THE UNIVERSITY OF ALBERTA

VALIDATION OF THE FOURTH EDITION OF THE STANFORD-BINET INTELLIGENCE SCALE: A FACTOR ANALYTIC STUDY OF THE STANFORD-BINET IV AND THE WISC-R

by

© BRUCE E. FRITZKE

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

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

FALL 1988

ISBN 0-315-45783-X

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Bruce E. Fritzke

TITLE OF THESIS: Validation of the Fourth Edition of the Stanford-Binet Intelligence Scale: A Factor Analytic Study of the Stanford-Binet IV and the WISC-R

DEGREE FOR WHICH THESIS WAS PRESENTED: Master of Education

YEAR THIS DEGREE GRANTED: 1988

(Signed) Bruce Fritzke

PERMANENT ADDRESS:

6310 - 96 Street,
Grande Prairie, Alberta

DATE: July 4, 1988

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 VALIDATION OF THE FOURTH EDITION OF THE STANFORD-BINET INTELLIGENCE SCALE: A FACTOR ANALYTIC STUDY OF THE STANFORD-BINET IV AND THE WISC-R submitted by BRUCE ERNEST FRITZKE in partial fulfilment of the requirements for the degree of Master of Education.

....
(Supervisor)

....

....

DATE : June 9, 1988

DEDICATION

To my son Jason, and my beloved wife Carol. Jason, as you grow and learn you teach me much. As I watch your intellect develop I am amazed: at the wonderful way that you were created; at the speed with which you acquire new knowledge; and at the joy with which you learn about and master your environment. And to Carol, my wife. I am constantly thankful for your desire and ability to teach and raise Jason in a positive environment filled with love. I learn much from your insightful teaching and encouraging of our son. You have both made many sacrifices in order to allow me to finish this thesis, and for this I am thankful.

Yet beyond all else, I am thankful for the love that we experience and share. I look forward to the journey that lies ahead of us.

Abstract

The purpose of this thesis was to examine the concurrent and construct validity of the Stanford-Binet IV. To this end, the component structures of the WISC-R, the Stanford-Binet IV, and the joint battery of the WISC-R and Stanford-Binet IV were explored. The clinical sample used included 168 children between the ages of 6-0 and 16-11 years; however, the adaptive format of the Stanford-Binet IV required the sample to be divided into three age groups: 6-0 to 8-11 years, 9-0 to 12-11 years, and 13-11 to 16-11 years; unfortunately, the exteme heterogeneity of the oldest age interfered with it utility and generalizability. Only subtests appropriate for each age level were included in the respective analyses. Using the principal components method of factoring, the number of components to retain was determined by the eigenvalue-one and scree test criteria; components were then rotated by the quartimax and varimax methods. The Composite score from the Stanford-Binet IV was found to have good concurrent validity with the WISC-R Full Scale IQ. Likewise, the results support a relatively strong general component (*g*) on the Stanford-Binet IV, and on the joint analysis of these two tests. As such, the construct validity of the Stanford-Binet IV as a test of general intelligence was confirmed. However, the organization of subtest into four Areas of Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning, and Short-Term Memory was not strongly supported, especially when the the effect of the general component was removed before group components were extracted., For the 6-0 to 8-11 age group, the three-component quartimax rotated solution inluded *g*, Verbal ability, and Abstract / Visual ability. In the 9-0 to 12-11 age group the three-component quartimax rotated solution included *g*, Auditory Short-Term Memory, and Visual / Conceptual Ability. The general component in these two groups accounted for 43.9% and 58.8% of the total subtest variance.

ACKNOWLEDGEMENTS

I would like to extend my thanks to those individuals on my committee: Dr. H. L. Janzen, Dr. J. G. Paterson, and Dr. G. V. Malicky. I will take a small part of you with me as my journey in life continues.

A very special thanks to Dr. H. L. Janzen for supervising this thesis. I know that the time you have taken to help me has been a sacrifice. Your knowledge, wisdom, and insight have helped to guide my thinking and writing. I am constantly amazed at the depth of your knowledge and the practical way that you apply it. Your stature has encouraged and challenged me to examine the depths of many intellectual crevices in an attempt to illuminate previously obscure areas. Your financial assistance has been greatly appreciated.

I also acknowledge the efforts of the students in the 1986-1987 class of Educational Psychology 524 for the testing that they conducted; this thesis would not have been possible without your participation. Finally, I would like to thank all the secretaries in the Education Clinic for their assistance--you made life much easier by helping in so many ways as I collected the data for this thesis.

Finally, a special note of thanks to my family. To my parents, for all the encouragement and support that was provided in innumerable ways; for the time that was taken to help speed the completion of this thesis. And to my mother- and father-in-law Dr. Joan Snyder and Dr. Vince Salvo--you have constantly been an encouragement and have remained as living proof that one can actually complete a graduate degree. Thanks for all your understanding and for all the tips that you have given. Most importantly, to my wife Carol, special thanks for watching and helping me through this degree.

TABLE OF CONTENTS

CHAPTER PAGE

CHAPTER PAGE

PAGE

PAGE

LIST OF TABLES

TABLE PAGE

TABLE		PAGE

LIST OF FIGURES

FIGURE PAGE

CHAPTER I

INTRODUCTION

The testing tradition owes much to the pioneering work of Alfred Binet. Starting with the Binet-Simon scale in 1905, the Binet tests have been periodically revised and renormed,to reflect changes in populations and improvements in psychometric methods. As Sattler (1982) points out, the Binet-Simon scale has been the "prototype of subsequent scales for the assessment of mental abilities" (p. 32). With the publication of the Stanford-Binet Intelligence Scale: Fourth Edition (Stanford-Binet IV), a break has been made from the old format of the Stanford-Binet scales. The Fourth Edition's structure and format flow from a theoretical model of intelligence. This hierarchical model of intelligence has a general factor at the apex and three group factors (Crystallized Intelligence, Fluid Intelligence, and Short-Term Memory) below this. Using this model, the subtests[1] on the Stanford-Binet IV are grouped into the following four Reasoning Areas: Verbal Reasoning, Abstract / Visual Reasoning, Quantitative Reasoning, and Short-Term Memory.

Although the *Technical Manual* for the Stanford-Binet IV (Thorndike, Hagen, & Sattler, 1986) reports the factor structure of the Stanford-Binet IV for three age groups, independent factor analysis is needed to validate the presence or absence of a general factor as well as group factors. This independent research is particularly important given that the confirmatory factor analyses done by Thorndike et al. (1986) do not clearly support the hierarchical model of intelligence which underpins the test's construction, nor do they provide firm support for the inclusion of four Reasoning Areas.

[1] In this thesis, the terms subtest and test will be used interchangeably; although Wechsler (1974) and Thorndike et al. (1986) both refer to subscales as "Tests", less confusion occurs if these scales are referred to as "subtests".

Furthermore, as this test will be used primarily for diagnostic purposes with clinically referred children, there is a need to determine the factor structure for children who have been referred for intellectual assessment. It is important to determine whether the factor structure of the Stanford-Binet IV is the same for clinically referred children as it is for normal children from the standardization sample. Identification of the factor structure underlying the Stanford-Binet IV is an important step both in providing clinicians with information on how to interpret this test, and also in making it a useful instrument for clinical diagnosis of children's intelligence.

Research on the Stanford-Binet IV is needed in order to validate its usefulness as a tool for measuring intelligence. The need to validate all new tests, including the Stanford-Binet IV, is clearly outlined by the American Psychological Association: "Validity is the most important consideration in test evaluation" (American Psychological Association, 1985, p.9). Two types of validity were addressed in this thesis. First, criterion-relate (concurrent) validity was examined by comparing the Stanford-Binet IV with the WISC-R. Second, construct validity was examined by factor analyzing the subtests on the Stanford-Binet IV.

Because the WISC-R has highly reliability (Wechsler, 1974), a fairly consistent factorial structure (Kaufman, 1975) and has become the standard against which new intelligence tests must be measured, it is an exemplar against which the Stanford-Binet IV can be validated. To ensure that the current sample is "typical" of other clinical samples, the WISC-R factor structure is also reported. Further, the joint factoring of the WISC-R and Stanford-Binet IV was conducted in order to provide additional information on both the construct and criterion validity of these two North American tests (Kieth, 1987). However, the major focus of this thesis is on the factor structure of the Stanford-Binet IV.

The existent literature on the Stanford-Binet IV is small, and relatively few published articles report research on this instrument; this is not surprising given that the Stanford-Binet IV was published in 1986. Fortunately, the *Technical Manual* for the Stanford-

Binet IV (Thorndike et al. 1986) provides a substantial amount of information on the reliability and validity of this new test. However, the need for independent research, especially in clinical populations, is urgent. The lack of research on the factorial structure of the Stanford-Binet IV necessitates that researchers rely on the literature that relates to the theoretical structure of this test. This requires a brief review of several theories of intelligence which are based on factor analytic findings. Fortunately, the Stanford-Binet IV was not developed in a vacuum; rather, the measurement of mental ability has a rich historical tradition that provides a foundation upon which the theoretical underpinnings of Stanford-Binet IV rest.

Earlier Binet scales have a rich tradition that provides a backdrop against which one's understanding of the Fourth Edition can be sharpened. Many eager researchers might be tempted to omit such a section on the historical background of previous Binet scales in that the Stanford-Binet IV has a vastly different layout than its predecessors. However, there is much continuity: the concept of general intelligence has been retained; several item types employed since the earliest Binet-Simon scales remain; and the inclusion of four group factors is partially supported by factorial work done on the earlier Stanford-Binet scales. Thus, by reviewing the literature on the previous Binet scales, it is easier to determine whether this is really a "Binet Scale" in the tradition set forth by Binet, Terman, and Merrill, or simply a new test that is getting mileage out of the Stanford-Binet name (Vernon, 1987). Also, the rationale for the new structure found in the Stanford-Binet IV, and its vast improvement over Form L-M, cannot be fully appreciated unless one carefully examines the history, development, and problems of the earlier Binet scales.

A review of the relevant literature would be incomplete without a lengthy section on the fundamentals of factor analysis. Indeed, the nature of factor analysis may seem mysterious, magical, or perhaps even esoteric to some. Nevertheless, the study of intelligence is increasingly becoming tied to factorial techniques. To prevent this discussion from detracting from the thesis, a section on factor analysis was provided in

Appendix A. Although this discussion was at a rudimentary level, the following pertinent issues were examined: impact of sample size and composition on the factor structure; description of the principal factoring method, with distinction between the principal components model and the common factor model; the number-of-factors problem; and also the principle of factor rotation[2]. Because the Stanford-Binet IV relies heavily on the theories of Vernon and Cattell, and both of these theories have resulted from factor analytic work, it is doubly important that an accurate understanding of factor analysis be achieved. Once the essentials of factor analysis are understood, one is better able to appreciate not only the subjectivity involved, but also the role that previous theoretical knowledge plays in guiding the interpretation of factors, and in infusing psychological meaningfulness into mathematically "objective" results based on matrices, vectors, and axes.

