman, C.G. (1975). What is IQ? London: Gordon Cremonesi Ltd.

Lutey, C. (1966). Individual intelligence testing: A manual. Greeley Colorado: Executory, Inc.

Maloney, M.P. & Ward, M.P. (1976). Psychological assessment: A conceptual approach. New York: Oxford University Press.

McNemar, Q. (1942). The revision of the Stanford-Binet Scale: An analysis of the standardization data. Boston: Houghton Mifflin Company.

Meeker, M.N. (1969). The structure of intellect: Its interpretation and uses. Columbus Ohio: Charles E. Merrill Publishing Company.

Nunnally, J.C. (1978). Psychometric theory. New York: McGraw-Hill Book Company.

Ramsey, P.H. & Vane, J.R. (1970). A factor analytic study of the Stanford-Binet with young children. Journal of School Psychology, 8(4), 278-283.

Rummel, R.J. (1970). Applied factor analysis. Evanston: Northwestern University Press.

Sarason, S. (1976). The unfortunate fate of Alfred Binet and school psychology. Teachers College Record, 77(4), 579-92.

Sattler, J. (1965). Analysis of functions of the 1960 Stanford-Binet Intelligence Scale, Form L-M. Journal of Clinical Psychology, 21, 173-178.

Sattler, J. (1974). Assessment of children's intelligence (Rev. reprint). Philadelphia: Saunders.

Silverstein, A.B. (1965). Comparison of two item-classification schemes for the Stanford-Binet. Psychological Reports, 17, 964.

Spearman, C. (1904). "General intelligence" objectively determined and measured. American Journal of Psychology, 15, 201-293.

Spearman, C. (1923). The nature of intelligence and the principles of cognition. London: Macmillan and Co. Ltd.

Spearman, C. (1927). The abilities of man: Their nature and measurement. London: Macmillan and Co. Ltd.

Stormer, G.E. (1967). Dimensions of intellect unmeasured by the Stanford-Binet. Dissertation Abstracts, 27(7-A), 2078-2079.

Terman, L.M. (1916). The measurement of intelligence. Boston: Houghton Mifflin.

Terman, L.M. & Merrill, M.A. (1937). Measuring intelligence. Boston: Houghton Mifflin.

Terman, L.M. & Merrill, M.A. (1960). Stanford-Binet Intelligence Scale. Manual for the third revision, Form L-M. Boston: Houghton Mifflin.

Terman, L.M. & Merrill, M.A. (1973). Stanford-Binet Intelligence Scale, 1972 Norms Edition. Boston: Houghton-Mifflin Co.

Thorndike, R.L. (1973). 1972 Norms Tables. In Terman L.M. and Merrill, M.A. Stanford-Binet Intelligence Scale, 1972 Norms Edition. Boston: Hought-Mifflin Co., 1973.

Thorndike, R. (1975). Mr. Binet's test 70 years later. Educational Researcher, 4(5), 3-6.

Thorndike, R. (1978). Correlational procedures for research. New York: Gardner Press Inc.

Thurstone, L. (1935). The vectors of mind: Multiple-factor analysis for the isolation of primary traits. Chicago: University of Chicago Press.

Thurstone, L. (1947). Multiple factor analysis: A development and expansion of the vectors of mind. Chicago: University of Chicago Press.

Tuddenham, R.D. (1962). The nature and measurement of intelligence. In L. Postman (Ed.), Psychology in the making. New York: Knopf, 1962.

Valett, R. (1964). A clinical profile for the Stanford-Binet. Journal of School Psychology, 2, 49-54.

Velicier, W.F. (1976). The relation between factor score estimates, image scores and principal component scores. Educational and Psychological Measurement, 36, 149-159.

Velicer, W., Peacock, A. & Jackson, D. (1982). A comparison of component and factor patterns: A monte carlo approach. Multivariate Behavioral Research, 17, 371-388.

Vernon, P.E. (1950). The stucture of human abilities. New York: John Wiley & Sons, Inc.

Waddell, D.D. (1980). The Stanford-Binet: An evaluation of the technical data available since the 1972 restandardization. Journal of School Psychology, 18(3), 203-209.

Wolf, T.H. (1973). Alfred Binet. Chicago: The University of Chicago Press.

Wright, R. (1939). A factor analysis of the original Stanford-Binet Scale. Psychometrica, 4(3), 209-220.