In summary, the purpose of this study was to examine the validity of the Stanford-Binet IV, especially the construct validity as revealed by the component structure[3] of this test in three different age groups. As Thorndike et al. (1986) acknowledge, there is great need to validate the findings of their factor analyses on independent samples--in order to bolster one's faith in the results they found in the standardization sample, and to extend these findings to special clinical groups. Because factor analysis is an "art" that has a subjective element, repeated verification of the Stanford-Binet IV's factor structure is needed; only with repeated findings can confidence be placed in the invariant nature of the factor structure of the Stanford-Binet IV (if it exists). In reading this thesis, one is well advised to first consider Vernon's (1961) warning that there are several problems associated with the use of factor analysis: (a) a strict hierarchical picture of mental structure

[2] For some, it may even be advisable to read the section on "Understanding The Essentials of Factor Analysis" before reading the other sections in the literature review. By doing so, those unfamiliar with factor analysis may gain a greater appreciation of the various theories that are examined.

[3] For the purpose of this thesis, principal components analysis is depicted as falling under the rubric of factor analysis; thus, "component structure" is analogous to "factor structure", with the difference being that principal components rather than factors were extracted. At times the terms components and factors are used interchangeably, though there is some technical differences between these (Gorsuch, 1983).

is an over-simplification, (b) the results depend upon the population tested, and (c) the results depend upon the test being used. Thus, a degree of skepticism is not unwarranted as one researches the factor structure of the Stanford-Binet IV.

This thesis has four basic objectives:

1. To provide descriptive statistics including means, standard deviations, and correlations for the subtests, Areas, and Composite scores on the Stanford-Binet IV as well as for the WISC-R subtests and scales; criterion validity was addressed by reporting both intercorrelations between the Stanford-Binet IV and WISC-R scores for each of the three age groups.
2. To factor analyze the WISC-R for the three different age levels (age 6-0[4] to 8-11 years; age 9-0 to 12-11 years; and age 13-0 to 16-11 years) in order to provide a reference structure against which the Stanford-Binet IV can be compared, as well as to verify whether the current sample has the prototypical WISC-R factor structure (Kaufman, 1975).
3. To provide construct validity for the theoretical structure of the Stanford-Binet IV by factor analyzing the standard scores from the appropriate subtests for each of the three different age levels.
4. To jointly factor the combined subtests from the Stanford-Binet IV and the WISC-R:
   a. to see whether the addition of different tests changes the component structure, or whether it will still conform to the factor structure of one or the other of the two tests.
   b. to provide a better understanding of which WISC-R subtests and Stanford-Binet IV subtests load on the same components, and whether or not the subtests from the two tests have loadings on the same general component.

[4] Read as six years and zero months. All ages will be expressed in this format of Years-months (i.e., 7-5 is seven years and five months).

c. to offer some suggestion to clinicians who may wish to use specific subtests or Areas from the Stanford-Binet IV in conjunction with the WISC-R subtests and scales (or vice versa) when testing children.

In this thesis, only exploratory factor analytic methods were used. Given Thorndike et al.'s (1986) failure to substantiate the theoretical structure of the Stanford-Binet IV using confirmatory factor analysis, it was deemed appropriate to use exploratory techniques in order to let the data "speak for themself". Nevertheless, some general hypotheses or trends were expected to emerge:

1. The correlations between subtests, Areas, and Composite IQ on the Stanford-Binet IV were expected to be similar to those reported by Thorndike et al. (1986).
   a. the intercorrelations between the WISC-R major scales and and the Stanford-Binet IV Area and Composite IQs were expected to be similar to those reported by Thorndike et al. (1986).
2. The component analyses ~~of the~~ WISC-R using varimax rotations, were expected to produce two or three components: Verbal Comprehension[5] (Information, Similarities, Vocabulary, and Comprehension); Perceptual Organization (Picture Completion, Picture Arrangement, Block Design, and Object Assembly); and possibly a Freedom From Distractibility factor (Arithmetic, Digit Span, and Coding).
3. The component structure for the Stanford-Binet IV was expected to demonstrate a strong general component at each of the three age levels; this general component is seen in the first unrotated component, or in the first rotated component when the quartimax method of rotation is used.

[5] As per A.P.A. style, all terms representing factors or components are presented with the first letter of the term in upper case (e.g., Fluid Intelligence rather than fluid intelligence).

4. In the Stanford-Binet IV component analysis using varimax rotations, three group components were expected at each age level. These three componentswere expected to correspond to (a) Crystallized Intelligence / Verbal-Educational , (b) Fluid Intelligence / Spatial-Mechanical-Practical , and (c) Short-Term Memory. The subtests from the Verbal and Quantitative Reasoning Areas were expected to load on (a), while the subtests from the Abstract/Visual Reasoning area were expected to load on (b), and the subtests from the Short-Term Memory Area were expected to load on (c).
5. When the Stanford-Binet IV and WISC-R subtests were factored jointly, a strong general component was expected; in the varimax rotations it was expected that the same three components would emerge as were expected from the factor analysis of the Stanford-Binet IV; the Verbal Scale subtests from the WISC-R were expected to load on Crystallized Intelligence component, the Performance Scale subtests were expected to load on the Fluid Intelligence component, and Digit Span was expected to load on the Short-Term Memory component.

To guide the reader through the literature review (Chapter II), it has been divided into eight major areas and three levels of headings have been used. The first major area deals with the early Binet-Simon Scales . The second major area provides a history of the American versions of the Binet-Simon scales, with the emphasis being on the Stanford-Binet scales. The third major area deals with the factor analytic studies done on the Stanford-Binet scales prior to the Fourth Edition. The fourth major area focuses on profile analysis done on the Stanford-Binet L-M. The fifth major area presents a brief synopsis of theoretical models of intelligence based on factor analytic work. The sixth major area provides a review of the factorial studies with the WISC-R. The seventh major area is concerned with the construction of the Stanford-Binet IV. Finally, the eighth major area covers some pertinent information on Short-Term Memory.

Because of the length of the literature review, the pertinence of each of these major areas (and sub-areas) to this thesis was addressed within the context of that section, though it may be elaborated in later sections. For example, the relevance of the Binet-Simon scales to the Stanford-Binet IV was addressed within the major section that deals with the early Binet-Simon scales; or as another example, the relevance of factor analytic procedures to the current research was discussed within the section dealing with the essentials of factor analysis.

Chapter III outlines the research design and procedure used in this thesis. It provide a description of the instruments, the subjects, the rationale for the age groupings, the procedure for dealing with missing data, a brief description of the factoring method used, and the limitations of this research.

Chapter IV presents the results from this research. To assist the reader, major and minor headings were used. The first major area deals with the effect that order of administration had on the test results. The second major area provides the means and standard deviations from the WISC-R and Stanford-Binet IV for the three age groups. The third major area gives the intercorrelations between (a) the Stanford-Binet IV subtests, Areas, and Composite at each age level, (b) the WISC-R subtests and Scales at each age level, and (c) the Stanford-Binet IV and WISC-R subtests, Areas/Scales and Composite/Full Scale IQ. The fourth major area presents the results of the principal components analyses for the WISC-R for each of the three age groups. The fifth major area reports the principal components analyses for the Stanford-Binet IV for each of the three age groups. Finally, the sixth major area provides the results from the joint analysis of the Stanford-Binet IV and the WISC-R for each of the three age groups. Because of the number of results presented, and in order to avoid confusion, a brief discussion of the results is given within the results section rather than being solely relegated to the discussion section (Chapter V).

Chapter V is devoted to discussion of the results, but focuses mainly on highlighting and addressing the objectives and hypotheses outlined in Chapter I, as well as identifying emerging patterns within and between age groups. For example, for the age group 6-0 to 8-11 years, the separate component structures of the Stanford-Binet IV and WISC-R were compared with each other and with the joint component structure of the Stanford-Binet IV and the WISC-R. Also, this chapter contains a synopsis and makes recommendations for future research.

After Chapter V, the references and various appendices are presented.

CHAPTER II

REVIEW OF THE LITERATURE

A. Historical Perspective of the Binet-Simon Scales

Binet was a true pioneer in the field of intelligence testing. He was one of the first to insist that the study of intelligence be based on the measurement of higher and more complex mental functions rather than simple sensory-motor measures as was common in the late 1800's and early 1900's. Binet's opportunity to develop a scale of intelligence occurred in 1904 when the French Minister of Public Instruction set up a commission and charged it with finding ways of identifying children with limited mental ability who could benefit from special education class As one of the commissioners, Binet realized the need for an objective method to measure the intelligence of childre [illegible] gmatic need led to the development of the 1905 Binet-Simon scale. As the ye [illegible] he 1905 scale was destined to be revised in 1908 and then again in 1911.

The 1905 Binet-Simon Scale

In a 1905 publication, Binet and Simon reviewed the necessity for a scientific diagnosis for mentally retarded children. In their review, they refer to work by Blin and Damaye who had developed a pre-arranged list of questions for twenty topics that could be used to diagnose mental deficiency (Binet & Simon, 1905/1916). Commenting on this list of questions, Binet and Simon state "let us say that in precision Dr. Blin's study seems to us superior to anything previously accomplished. The criticisms which we shall make will not cause us to forget that we have here a first attempt to apply a scientific method to the diagnosis of mental debility" (Binet & Simon, 1905/1916, p. 28). Thus, the ingenuity of Binet and Simon was their ability to discern that a standardized set of questions could be used to assess the intellectual potential of children. This concept has continued until this day.

The earliest versions of the Binet-Simon scales were not designed around a theoretical basis. Rather, Binet and Simon constructed an instrument that would empirically identify children with different scholastic abilities. The pragmatic nature of Binet and Simon's work can be seen in a 1905 publication by these two authors: "The scale that we shall describe is not a theoretical work; it is the result of long investigations, first at the Salpetriere, and afterwards in the primary schools of Paris, with both normal and subnormal children" (Binet & Simon, 1905/1916, p. 41). Without a theory to guide the construction of their scale, Binet and Simon proceeded in what has been called a "trial and error" process, refining their idea of intelligence as they went along (Cronbach, 1960, p. 164).

Nevertheless, as the year passed Binet and Simon's conception of intelligence was clarified. After testing many children, and after hours of direct clinical observation, Binet and Simon (1905/1916) arrived at a tentative definition of intelligence:

> It seems to us that in intelligence there is a fundamental faculty, the alteration or the lack of which, is of the utmost importance for practical life. This faculty is judgment, otherwise called good sense, practical sense, initiative, the faculty of adapting one's self to circumstances. To judge well, to comprehend well, to reason well, these are the essential activities of intelligence (p. 42).

From this definition it is clear that Binet and Simon believed that intelligence would be manifest in practical ways and that judgment, comprehension, and reasoning were to be considered of foremost importance.

The content of the 1905 Binet-Simon scale was not totally new; many of the items had been used on earlier tests or experimental research. However, the combination of previously existing items into a single scale was new. More importantly, Binet and Simon utilized an empirical / psychometric approach in constructing their scale. In total, the 1905 Binet-Simon scale contained 30 items (interchangably called "tests" by Binet & Simon).

These were arranged in ascending order of difficulty as determined by testing of normal and mentally retarded children. The items covered topics such as visual coordination, recognition of food, performance of simple instructions, verbal knowledge of pictures, naming objects, repeating digits, repeating sentences, definitions of objects, immediate recall of pictures, sentence completion, paper cutting, and definitions of abstract terms (for a c[illegible] listing of the 30 items see either Freeman, 1955, pp. 102-104; or Binet & Simon, 1905/1916, pp. 45-68). Several of these item types, though perhaps modified, have been retained in subsequent Binet scales--including the Stanford-Binet IV. This heterogeneous sampling of items that began with 1905 scale was continued through the various Stanford-Binet scales (up to and including Form L-M) and was a hallmark of the Binet tests.

Before moving on to the 1908 Binet-Simon scale, it is important to note the relationship that Binet and Simon saw between learned knowledge (e.g., Crystallized Intelligence) versus more innate ability (e.g., Fluid Intelligence). In 1905, Binet and Simon made the following comment: "Our purpose is to evaluate a level of intelligence. It is understood that we here separate natural intelligence and instruction. It is the intelligence alone that we seek to measure, by disregarding in so far as possible, the degree of instruction which the subject possesses" (Binet & Simon, 1905/1916, p. 42). Lest the reader be fooled into thinking that Binet and Simon did not regard acquired knowledge as part of intelligence, a digression to a 1908 publication is called for, as it clearly elaborates on this issue: "In previous publications we have shown that it is possible to divide the methods of measuring intelligence into three groups: (1) the anatomical method, (measurement of the cranium, of the face, of corporeal development; observation and interpretation of stigmata of degeneracy, etc.); (2) the pedagogical method (measurement of knowledge acquired at school, principally in spelling and arithmetic); (3) the psychological method (measurement of the uncultured intelligence)" (Binet & Simon, 1908/1916, p. 183).

Binet and Simon focussed their attention on the "psychological method" as they called it. They believed that their tests indeed tapped this uncultured intelligence rather than knowledge taught in school. The "pedagogical method" described by Binet and Simon can be conceptualized in two ways: first, it can be thought of as scholastic achievement; second, it can be conceptualized as learned or Crystallized Intelligence (e.g., Cattell, 1971). Whether or not the 1905 Binet-Simon scale measured acquired knowledge, the important point is that Binet and Simon hoped that their scale would measure "uncultured intelligence"--or what Cattell (1971) calls Fluid Intelligence.

The 1908 Binet-Simon Scale

The 1905 scale was a bold step, but both Binet and Simon recognized several problems with this crude scale and consequently they revised it in 1908. Like its forerunner, the 1908 scale provided a pragmatic measurement of intelligence. The authors were not pretentious, and acknowledged that the nature of children's intelligence was still not clear to them: "The child differs from the adult not only in the degree and quantity of his intelligence, but also in its form. What this childish form of intelligence [illegible] do not yet know" (Binet & Simon, 1908/1916, p. 183). They were, however, becoming more confident that intelligence was primarily innate rather than a learned ability: "Furthermore, the intellectual faculty appears to us to be independent not only of instruction but of that which may be called the scholastic faculty, that is to say, the faculty of learning at school, the faculty of assimilating the instruction given in school with the methods used in school" (Binet & Simon, 1908/1916, p. 254).

Structurally, the number of items on the 1908 scale had increased from the 30 (on the 1905 scale) to 58, and the age range was broadened to allow testing of children between the ages of 3 and 13 years. However, as Binet and Simon stated, the scale was designed primarily for school age children. As a result, the questions at age three were of somewhat questionable difficulty for that age. Based on the results of previous trials and administrations, the tests were once again arranged in order of increasing difficulty. Two new features introduced on the 1908 scale have made a substantial contribution to the area of mental measurement: first, the tests were grouped together on the basis of age-levels; second, the concept of mental age was employed for the first time (Freeman, 1955). Other improvements over the 1905 scale included better details for administration and scoring.

The 1911 Binet-Simon Scale

Criticisms of the 1908 scale focussed on the appropriateness of item placement in terms of age levels. According to Freeman (1955), one of the major changes in the 1911 revision was relocating several tests to higher age-levels. Also, several tests were omitted because "they seemed to depend too much on school learning or on very incidental information" (Freeman, 1955, p. 110). An important change in the 1911 scale was its extended age range. It was designed for use with children between the ages of 3 and 15, though a limited category for adults was provided. At each age level there were five items (Anastasi, 1982; Fancher, 1985b). This expansive age coverage has been retained in future Binet scales, and has held much appeal in that a single test can be used at different ages. Also, the idea of placing five items at each age level was adopted on the American revision of the Binet scales.

Implications of the Binet-Simon Scales for the Stanford-Binet IV

From the Binet-Simon scales several concepts like general mental ability, mental age, age scale, and basal age / testing ceiling became part of the testing tradition (French, 1986). Furthermore, several item types from the Binet-Simon scales survived through the various American revision of the Binet scales and appear on the Fourth Edition. Although there has been this continuity in the content, the theoretical underpinnings of Stanford-Binet IV have finally broken from the Binet-Simon conception that intelligence tests measure only innate intelligence at the exclusion of learned abilities. Thus, the authors of the Stanford-Binet IV (Thorndike et al., 1986) depart from Binet and Simon's (Binet & Simon, 1905/1916; 1908/1916) belief that innate abilities are independent of abilities learned in school. This change fits with the trend in the last few decades of questioning whether intelligence tests involve a scholastic/educational component (e.g., Sattler, 1982). In this respect, the Stanford-Binet IV has broken with the tradition of the older Binet scales and has adapted to a more workable differentiation between learned versus innate intelligence (Horn & Cattell, 1966).

B. The American Versions of the Binet Scales

The early Binet-Simon scales were of considerable interest to psychologists and educators of the day in that they promised to objectively distinguish between children of different mental abilities. In the United States, several different revisions of the Binet-Scale were undertaken between 1908 and 1923. By far the most popular American version was done at Stanford University in 1916 by Terman.

The 1916 Stanford-Binet

The most successful American revision of the Binet-Simon scales was done at Stanford University by Terman in 1916. This revision was appropriately called "The Stanford Revision of the Binet-Simon Intelligence Scale". From this revision, a testing tradition in the United States was destined to follow. Not only were many new questions developed for the Stanford scale, but the entire scale was also standardized on an American population. A total of 90 items were included on the 1916 Stanford-Binet. Of these, 54 had been adapted from the 1911 Binet-Simon scale, 5 were from earlier Binet-Simon scales, 4 were from contemporary American tests, and 27 were new items (Freeman, 1955). It is not surprising, then, that Anastasi suggests that, "This revision introduced so many changes and additions as to represent virtually a new test" (1982, p. 228). Nonetheless, in recognition of the guiding principals set down by Binet, his name was retained on the 1916 Stanford test.

Although the major age range covered on this new Stanford-Binet was from 3 to 14 years, there was also a group of items for an "average-adult" level and another group at the "superior-adult" level. Despite this attempt to cover all age levels above three years, sampling inadequacies essentially made it unadvisable to use the scale for anyone older than 14 years of age. For ages 3 to 10, there were six test items at each age. Surprisingly there were no items for ages 11 and 13, though there were eight items for age 12, and six items for age 14. These gaps are interesting in that the 1908 Binet-Simon scale also had some problems with placement of items at the higher age levels. Freeman (1955) suggests that these gaps may have occurred because the test constructors found it difficult to devise tests that would indicate a one year age difference at these stages of mental development, or perhaps test items for these ages did not exist in 1916.

The 1916 Stanford-Binet had problems at the older age levels, especially in rating bright students over the age of twelve to fourteen years (Merrill, 1938); it was apparent that the 1916 Stanford-Binet was not finely tuned for use at the upper age levels. Another problem inherent in the scale was that the concept of mental age beyond age 16 was not the same as it was below age 16[6]. Although the 1916 Stanford-Binet remained basically an age-scale yielding a mental age, it also provided an intelligence quotient (IQ). Following Stern's suggestion, IQ was calculated by the formula IQ = MA / CA X 100. Scrutiny of the distribution of IQs indicates that they were basically normally distributed at each age level. Unlike later Stanford-Binets, however, the 1916 version had a standard deviation of 12 points.

The concept of a "general intelligence" was retained from the Binet-Simon scales, and "no attempt was made to measure separate mental faculties such as memory, attention, sensory discrimination etc" (Terman & Merrill, 1973, p. 6). Indeed, the 1916 Stanford-Binet was destined to go through several more revisions until these types of separate abilities would finally be measured on the Stanford-Binet IV. As for the 1916 Stanford-Binet, it was designed as a measure of global intelligence, and this was to have a pr[illegible] effect in the area of testing for many years.

---

6 Age 16 was taken to be the peak of intellectual growth on the 1916 scale (Terman & Merrill, 1938).

The criticisms of the 1916 Stanford-Binet have been summarized effectively by Freeman (1955). First, the scale was not useful as a measure of adult intelligence or as a measurement tool for young children--rather it was most useful for assessment of children between the ages of 5 and 14 years. Second, the standardization group (1,000 Californian children) was not necessarily representative of the American population. Third, the scale was very heavily weighted with verbal and abstract materials. Fourth, some procedures for administering and scoring the scale were rather subjective and thus somewhat problematic.

The 1937 Stanford-Binet Forms L & M

The second Stanford-Binet revision occurred in 1937. The impetus behind this revision was to clarify instructions for administration and scoring, to improve low validity associated with certain tests in the scale, to restandardize the test, and to improve inadequacies at ages below 4 years and above age 15. According to Terman and Merrill, the three aims of the 1937 revision were defined by the following characteristics: "(1) expresses test scores in terms of age levels of performance, (2) employs a variety of tasks designed to gauge problem-solving ability along many lines not primarily dependent upon specific training, and (3) aims to measure 'general intelligence'" (1953, p. 510).