APPENDIX I

Rating Form For The Clinician Judgement Study

INTRODUCTION

The Stanford-Binet, as you are aware, is one of the oldest, best validated, and most widely used individual tests of intelligence. It is rivaled only by the Weschler Scales which, along with many other tests of intelligence and achievement, have borrowed extensively from the theoretical framework upon which the Binet was developed. Although the Binet was designed to provide a single index of overall intellectual potential (i.e. and IQ score), attempts have been made to examine more specifically the item content in hopes of better identifying the various abilities being measured. This has come about not only through the more theoretical research on the nature of intelligence but also out of a real clinical need. Assessment of intellectual ability has served as a means of predicting performance. When a child's rate of learning, however, is less than what would be predicted, an IQ score neither explains nor accounts for this failure. Hence, the need for much more detailed diagnostic information which would ideally allow the clinician to profile individual strengths and deficits. There are many diagnostic tests available which presume to measure one or more of the various abilities thought to be vital to learning and such tests are invaluable. Given, however, that the administration of an individualized intelligence scale is usually routine in any thorough assessment, that the Stanford-Binet has a history of thorough validation, that many items on the Binet are similar in content to items on other diagnostic tests and finally, the practical need for efficient, streamlined assessment routines, it would be helpful if the Binet could be validly used to obtain more specific information which would aid the clinician in directing the diagnostic process. Several content and factor analyses of the Binet indicate considerable potential in this direction. As you, being users of the Binet, are aware, clinicians are currently taking advantage of this research. A considerable amount of work remains to be done, however, and that is to be the purpose of the study in which your participation is being requested.

INSTRUCTIONS

In the following pages you will find a description of a number of intellectual processes or abilities along with a list of all the Stanford-Binet items, excluding alternates. Your task is to rate each of the items according to which of the abilities you feel it is measuring at the particular age level. This is to be done by placing the appropriate ability number (i.e. 1-10) beside each item. You may feel that some of the items are measuring more than one of the given processes. In the first blank place the number of the ability you feel the item is primarily measuring. You may then use the other two blanks for any other abilities you feel that the item is significantly tapping. Rate these 2 and 3 on a priority basis.

**Please consult the Stanford-Binet manual (1972 Norms Edition) for item descriptions where needed.

ABILITIES (After Sattler, 1965)

1. LANGUAGE:	Maturity of vocabulary in regard to the pre-kindergarten level, extent of vocabulary referring to the number of words which can be defined, quality of vocabulary use, and comprehension of verbal relations.
2. MEANINGFUL MEMORY:	Also may be referred to as ideational memory and involves the ability to recall meaningful information.
3. NONMEANINGFUL MEMORY:	The ability to recall material which contains no inherent meaning. May also be thought of as rote memory.
4. VISUAL MEMORY:	Memory for visual information.
5. CONCEPTUAL THINKING:	While closely associated with language ability, this is primarily concerned with abstract thinking. Such functions as generalization and classification are included:
6. VERBAL REASONING:	This includes the perception of logical relations inherent in verbal material as well as analysis and synthesis.
7. NONVERBAL REASONING:	This involves the perception of logical relations inherent in nonverbal information, discrimination ability, and analysis and synthesis. The understanding of spatial relationships is also frequently involved.
8. NUMERICAL REASONING:	This involves the perception of numerical relations and the ability to generalize from numerical data.
9. VISUAL-MOTOR INTEGRATION:	This involves manual dexterity, eye-hand coordination, and the perception of spatial relations.
10. SOCIAL INTELLIGENCE:	Social maturity and social judgement.

**Please remove this and the preceeding page for ease of reference.

YEAR	ITEM	RATING 1	2	3
II	1. Three-Hole Form Board	____	____	____
	2. Delayed Response	____	____	____
	3. Identifying Parts Of The Body	____	____	____
	4. Block Building	____	____	____
	5. Picture Vocabulary	____	____	____
	6. Word Combinations	____	____	____
II-6	1. Identifying Objects By Use	____	____	____
	2. Identifying Parts Of The Body	____	____	____
	3. Naming Ojbects	____	____	____
	4. Picture Vocabulary	____	____	____
	5. Repeating 2 Digits	____	____	____
	6. Obeying Simple Commands	____	____	____
III	1. Stringing Beads	____	____	____
	2. Picture Vocabulary	____	____	____
	3. Block Building: Bridge	____	____	____
	4. Picture Memories	____	____	____
	5. Copying A Circle	____	____	____
	6. Drawing A Vertical Line	____	____	____
III-6	1. Comparison Of Balls	____	____	____
	2. Patience: Pictures	____	____	____
	3. Discrimination Of Animal Pictures	____	____	____
	4. Response To Pictures: Level I	____	____	____
	5. Sorting Buttons	____	____	____
	6. Comprehension I	____	____	____
IV	1. Picture Vocabulary	____	____	____
	2. Naming Ojbects From Memory	____	____	____
	3. Opposite Analogies I	____	____	____
	4. Pictorial Identification	____	____	____
	5. Discrimination Of Forms	____	____	____
	6. Comprehension II	____	____	____
IV-6	1. Asthetic Comparison	____	____	____
	2. Opposite Analogies	____	____	____
	3. Pictorial Similarities And Differences I	____	____	____
	4. Materials	____	____	____
	5. Three Commissions	____	____	____
	6. Comprehension III	____	____	____