Like the 1916 scale, the 1937 Stanford-Binet used a wide variety of items rather than a few item types grouped into subscales. The rationale behind the design of the 1937 Stanford-Binet can be seen in a comment by Terman and Merrill (1937):

> The variety provided by the ever-changing tasks insures the zestful cooperation of subjects and is at the same time based upon what we believe to be sound psychological theory. It is a method which, to paraphrase an oft-quoted statement by Galton, attempts to obtain a general knowledge of the capacities of a subject by the sinking of shafts at critical points. In our revision we have greatly increased the number of shafts and have sunk them at points which wider experience with tests has shown to be critical (p. 4).

In the process of revising the Stanford-Binet, Terman and Merrill were not unaware of criticisms against tests that only provide a global measure of intelligence. Yet they steadfastly held to the Binet tradition of measuring intelligence by taking a wide sampling of a person's knowledge and ability. In McNemar's (1942) book, Terman provides an introductory chapter where he defends the design of the 1937 Stanford-Binet:

> The Binet scale has often been criticized because of its great variety of brief, disconnected tests--a 'motley array,' as Spearman scornfully refers to them. According to some critics, if there is any large general factor measured by such tests it is purely accidental. They contend that the logical way to proceed is to devise a few series of tests, each series containing many items of a given kind and so, presumably, measuring thoroughly a given aspect of intelligence. The method they recommend has its advantage in group testing in that it simplifies administration procedures, but no test of this type has ever been devised that rivals the Binet test for clinical use with children. The latter, for all its 'motley array' of variegated tasks, not only is more interesting but also affords a better measure of all-round intellectual development than any of the substitutes that have been suggested. Binet's abandonment of the attempt to test the intellectual 'faculties' as such was his outstanding contribution to psychometrics (McNemar, 1942, p.6).

In 1942, Terman may have been correct in his assertion that the Stanford-Binet had no rival, but with the introduction of the WISC in latter years, Terman's words stand as an antiquated, perhaps elitist statement. In fact, as needs of educators and clinicians changed tests providing only global measures of intelligence fell into disrepute (Sattler, 1982) and tests with subscales became more prominent (Reynolds & Sundberg, 1976).

The structure of the 1937 test, as can be seen clearly from the previous quotes, was designed to measure general intellectual ability rather than specific or group abilities (Terman & Merrill, 1973). In revising the test, the authors claimed that they surveyed the literature and examined "thousands of test items" that had previously been used or suggested as measures of intelligence. To ensure that a general factor would emerge, one of the criteria for selecting items was that "the biserial correlation of each subtest with the total score had to be high enough to indicate that each contributed to what the scale as a whole was measuring" (Terman & Merrill, 1973, p. 9). Thus, the procedure for the 1937 revision virtually guaranteed that a general factor would be measured. Items that did not correlate highly with the total scale, and thus presumably did not measure general intelligence well, were not included.

The 1937 Stanford-Binet had two parallel forms (L and M). Whereas the 1916 Stanford-Binet had contained only 90 items, there were 129 items on each of the two forms. As noted by Terman and Merrill (1937) Form L bore slightly more resemblance to its predecessor than did Form M. Consequently, in the following years Form L tended to be used substantially more than Form M. As noted by Freeman (1955), of the 258 coefficients obtained when Forms L and M were correlated, 78 percent were .50 or higher; this does lend some support to the conception that general intelligence (or some common construct) was being measured on both of Forms L and M.

In response to previous criticisms that the 1916 Stanford-Binet had been too verbal in nature, Terman and Merrill attempted to include more non-verbal item types on the 1937 revision. However, they acknowledged that they were unable to develop effective non-verbal tests for any but the lowest age levels. As such, the upper levels remained highly verbal in nature. Referring to the older age levels, Terman and Merrill (1937) make the following comment:

> At these levels the major intellectual differences between subjects reduce largely to differences in the ability to do conceptual thinking, and facility in

dealing with concepts is most readily sampled by the use of verbal tests. Language, essentially, is the short-hand of the higher thought processes, and the level at which this shorthand functions is one of the most important determinants of the level of the processes themselves (p. 5).

Although it is possible that verbal skills are more or less synonymous with intelligence in older children and adults, it is also wise to consider Vernon's (1961) suggestion that the emergence of a large Verbal factor may be an outcome of the education system in North America; put another way, the Verbal-Educational factor (*v:ed*, Vernon, 1961) on most intelligence tests is larger than and accounts for more variance than non-verbal factors. With this in mind, consider Terman and Merrill's (1973) statement: "Many of the so-called performance test items tried out for inclusion in the scale were eliminated because they contributed little or nothing to the total score. They were not valid items for this scale" (Terman & Merrill, 1973, p. 8). Insistence on measuring only a general factor (which appeared largely verbal in nature) may have ruled out the inclusion of non-verbal item types.

The age range for the 1937 scales was extended downward to age 2 and upward to include three adult levels including one labelled "superior adult". The arrangement of items at each age level was much improved over the 1916 scale. Below the five-year-old level tests were spaced at half-year intervals and the gaps at ages 11 and 13 had been filled. Although the standardization sample included only white American-born subjects, it was relatively large with 3,184 children ranging in age from 2 to 18 years. The sample was taken from 11 states and some attempt was made to obtain a sample that was representative of the United States population in terms of social situation and father's occupation. Despite these attempts, the children tended to have a slightly higher socioeconomic status than that indicated by the 1930 census data and rural children were under represented. Thus, although the 1937 standardization sample was considerably improved over the 1916

The 1937 Stanford-Binet remained an age scale, with items grouped by age. Item were retained only if the percentage of children passing that item increased with age. Because the adult mental age was set at sixteen years, problems emerged in calculating the IQ of older subjects. Evidence indicated that gains in IQs began to drop after the age of thirteen rather than sixteen and that there was little increase in ability after age sixteen (Terman & Merrill, 1937, 1953). Taken together, these problems indicate that there was an inadequate ceiling for older children with above average ability, and also for adults. Even when mental ages were calculated at the upper age levels, their meaning was opaque. Thus, even though the 1937 Stanford-Binet was an age scale, the items were arranged so that a mean IQ of approximately 100 was obtained at each age.

Several criticisms of the 1937 scale exist and have been outlined by Freeman (1955). First, there was a considerable variation (from 12.5 to 20.0) in the standard deviations at different ages. This created considerable problems in interpreting IQs as those obtained at different ages did not mean the same thing, as the same score at different ages might represented different relative standing (i.e., different z-score). Second, there was some concern as to whether the 1937 Stanford-Binet was measuring different abilities at different age levels. Third, though theoretically usable for testing adults, the oldest subjects in the normative sample were 18; therefore, when testing adults, no direct comparison with a normative group was available; further, the difficulty level for older subjects was still too easy, especially for brighter adults. Finally, McNemar (1942) found that items were more reliable for older children than for younger children, and more reliable for children with lower IQs than for those with higher IQs.

The 1960 Stanford-Binet Form L-M

The third revision of the Stanford-Binet occurred in 1960. The purpose for the 1960 revision was clearly stated by Terman & Merrill (1973): "The present revision of the Stanford-Binet Tests aims at providing test users with a single scale that, while preserving the characteristic features of previous revisions, eliminates out-of-date content and improves general structure" (p. v). In this revision the best subtests from forms L and M were incorporated into a single scale with 138 items. At each age level, the six test items were supplemented by an alternate item that could be used if one of the six test items was spoiled. Appropriately, the third revision became known as the Stanford-Binet Form L-M--reflecting the merger of Forms L and M. No new content was introduced on this revision, and the overall structure of the test remained the same as the 1937 scales.

The selection of subtests for Form L-M was based on a sample of 4,498 subjects who had been given the Stanford-Binet between 1950 and 1954. Using this sample, changes in item difficulty from the 1937 scale were noted; these changes were used to determine placement of items on Form L-M. As in previous revisions, a subtest was kept only if its biserial correlation with the total score was relatively strong--the mean biserial correlation for items on the 1960 Stanford-Binet was .66 whereas on the 1937 scales it was .61. The implication is that the 1960 Stanford-Binet may have been a more consistent measure of general intelligence than either Forms L or M were.

The content of the 1960 revision had much in common with earlier Binet Scales. According to Popplestone and McPherson (1974) the origin of Form L-M's content can be traced to the following sources: 8% to the Binet-Simon scales of 1905, 1908, and 1911; 27% to the 1916 Stanford-Binet; and 8% to Terman's research in 1918. Thus, approximately 43% of the Stanford-Binet Form L-M was developed during or before 1918, and just as important, 100% of the content was present on the 1937 revision.

Perhaps the most important change on Form L-M was the adoption of deviation IQs (Mean of 100 and standard deviation of 16) rather than ratio IQ's which had been used in the 1937 revision. Also, the variability of scores (different standard deviations for different ages) on the 1937 scale had been taken into account when the new IQ tables were constructed for the 1960 scale. The result of these changes was that the obtained IQ's on the 1960 scale were comparable across age levels. Because of data suggesting that mental growth on the Stanford-Binet continued beyond age 16, these new IQ tables had been extended to provide norms for children between the ages of 17 and 18. Despite this change to deviation IQs, Terman and Merrill were not willing to abandon the concept of mental age; they felt that mental age was useful in helping educators understand the level at which children function.

In an unexpected fashion, and apparently contrary to their steadfast position that the Stanford-Binet measures general ability, Terman and Merrill (1973) digress to a discussion of three possible categories of items on the 1960 scale. These three categories represented a verbal scale, a non-verbal scale, and a memory scale as suggested by McNemar (1942) in his factor analysis of the 1937 scales. The proposed items on the "verbal scale" had an average correlation of .65 with the total score; items on the "non-verbal scale" had an average correlation of .58 with the total score; and items with a memory content correlated .61 with the total score. Although Terman and Merrill do not deny the possibility of group factors as pointed out by both McNemar (1942) and Jones (1949), they emphasized that items were selected in order to maximize first factor loadings--that is, to measure general intelligence. What can one conclude from Terman and Merrill's reported correlations on these three categories? If nothing else, it appears to be an indication that they were willing to consider the possibility that group factors exist in intelligence, perhaps even on the Stanford-Binet Form L-M.

The 1972 Renorming of the Stanford-Binet Form L-M

Despite the combining of Forms L and M into a single scale, the 1960 revision did not involve a restandardization. This restandardization did not occur until 1972. In 1970, the Cognitive Abilities Test (CAT) had been normed using a stratified sample of the U.S. population. This stratification was based on community size, geographic region, and economic status. The 1972 norming sample for the Stanford-Binet L-M rode piggy-back on the CAT norming sample. That is, from the CAT sample, 100 children at each of 21 age levels were chosen so that their CAT scores had a mean of 100 and a standard deviation of roughly 16. In actuality, the number of children per age level varied from 43 to 150

(Waddell, 1980). Furthermore, as several authors have pointed out, the 1972 standardization did not ensure that ethnic minorities, social economic status, urban / rural communities, and sex were adequately controlled in the sample, even though they may have been controlled in the Cognitive Ability Test norming (Davis & Rowland, 1974; Waddell, 1980). The haunting conclusion is that the 1972 norming sample may not have adequately represented the U.S. population. Nevertheless, it was certainly a vast improvement over the antiquated sample from the 1937 revision that provided the norms for the 1960 revision.

With the new IQs based on the 1972 norming sample, the concept of mental age was no longer useful because it no longer corresponded to a child's chronological age (Salvia, Ysseldyke, & Lee, 1975). Salvia et al. provide "test ages" that correspond to the outdated mental ages, and suggested that examiners abandon the reporting of mental age in favor of the test age. In essence, these test ages are simply mental ages that have been revised in order to take into account the decrease in item difficulty from 1937 to 1972. For example, a mental age of 5-0 is translated into a test age of 4-6 which corresponds to the child's chronological age.

Corresponding to changes in mental ages, item difficulties decreased from 1937 to 1972. That is, children in 1972 found items to be easier than did children in 1937. It was also noted that items tended to be more age specific in the 1972 sample; put another way, there was a more rapid age change in the percentage of children passing an item in the 1972 sample (Garfinkel &Thorndike, 1976). The rate of change in item difficulty was not uniform across age or ability l[illegible] Children between the ages of two and six appeared to have the largest gains in IQ; between ages six and ten gains in IQ were small; after age ten gains in IQ gradually increased again (Flynn, 1984; Waddell, 1980). Similarly, there was a greater change in IQs for children with average or above average ability than there was for children with below average ability. Given these patterns, Holroyd and Bickley (1976) suggest that discrepancies between the 1960 norms (which in essence are 1937 norms) and the 1972 norms can not be attributed to educational and environmental advantages for children in the 1970's; rather, they suggest that the discrepancies are most likely methodological in nature. Overall, these disconcerting findings cast a certain amount of suspicion on the 1972 norms.

C. Factor Analytic Studies of the Stanford-Binet Scales

Despite the insistence by Terman and Merrill (1937; 1973) that the Stanford-Binet scales were measuring general intelligence, several researchers have persisted in attempting to find group factors on these scales. The factor analytic work done on the Stanford-Binet scales and its relationship to the Stanford-Binet IV are examined next. For the sake of clarity the analyses of the 1916, 1937, 1960, and 1972 Stanford-Binet have been treated separately, and in chronological order. Although each analysis contributes to our understanding of the Binet scales, McNemar's (1942) analysis of the 1937 Stanford-Binet deserves special attention because his work on the standardization sample from the 1937 scale was extensive. The goal of reviewing the factor analytic studies of the earlier Stanford-Binet scales is to determine whether there is a precedent for Thorndike et al.'s (1986) organization of the Stanford-Binet IV into four Reasoning Areas.

General Intelligence On the Stanford-Binet Scales

Analysis of Terman and Merrill's position on the structure of the Stanford-Binet scales is a good starting point. Although the possibility of group or specific factors was not ruled out on Form L-M, factor analysis done by McNemar on scales L and M had indicated that "a single common factor would explain performance on the Stanford-Binet" (Terman & Merrill, 1973, p. 35). Because Terman and Merrill (1973) used the results from McNemar's factor analysis of the 1937 scales to determine which items would be retained on Form L-M, the unitary nature of the latter scale and the exclusion of group factors was to be expected[7]. Thus, the development of the Stanford-Binet scales after 1916 was aimed at honing a general factor and likely reducing group factors.

Before jumping to the conclusion that Terman and Merrill were diametrically opposed to the idea of group or specific factors on the Stanford-Binet scales, consideration of their openness on this issue is appropriate: "The interpretations of these other statistical investigations of Stanford-Binet test data indicate that the 'organization or intelligence' that is measured by the test can be described in terms of general, group, and specific factors" (1973, p. 35). Thus, although Terman and Merrill explicitly adhered to a measurement of a general factor of intelligence, they were also willing to admit that group or specific factors might be part of this general intelligence as measured on the Stanford-Binet Form L-M. Through this acknowledgement, the authors of the Stanford-Binet scales set the stage for factor analytic work on these scales.

Wright's Analysis of the 1916 Stanford-Binet

Wright's (1939) analysis of the 1916 Stanford-Binet was rather incomplete and her interpretation of the results was somewhat questionable. Nonetheless, her findings are worth mentioning. Her sample consisted of 456 children with mental ages ranging from 6 to 14 years, and chronological ages between 10 and 10-11 years. Only items that were passed by more than 10% and less than 90% of the children were used in the factor analysis--a total of 31 items. The centroid method of factoring was used followed by graphical rotation of factors. The first unrotated factor appeared to represent general intelligence, with item loadings ranging from .57 to 1.04[8].

Though Wright extracted and rotated seven factors only five were named: Number, Spatial, Verbal Relations, Reasoning, and Induction. Meanwhile a "common factor"

---

[7] Terman and Merrill (1973) report that a weighting procedure was used to decide which items to retain on Form L-M; items were selected on the basis of their contribution to the measurement of general intellectual ability.

[8] The methodology of using the centroid method and tetrachoric correlations sometimes results in Heywood cases.

emerged as the third rotated factor. Although Wright acknowledged that this might represent general intelligence she strongly believed that it was a "maturational" factor (Wright, 1939). She reasoned that because Thurstone (1938) had not found a general factor of intelligence that it probably didn't exist in her data either; this left her with the conclusion that the general factor in her data must be a maturational factor. Reasoning like this demonstrates how previous research can influence interpretation of factors; it also serves as a warning to carefully consider a broad range of relevant literature before interpreting the results from a factor analysis.

Burt and John's 1942 Analysis of the 1916 Stanford-Binet

Burt and John's (1942a, 1942b) sample included 483 children with chronological ages between 10 and 14 $^1/_2$ years, and mental ages between 10 and 11 $^1/_2$ years. Centroid factoring using tetrachoric correlations was used and four factors emerged: the first factor (accounting for 45% of the variance) was interpreted as general intelligence; the second appeared to be a maturity factor[9] ; the third factor was bipolar with one part verbal in nature and the other nonverbal (numerical and spatial); finally, the fourth group factor, again bipolar, included memory, vocabulary, and comprehension.

The interpretation of Burt and John's (1942a) results was complicated by the bipolar nature of several of their factors. To deal with this problem, Burt and John (1942b) reanalyzed their original (1942a) data using a group-factor method; that latter method produces positive group factors in place of bipolar factors. This time seven factors besides general intelligence (the latter accounting for 35% of the variance) emerged: Maturity, Vocabulary, Comprehension, Number, Spatial, and Memory. A major problem was that these seven group factors tended to overlap and were not all of the same magnitude; that is, four were broadly defined while three were fairly narrow. Conclusions based on Burt and John's 1942b work need to be taken cautiously, as over-factoring may have occurred.

---

9 As noted by Burt and John (1942a) this maturational factor may have emerged because the 1916 Stanford-Binet did not have any items for children with a chronological age of 11.

McNemar's 1942 Analysis of the 1937 Stanford-Binet Standardization Sample

McNemar's (1942) factor analysis of the 1937 Stanford-Binet standardization sample is one of the most conscientious, comprehensive, and straightforward analyses done on a Stanford-Binet scale. As was common in that era, McNemar used the centroid method of factoring. Analyses were done separately for each of the 14 different age groups in the 1937 standardization sample. The number of people in each age group ranged from 99 to 200, and the number of items ranged from 19 to 30. Lacking a definite criteria to determine the appropriate number-of-factors, and despite stated reservations, McNemar chose to extract three factors at each age level.

Before looking at the results obtained by McNemar, an inherent limitation of factor analyses of the Stanford-Binet scales (1916-1972) deserves consideration. This limitation deals with the reliability of individual items on the Stanford-Binet scales. McNemar makes the following astute observation: "In general, one cannot expect to find large factor loadings or large common factor variances for individual items because of their relatively low reliabilities. In other words, each variable or item will yield a rather large error variance which coupled with possible specific variance will tend definitely to limit its communality" (p. 101). The problem of low reliabilities presents a problem for factor analysis, in that the communality of an item generally cannot exceed its reliability. Thus, with an average reliability of about .65, the analysis of items is somewhat problematic. The major problem this creates is that the percentage of total variance accounted for by the factors tends to be lower than the error variance combined with the specific variance. Put simply, too much of the variance from the items is unaccounted for by the factors.

McNemar (1942) clearly outlines his conception of intelligence before presenting his factor analytic results: "Without committing ourselves to any particular theory as to the organization of intellectual abilities, we are inclined to the position that a useful measuring scale should be highly saturated with one common factor to the exclusion of all conspicuous group factors" (McNemar, 1942, p. 99). Because a researcher's expectations influence interpretation of factors, it is important to know his or her bias. Indeed, McNemar's expectation of a strong general factor was not unfounded but almost guaranteed by the way items were selected for the 1937 Stanford-Binet scales: (a) a priori selection of items for the initial tryout, (b) checking of tryout items against the 1916 Stanford-Binet, and (c) insistence that each item show a substantial correlation with the composite (McNemar, 1942). Not surprisingly, McNemar found that a general factor emerged at each age level: "The results indicate that the several items included in a given analysis tend to be saturated, though in varying degrees, with a common factor" (McNemar, 1942, p. 168). Based on overlap of items in his analyses, he further concludes that this general factor is probably measuring the same construct at each age level.

In contrast to the strong general factor found at each age level, McNemar (1942) states that no "conspicuous" group factors were found. However, he does acknowledge the existence of small "minor" group factors for several ages. McNemar (1942) found that on the average, the first factor accounted for about 40% of the variance; the second factor accounted for 5-11% of the variance; and the third factor accounted for 4-7% of the variance. According to Thorndike (1975), reanalysis of McNemar's correlations with newer factoring procedures indicates that approximately 80% of the common-factor variance is extracted by the first factor. Thus, the general factor appears to have saturated the 1937 Stanford-Binet scales. With due caution, McNemar warns against the overinterpretation of the group factors in that their structure is not always clear. With this warning in mind, and remembering that the magnitude of the general factor is far greater than any of these group factors, it is now time to review the minor factors that McNemar found at some of the age levels.

For ages 2, and 2 $^1/_2$, the second factor[10] involved "identifying" or "knowing" of objects while the third factor included items related to "motor" and "memory". For ages 3, 3 $^1/_2$, 4, and 4 $^1/_2$ McNemar concluded that no logical groupings appeared in the second and third factors, and hence no attempt was made to interpret them. For ages 5 and 6 there were two separate abilities: one being verbal, the other numerical. For age 7 there were no distinct groupings. For age 9 and 11, the second factor appeared to have two components, one part being verbal, and the other being memory oriented; although unclear, the third factor was also bipolar with one pole involving memory for designs while the other involved repeating of digits. For age 13 there appeared to be a groupings for verbal items, and a grouping for "problem" items and memory items. For age 15, there was once again a differentiation between "problem" and verbal items. Finally, for age 18 the second factor involved verbal and immediate memory, while the third factor was difficult to interpret.