YEAR	ITEM	RATING 1	2	3
V	1. Picture Completion: Man	____	____	____
	2. Paper Folding: Triangle	____	____	____
	3. Definitions	____	____	____
	4. Copying A Square	____	____	____
	5. Pictorial Similarities And Diff.	____	____	____
	6. Patience: Rectangles	____	____	____
VI	1. Vocabulary	____	____	____
	2. Differences	____	____	____
	3. Multilated Pictures	____	____	____
	4. Number Concepts	____	____	____
	5. Opposite Analogies II	____	____	____
	6. Maze Tracing	____	____	____
VII	1. Picture Absurdities I	____	____	____
	2. Similarities: Two Things	____	____	____
	3. Copying A Diamond	____	____	____
	4. Comprehension IV	____	____	____
	5. Opposite Analogies III	____	____	____
	6. Repeating 5 Digits	____	____	____
VIII	1. Vocabulary	____	____	____
	2. Memory For Stories: The Wet Fall	____	____	____
	3. Verbal Absurdities I	____	____	____
	4. Similarities And Differences	____	____	____
	5. Comprehension IV	____	____	____
	6. Naming Days Of The Week	____	____	____
IX	1. Paper Cutting	____	____	____
	2. Verbal Absurdities II	____	____	____
	3. Memory For Designs I	____	____	____
	4. Rhymes: New Form	____	____	____
	5. Making Change	____	____	____
	6. Repeating 4 Digits Reversed	____	____	____
X	1. Vocabulary	____	____	____
	2. Block Counting	____	____	____
	3. Abstract Words I	____	____	____
	4. Finding Reasons I	____	____	____
	5. Word Naming	____	____	____
	6. Repeating 6 Digits	____	____	____
XI	1. Memory For Designs I	____	____	____
	2. Verbal Absurdities IV	____	____	____
	3. Abstract Words II	____	____	____
	4. Memory For Sentences II	____	____	____
	5. Problem Situation II	____	____	____
	6. Similarities: Three Things	____	____	____

YEAR	ITEM	RATING 1	2	3
XII	1. Vocabulary			
	2. Verbal Absurdities II			
	3. Picture Absurdities II			
	4. Repeating 5 Digits Reversed			
	5. Abstract Words I			
	6. Minkus Completion I			
XIII	1. Plan Of Search			
	2. Abstract Words II			
	3. Memory For Sentences III			
	4. Problems Of Fact			
	5. Dissected Sentences			
	6. Copying A Bead Chain From Memory			
XIV	1. Vocabulary			
	2. Induction			
	3. Reasoning I			
	4. Ingenuity I			
	5. Orientation: Direction I			
	6. Reconciliation Of Opposites			
AA	1. Vocabuary			
	2. Ingenuity I			
	3. Differences Between Abstract Words			
	4. Arithmetical Reasoning			
	5. Proverbs I			
	6. Orientation: Direction II			
	7. Essential Differences			
	8. Abstract Words III			
SAI	1. Vocabulary			
	2. Enclosed Box Problem			
	3. Minkus Completion II			
	4. Repeating 6 Digits Reversed			
	5. Sentence Building			
	6. Essential Similarities			
SAII	1. Vocabulary			
	2. Finding Reasons II			
	3. Proverbs II			
	4. Ingenuity I			
	5. Essential Differences			
	6. Repeating Thought Of Passage I: Value Of Life			

YEAR	ITEM	RATING 1	2	3
SAIII	1. Vocabulary	______	______	______
	2. Proverbs III	______	______	______
	3. Opposite Analogies IV	______	______	______
	4. Orientation: Direction III	______	______	______
	5. Reasoning III	______	______	______
	6. Repeating Thought Of Passage II: Tests	______	______	______