Overall, McNemar states "For any one given analysis there is considerable variation from item to item in first factor loadings" (p. 110). Fortunately, he summarizes which items have high and low loadings on the general factor. Because several of these item types have been retained on the Stanford-Binet IV, it is relevant to consider their relationship to general intelligence. Items with high loadings included comprehension, vocabulary, similarities, verbal absurdities, and arithmetic reasoning; those with low loadings included copying of figures, building with blocks, motor coordination, paper folding, copying of beads from memory, picture absurdities, memory for stories, paper cutting, and repeating digits. Thus, some items appeared to be better measures of general intelligence while others had more unique content.

Attempting to integrate the overall pattern of factor loadings at the different age levels, McNemar (1942) suggested that three "special scales" might be formed from the

10 In McNemar's (1942) analyses at each age level, general intelligence always emerged as the first factor.

Stanford-Binet items. These special scales do not seem to have come directly from McNemar's factorial work but rather from an a priori conception. These three scales were (a) vocabulary, (b) non-verbal, and (c) memory. For each scale McNemar gave not only the items from the Stanford-Binet that should be included, but also means and standard deviations for each of the items involved. Given McNemar's commitment to the general-factor, it is interesting that he even included a chapter on these special scales.

Although discussing these special scales, McNemar warns against using the results from one or two Binet items in order to make judgments about special abilities like memory, visual imagery, or perception. However, McNemar's special scales represent an early attempt to organize the Stanford-Binet items into separate, meaningful categories. Of special interest, McNemar's three scales (vocabulary, non-verbal, and memory) correspond, at least vaguely, to three of the Reasoning Areas on the Stanford-Binet IV (respectively, Verbal Reasoning Area, Abstract / Visual Reasoning Area, and Short-Term Memory Area). Add to this Thorndike's (1975) praise of McNemar for generating multiple factors that have been more-or-less interpretable on the Stanford-Binet scales. Taken as a whole, McNemar's findings do seem to provide a justification for the new structure on the Stanford-Binet IV.

In 1964 McNemar suggested that neither the Stanford-Binet L-M nor the WISC provided a "factorial pure, unidimensional measure of a *g* " (p. 873). Referring to the Stanford-Binet L-M he made the following statement: "The current Stanford-Binet was in reality constructed too early to benefit from the implication of factor analysis for test purity, whereas the Wechsler scales were based on the impossible premise that 10 or 11 subtests can simultaneously provide diagnostic subscores and a meaningful total score" (p. 873). As pointed out by McNemar (1964), the older Binet scales as well as the WISC (and now the WISC-R) were not designed on the basis of factorial results. In contrast, the Stanford-Binet IV's theoretical underpinnings provide a factorial base against which this latest revision can be assessed. Indeed, a major issue addressed in this thesis is whether the design of the Stanford-Binet IV with four Reasoning Areas is supported by factor analytic findings.

Jones' Analysis of the 1937 Stanford-Binet

In analyzing the 1937 Stanford-Binet, Jones (1949) used the centroid method of factoring on a matrix of tetrachoric correlations. Factors were rotated graphically (presumably orthogonally) with attention being given to three criteria: (a) maximizing simple structure, (b) maintaining positive manifold, and (c) psychological meaningfulness of factors. Four age levels (seven, nine, eleven, and thirteen years) were included in the analysis, and each group had 200 children. The number of items included in the analysis at each age level was 30.

For the 7 year-olds, Jones extracted four factors. The first, accounting for 21.1% of the total variance, was a Verbal factor. A Reasoning factor emerged next, accounting for 18.3% of the total variance. The third and fourth factors were considerably smaller: Memory accounting for another 9.4% of the total variance, and Number accounting for 8.8% of the total variance.

At the 9 year-old level, Jones extracted five factors. Once again the first factor was Verbal in nature, accounting for 15% of the total variance. A Reasoning factor emerged next, accounting for an additional 13.7% of the total variance. In the same pattern as for the 7 year-olds, the third factor related to Memory, explaining 9.6% of the total variance. Fourth, a Spatial factor accounting for 8.9% of the total variance, was found. Finally, a "Residual factor" (3.1% of the total variance) was found.

In the 11 year-old group, the factors that emerged were similar to those found at the 7 and 9 year-old levels. Once again four factors were extracted. As expected, the first factor was Verbal in nature, accounting for 25.2% of the total variance. In an unexpected fashion a Memory factor emerged next, accounting for 11.5% of the variance. As in the 9 year-old group, a Spatial factor was found as the last meaningful factor, accounting for 12.2% of the total variance. Finally, another "Residual factor" (2.8% of the total variance) was found, though it had no apparent utility. Of interest, no reasoning factor appeared at this age level.

For the 13 year-old level Jones extracted six factors. Not surprisingly, the first factor to emerge was Verbal in nature, accounting for 19.4% of the total variance. The next two factors, Reasoning I and Reasoning II, were quite similar accounting for 11.8% and 6.4% of the total variance respectively. Fourth, a Memory factor emerged, accounting for 8.3% of the total variance. The final two factors both involved visual perception and were called Visualization and Spatial, respectively accounting for 7.8% and 7.0% of the total variance.

In 1954 Jones reanalyzed the data of the 13 year-old group from his 1949 sample. This time he used oblique rather than orthogonal rotations, and extracted ten rather than six factors[11]. A list of the ten factors found by Jones (1954) follows: Three Verbal Comprehension factors (supplying previously learned linguistic meanings, ability to manipulate words such that a meaningful relationship is imposed, and definition and manipulation of words and ideas); Memory for Disconnected Elements; Memory for Verbal Material; Visualization of Configurations; General Reasoning; Closure (the ability to fuse a perceptual field into a single percept); Carefulness (ability to accurately perceive details of spatial configurations); and finally a Residual factor. From these ten first-order factors Jones extracted three second-order factors: (a) Ability to Profit From Scholastic Experience, (b) Ability to Synthesize Perceptions and Perceive the Gestalt, and (c) Facility for Dealing with Relationships, both verbal and spatial. Given the complexity of Jones' second-order factors, many alternate interpretations could have been made; for example, the three factors could have been called (a) Verbal, (b) Visual/Spatial, and (c) Reasoning. Thus, Jones' 1954 analysis does not stand in direct contradiction to his 1949 analysis. It

11 Extraction of ten factors without an adequate criteria for deciding the number-of-factors to extract brings up the question of whether over-factoring occurred; when over-factoring occurs, the variance from the major factors is often splintered into very narrow factors which are not replicable.

does, however, demonstrate that interpretation of factorial results may be biased both by the number of factors extracted and by the type of rotation used.

Before looking at the general trends in Jones' (1949) work[12] it is important to note that Jones qualifies his findings by stating that many of the emergent factors were "not clearly defined" (p. 314). Indeed, the combination of item loadings on a given factors was often puzzling. It is also worth noting that a general factor was not extracted before rotations were conducted; in contrast, the first unrotated centroid for the different age levels accounted for 35-38% of the total variance and represented a general factor.

Thus, the overall pattern of factors found by Jones (1949) across the four age levels suggests that there are fewer latent factors for younger children than there are for older children. He therefore comes to the following conclusion: "The battery of items included at one age level of the Stanford-Binet is not factorially identical to the batteries of items at other age levels" (p. 318). However, some pattern differences may have resulted from the number of factors that were extracted. Despite Jones' suggestion of factorial variation across age levels, there does appear to be much similarity in the major group factors. At all age levels there was a Verbal factor, and a Memory factor; at almost all of the levels a Spatial factor and a Reasoning factor were also found.

Ramsey and Vane's Analysis of the 1960 Stanford-Binet L-M

Ramsey and Vane (1970) factored the Stanford-Binet L-M using the principal components method and the eigenvalue-one criteria for determining the number of components. Their sample included 152 children between the ages of three and seven years. Whether the first unrotated principal component was a general factor is debatable in that only two-thirds of the 18 items loaded on this component. The seven factors rotated by the varimax criterion are listed in the order that they were extracted: Verbal, Visual-Motor, an unnamed factor perhaps involving Persistence, Visual Ability and Judgment,

---

12 Jones' 1954 results will not be included here as they are somewhat prolematic and do not allow for comparison across the different age groups.

General Knowledge, Control of Impulsivity / Memory factor, and Visualization. The Ramsey and Vane study is interesting in that it appears to have been the first factor analytic study of the Stanford-Binet scales to use an objective criteria for deciding the number of factors to extract. However, a potential flaw in this study relates to the use of principal components factoring with items that have relatively low reliabilities (this problem is treated in more detail when Thompson's (1984) factor analysis of the Stanford-Binet L-M is reviewed).

Hallahan, Ball, and Payne's Analysis of the 1960 Stanford-Binet L-M

Hallahan , Ball, and Payne's (1973) factor analysis of the Stanford-Binet L-M was restricted to 363 preschool age children ranging in age from 3-0 to 5-8 years. These children were from culturally disadvantaged backgrounds and had been enrolled in the Head Start program. Because a short form of the Stanford-Binet L-M had been administered to the children, only 15 items were included in the factor analysis. Principal axes factoring was used and initial communalities were estimated by squared multiple correlations (R^2). Using the scree test criterion, three factors were extracted and rotated by the varimax criterion. In the order of extraction these three factors were General Knowledge, Visual-Motor, and Verbal. Although Hallahan et al. interpreted their first rotated factor as General Knowledge, it should be noted that only 12 of the 15 items loaded on this factor. Failure to find a strong general factor is not surprising in that rotation by the varimax criterion usually results in the loss of a general factor. However, though the nature of the first factor is not totally clear, it does seem to approximate a general factor or perhaps a reasoning factor.

Thompson's Analysis of the 1972 Stanford-Binet L-M

In her analysis of the 1972 renormed Stanford-Binet L-M Thompson used the principal components method of factoring; however, unlike previous factor analyses of the Stanford-Binet scales, she used phi correlations between items rather than tetrachoric correlations. Her sample was divided into two age-groups each with 110 subjects. Sixteen items were included in the analysis for the younger group while 22 items were included for the older group .

Compared to previous studies, the correlations between individual items were surprisingly low. For the younger group, these correlations ranged from -.22 to .64 with most being between .10 and .35. Similarly, for the older group the correlations ranged from -.11 to .58 with most being between .10 and .50. Thompson suggested that her use of phi rather than tetrachoric correlations probably explains why her correlations were lower than those usually reported. As she points out, tetrachoric correlations are designed to give estimates of what the product-moment correlation would be if the item was continuous even though in reality it is dichotomous. Thus Thompson suggests that tetrachoric correlations are often poor estimates (often inflated) of product-moment correlations when the variable is dichotomous. Her rationale for using phi correlations rather than tetrachoric appears to be sound in that each item from the Stanford-Binet is treated as a variable which is either passed or failed, and no partial credits are given.

An important point to remember in interpreting the results from Thompson's work is that she used a principal components analysis. As was mentioned briefly when considering McNemars (1942) work, the low reliabilities of individual items on the Stanford-Binet scales cause some problems because the communality of an item generally cannot exceed its reliability. Consequently, the low item reliabilities on the Stanford-Binet dictate that communalities will be fairly low and conversely that a fairly large amount of the item variance will be unique[13]. Hence the principal components method may not be the

13 That is, specific + error variance.

most appropriate to use in factoring the Stanford-Binet L-M because it assumes that all of the variance of the variable will be reliable; the principal factoring[14] method used by Hallahan et al. (1973) would probably have been a better choice.

Though one can question Thompson's (1984) use of the principal components rather than the principal factoring method, her results are still informative and probably reasonably accurate. For her younger group the first unrotated component seemed to be a general component, though loadings were between .29 and .63. The five rotated components that emerged were similar regardless of whether the varimax or oblique promax methods of rotation were used. She gave the following names to her factors: (a) Verbal / Nonverbal Reasoning, (b) Verbal, (c) Visualization / Visual Judgment, (d) Control of Impulsivity, and (e) Visual-Motor.

For the older group the first unrotated component again appeared to be a general component with loadings ranging from .22 to .75. Though six components had eigenvalues greater than one, the five component solution was more conducive to interpretation. In both the varimax and oblique promax rotated matrices the following five factors were named: (a) Verbal, (b) Verbal / Nonverbal Reasoning, (c) Difficulty, (d) Visualization, and (e) Meaningful / Nonmeaningful Auditory Memory.

Comparison of the components found by Thompson (1984) in her older and younger groups is worthwhile. First it should be noted that Control of Impulsivity from the younger group may alternately be considered as a Memory component (as indicated by loadings from repeating of 5 digits, and naming days of the week); using this alternate interpretation the similarity between the components in the two different age-groups is remarkable. For both age-groups, there is a Verbal factor, a Reasoning factor, a Visualization factor, and some type of Memory factor.

---

14 For the sake of brevity, the application of the common factor model to the principal axes method will be referred to as "principal factoring" to distinguish it from the principal components method.

Summary of the Factorial Studies on the 1916-1972 Stanford-Binet Scales

The importance of the factor analytic studies on the earlier Stanford-Binet scales is perhaps unclear to some. Why in a test designed to measure general intelligence have researchers persistently conducted factor analyses in order to find group factors? Furthermore, now that several researchers have demonstrated that group factors appear to exist on the Stanford-Binet scales, what are the implications? To answer the first question, the persistent attempts to find group factors on the Stanford-Binet are basically an attempt to aid in its interpretation and general utility. Though general intelligence may be one of the best predictors of scholastic achievement, development of specialized programs for children requires some idea of their specific strengths and weaknesses, not just an indication of their overall cognitive ability. Therefore, the rationale for searching for group factors follows from the belief that individuals may possess differential levels of ability on different areas that we consider as part of intelligence (e.g., Verbal and Nonverbal ability). Thus, in order to use the Stanford-Binet scales for designing educational plans, it has been desirable to find factors on which profiles of strengths and weaknesses can be assessed. In short, there was a need to find group factors on the Stanford-Binet.

Another reason for factoring the Stanford-Binet scales has a theoretical origin. Simply, is there really a general intelligence like that suggested by Spearman (1904, 1927), or is a model like Thurstone's (1938) more workable in that it suggests primary factors rather than a single general factor. In this respect, the previous review of the factor analytic studies of the Stanford-Binet is important in terms of theoretical implications for the Stanford-Binet IV. That is, is the fourth Edition really a Stanford-Binet type test or, as suggested by Vernon (1987), is it so different that it really should not be called a Stanford-Binet scale? This is an important question which deserves a brief comment. In the old Stanford-Binet scales we had a test that was designed to measure general intelligence. However, despite attempts by Terman and Merrill to construct a scale with a unitary focus, factor analytic studies have demonstrated the presence of group factors on the Stanford-

Binet. Perhaps then it is time to recognize the presence of group factors, and to move away from the idea that intelligence is unifocal. Although the Stanford-Binet Scales have indeed demonstrated that they contain a strong general factor, the presence of group factors on these scales does provide a precedent for developing Reasoning Areas on the Stanford-Binet IV.

As a basis for answering whether the Fourth Edition of the Stanford-Binet is really a Binet type test, it is imperative that one not impulsively conclude that it is not because it has broken from the concept of measuring only general intelligence. Rather, the thoughtful psychometrician will carefully weigh the structure of the Stanford-Binet IV against the factorial results that have emerged from the factoring of the various Stanford-Binet scales preceding the Fourth Edition.

From the review of the factorial studies covered in this section, some generalizations can be made about what should be expected in a Stanford-Binet scale. First, it should contain a strong general factor (Burt & John, 1942a, 1942b; Hallahan et al., 1973; Jones, 1949; McNemar, 1942; Wright, 1939). Second, several group factors will probably be present: a Verbal factor (Hallahan et al., 1973; Jones, 1949; McNemar, 1942; Ramsey & Vane, 1970; Thompson, 1984); a Visual-Spatial or Nonverbal factor (Hallahan et al., 1973; Jones, 1949; McNemar, 1942; Ramsey & Vane, 1970; Thompson, 1984); a Memory factor (Jones, 1949; McNemar 1942; Ramsey & Vane, 1970; Thompson, 1984); a Reasoning factor (Jones, 1949; Thompson, 1984); and a Number factor (Burt & John, 1942a, 1942b; Jones, 1949; Wright, 1939). Further, given the emergence of group factors, a model of intelligence that retains a general factor, and also allows group factors must be adopted to provide a theoretical explanation of the empirical findings from the factorial studies.

Should all of the factors found in previous factor analytic studies of the Binet scales be present on a new Binet type test? Probably not. The reason they all need not be present is complex. Because the factoring done in the early studies of the Stanford-Binet (Burt &

John, 1942a, 1942b; Jones, 1949, 1954; McNemar, 1942; Wright, 1939) lacked a formal method for deciding the number of factors to rotate, some overfactoring or underfactoring may have occurred; the implication is that some of these factors may have been too narrow or specific. Second, because each item on the Stanford-Binet is treated as a variable, their low reliabilities makes the principal components analysis potentially flawed (applicable to Ramsey & Vane's 1970 analysis and Thompson's 1984 analysis).

Another unaddressed but important issue regarding the factoring of the Stanford-Binet Scales is whether different factors are measured at different ages. Determining the answer to this question is problematic in that the entire scale could not be used in any analysis--because the 1916, 1937, 1960 and 1972 Stanford-Binets are basically an age scale, not all children received the same tests. In her overview of the factor analytic studies on the Stanford-Binet scales Thompson (1984) provides a quick tabular summary (Thompson, 1984, Table 2.1, pp. 54-55) of the factors found in most of the studies that have been reviewed in this section. Her main conclusion is that there may be several group factors embedded in the Stanford-Binet scales. However, a second conclusion is that the nature of the group factors may change with age. Hence, following the Stanford-Binet tradition it is possible that different factors may emerge at different age levels on the Stanford-Binet IV.

To summarize, previous factor analytic work on earlier Stanford-Binet scales suggests that in designing a new Stanford-Binet, the test constructors should retain a strong commitment to general intelligence, and secondarily a commitment to the presence of several group factors which have consistently been found on the older Stanford-Binet scales. In essence, Thorndike et al. (1986) have adhered to these criteria.

D. Profile Analysis of the Stanford-Binet Form L-M

Commenting on the transition between the second and third edition, Terman and Merrill state that "Attempts to construct profiles that are psychologically meaningful on tests designed to yield a single measure of general mental ability have been very discouraging" (1973, p. 13). Terman and Merrill further warned that attempts to group the items from the Stanford-Binet onto meaningful scales appeared to have little psychological justification, because of the way the the Binet scales had been created and honed. For all intents and purposes, the Stanford-Binet L-M is mainly unitary in nature--most strongly measuring general intellectual ability. Despite this, many attempts have been made to conduct profile analysis based on grouping of individual items from the Stanford-Binet.

The utility of the Stanford-Binet L-M in educational settings has been hampered by the lack of subtest/subscale scores on which profiles can be based. This lack of subtests has been a concern to those seeking to identify individual strengths and weaknesses of children. Faced with a choice of interpreting only the global IQ on the Stanford-Binet L-M or resorting to some classification scheme to facilitate interpretation of differential ability, most clinicians have chosen the latter. Fortunately, two prominent classification schemes have been developed, one by Valett and one by Sattler. The purpose of a classification scheme has been succinctly described by Sattler (1982): "A classification scheme is a convenient way of describing clusters of items in categories that have some face validity. Classification schemes aid in evaluating the child's performance, especially the pattern of successes and failures, and enable the examiner to make recommendations based on these patterns" (p. 133).

In the next section, these two schemes and their criticisms and proposed revisions will be considered. Although these classification schemes have not been derived from factor analytic work, the categories are very similar to the factors found in several studies, and they have proven to be very helpful in interpreting the Stanford-Binet L-M. Further, the prevalent use of profile analysis on the Stanford-Binet L-M warrants some understanding of how these profiles are constructed. The major utility of using a classification scheme is not the definitive identification of strengths and weaknesses but rather to generate hypotheses: "The classifications are to be used to generate hypotheses about the child's performance in order to make more meaningful recommendations; they are not to be used to make diagnoses about special abilities" (Sattler, 1982, p. 136).

The importance of looking at these classification schemes is twofold. First, it should alert the reader to the great need for categories by which the Stanford-Binet L-M can be meaningfully interpreted. Second, it will hopefully awaken in the reader an elementary understanding that factorially derived categories are superior to subjectively derived categories such as Valett's and Sattler's. Even more importantly, it is hoped that discussion of these schemes will provide the reader with a greater appreciation of why it was necessity to replace the Stanford-Binet L-M with a new Stanford-Binet that has not only four major group factors[15] but also subtests within these four factors.

Valett's Six Category Classification Scheme

Valett (1964) proposed a classification scheme with six categories: general comprehension, visual-motor ability, arithmetic reasoning, memory and concentration, vocabulary and verbal fluency, and judgment and reasoning. Placement of items into appropriate categories was done subjectively by "experts". Using these categories, a clinician can construct a profile which is useful in identifying children's strengths and weaknesses. Although some items from the Stanford-Binet appeared in more than one category, Valett defends this by suggesting that abilities overlap in "real" life as well.

Sattler's Seven Category Classification Scheme

Sattler (1965) also constructed his classification scheme on "somewhat arbitrary groupings according to item content" (p. 173). Sattler's scheme has seven major categories (with subcategories given here in parentheses): Language, Memory (Meaningful Memory,

[15] On the Stanford-Binet IV these group factors are referred to as Reasoning Areas.

Nonmeaningful Memory, and Visual Memory), Conceptual Thinking, Reasoning (Nonverbal Reasoning, and Verbal Reasoning), Numerical Reasoning, Visual-Motor, and Social Intelligence. Sattler refers to this classification schemes as a "Binetgram" (1982, p. 134). In constructing his scheme and analyzing the items at each level according to this scheme, Sattler suggests that his categories substantiate the suspicion that the Stanford-Binet L-M does not measure the same factors at each age. At the younger ages, stress is placed on the categories of visual-motor capacities, nonverbal reasoning, social intelligence, and language functions; during the middle years of childhood, more emphasis is placed on memory functions, and reasoning; and finally, during adolescence emphasis is placed on verbal reasoning (Sattler, 1965).

Comparison of Valett's and Sattler's Classification Schemes

Silverstein (1965) compared Valett's (1964) and Sattler's (1965) classification schemes. He found that several areas overlapped considerably (coefficient of overlap being as high as .85 to .89 in three areas). Thus, Silverstein concludes that "If the categories with the greatest amount of overlap are taken as equivalent, the schemes agree in classifying 75% of the total number of test items" (p. 964). Silverstein (1969) further examined Valett's and Sattler's classification schemes by applying the schemes to the Stanford-Binets of 80 retarded children. He calculated correlations between the different categories of the two schemes and again concluded that the schemes have a great deal in common. However, he quickly points out a hazard of using item-classification schemes: namely that the use of abstract categories like "Reasoning" may cause the examiner to lose sight of what individual items actually requir of examinees.

Sattler's SD Method of Determi Strengths and Weaknesses

Although the two classification schem ned above are useful heuristics, it is difficult to tell whether performance in one area ficantly better or worse than for another area. Because of this problem, Sattler (1982) devised a method that can be used to evaluate the scatter in a Binet profile. This procedure is used in connection with the examinee's CA and MA and is known as the standard deviation technique (SD). The SD method provides guidelines that help to determine whether successes or failures result from normal variability or from meaningful differences in abilities; thus the SD technique helps guard against overinterpretation of minor fluctuations in a child's performance.

The major assumption behind the SD method is that IQs that are less than one standard deviation above or below a score of 100 are normal, and represents the normal range of functioning. In short, the goal is to identify a band that surrounds an age level (MA or CA) for which one would expect normal variation. Strengths and weaknesses are then inferred from successes or failures outside of this band (Chase & Sattler, 1980). Once specific items are found to be strengths or weaknesses, they can be interpreted by either Sattler's (1965) or Valett's (1964) classification schemes, or can be used singularly to generate hypotheses about broader strengths and weaknesses (Kaufman & Waterstreet, 1978).

Because Sattler (1974) did not provide reference tables for use with the SD technique, Kaufman and Waterstreet (1978) made an attempt to correct this deficiency. They present a table that gives guidelines for both CAs and MAs in terms of normal variability using Sattler's SD method. Although Chase and Sattler (1980) agree with the basic methodology outlined by Kaufman and Waterstreet (1978), the former point out that CAs and MAs provided by the latter have been rounded to half-years, and therefore may give improper estimates. Also, at some age levels the Kaufman and Waterstreet tables do not accurately reflect the variability as outlined by Sattler. To correct these errors, Chase and Sattler (1980) provide a refinement of the Kaufman and Waterstreet table as does Sattler (1982). They also clarify that using the CA provides normative information while using the MA provides ipsative information as to strengths and weaknesses. Again, the warning is that information gained through the SD method should only be treated as a working hypothesis.

Although Sattler's (1974) SD method is a noble attempt to salvage information from the Stanford-Binet L-M, it is far more cumbersome than a similar method provided by Sattler (1982) for use with the WISC-R. To illustrate the superiority of multi-scale subtests (like the WISC-R or Stanford-Binet IV) for the purpose of profile analysis and planning of individual educational plans, Sattler's (1982) method for profile analysis of the WISC-R is briefly described. For the WISC-R, Sattler provides tables for determining when a subtest scaled score is significantly different from the mean scaled score for the appropriate scale (Verbal or Performance). Because the WISC-R is divided into subtests with a mean of 10 and standard deviation of 3, it is relatively easy to see which subtests are weak and which are strong. Further, the WISC-R manual lists the minimum differences between subtests that are needed for determination of significant differences. Also, for the Verbal and Performance scales, the manual provides information on the size of differences that are required in order to conclude that the two scales are significantly different.

As such, the WISC-R provides not only information that lends itself to profile analysis, but also provides statistical information useful in examining profiles (Evans & Richmond, 1976). In the same respect, the Stanford-Binet IV lends itself to profile analysis in that Areas have a mean of 100 and standard deviation of 16 while subtests have a mean of 50 and a standard deviation of 8. Furthermore, in the *Technical Manual* for the Stanford-Binet IV Thorndike et al. (1986) provide several tables which provide information need for determining which subtests and areas are significantly different from one another. This new feature on the Stanford-Binet IV represents a great improvement over the Stanford-Binet L-M in terms of usefulness for profile analysis.

Classification of the Stanford-Binet According to Guilford's SI Model

As outlined by Sattler (1982) it was Meeker in the late 1960s who developed an approach for classifying Stanford-Binet items in terms of Guilford's Structure of Intellect (SI) model. The belief was that each item could be classified according to the operation, content, and product that it measures in Guilford's SI model (see review of Guilford for a brief description of the SI model). However, when psychologists attempted to classify the Stanford-Binet items according to Guilford's SI model, there is considerable disagreement as to the appropriate category that an item belongs in. In fact, Sattler (1982) concludes that some of the categories have such poor agreement across different raters that this classification scheme is probably not justified.

Despite the problems with this SI classification system, some researchers like Silverstein (1974) have attempted to classify the items of the Stanford-Binet L-M according to the SI model. Though Silverstein suggested that 28 of Guilford's SI categories can be found on the Stanford-Binet L-M, only 14 of Guilford's categories were used in Silverstein's study. An important finding was that "Although conceptually the structure-of-intellect abilities are independent, the average *r* s were .89, .80 and .80 for the operation, content, and product categories, respectively" (Silverstein, 1974, p. 763). These high correlations suggest that the categories based on Guilford's SI model are probably not independent enough to be helpful in constructing profiles. Thus, for the Stanford-Binet L-M there appears to be little support for the use of a classification scheme based on Guilford's SI model.

Summary of Profile Analysis of the Stanford-Binet L-M

Indeed, as noted by Terman and Merrill (1973) attempts to construct profiles based on the Stanford-Binet L-M have been disappointing. Although classification schemes have been developed by Valett (1964) and by Sattler (1965), the clinician is still left with a web to untangle. Both of these schemes were constructed subjectively, and lack the empirical support that could be offered by more precise methods (e.g., classification of the WISC-R subtests according to factor analytic findings). Neither of these classification schemes allows the clinician to determine whether one category in the scheme is significantly different from another. Although Sattler's standard deviation technique has been useful in identifying individual items that are strengths or weaknesses, the clinician still needs some type of scheme for identifying what each item measures. Further, because individual items do not have high reliability, interpreting a single item may be unwise, and therefore groupings of items that are strengths or weaknesses must be checked by some classification scheme such as Valett's or Sattler's. Overall, profile analysis based on the Stanford-Binet L-M has proven to be cumbersome.

In comparison to the Stanford-Binet L-M, the WISC-R has proven extremely useful for constructing profiles. To a large extent, this is probably one of the biggest reasons why the WISC-R has overshadowed the Stanford-Binet L-M in recent years. Fortunately, the structure of the Stanford-Binet IV is now designed to facilitate profile analysis, and important statistical differences between subtest and Area scores are available in the *Technical Manual* (Thorndike et al., 1986). The key question that needs to be addressed is whether or not the hypothesized Areas and corresponding subtests on the Stanford-Binet IV are substantiated through factor analysis and hence are viable in developing profiles.

Correlations Between the Stanford-Binet Form L-M and the WISC-R

In general, the correlations between the Stanford-Binet L-M and the WISC-R were moderately strong. In the manual for the WISC-R Wechsler (1974) reports that for four age groups, the mean coefficients of correlation of the WISC-R Verbal, Performance, and Full Scale IQs with the Stanford-Binet IQ were .71, .60, and .73 respectively. Covin and Sattler (1985) did a short review of the literature comparing the correlations between the Stanford-Binet and the WISC-R; they found that reported correlations between the Stanford-Binet IQ and WISC-R Full Scale IQ ranged from .58 to .95 with a median of .81.

In order to show some interesting relationships between the WISC-R and the Stanford-Binet, three correlational studies will be reviewed. The first, by Sewell and Manni (1977), is of interest because it shows that for normal children the correlation between the WISC-R Verbal IQ and the Stanford-Binet IQ is higher than the correlation between the WISC-R Performance IQ and the Stanford-Binet IQ. Sewell and Manni's sample included 106 children between the ages of 6-0 and 16-10 years. The correlations between the Stanford-Binet L-M IQ and the WISC-R Verbal, Performance, and Full Scale

IQs were .86, .71, and .86 respectively. While there was no significant difference between the mean Stanford-Binet IQ and either of the WISC-R Verbal IQ (VIQ) or the WISC-R Full Scale IQ (FSIQ), there was a significant difference between the mean Stanford-Binet IQ and the mean WISC-R Performance IQ (PIQ). The implication is that the Stanford-Binet L-M may be quite verbal in nature.

A second study, by Covin and Sattler (1985) is of interest because it found that for retarded children, the WISC-R PIQ was more closely related to the Stanford-Binet IQ than was the WISC-R VIQ. Each child was assessed with the Stanford-Binet L-M and three years later reassessed with the WISC-R. These 147 children ranged in age from 9 to 16 years, were from homes with low socioeconomic status, and were labelled as trainable or educable mentally retarded. With this methodology and time span, the correlations between the Stanford-Binet L-M and the WISC-R VIQ, PIQ, and FSIQ we [illegible] .56, and .60 respectively. Thus, for children with fairly low ability, it appears [illegible] WISC-R Performance scale correlates more strongly with the Stanford-Binet L-M than does the Verbal Scale. However, one must be careful in making this statement because design of the Covin and Sattler study allows for possible confounding of time and instrument used. A better design would have been to assess the children with both tests at the same time and then reassess them with both tests after a period of three years.

Finally, the third study by Bloom, Reese, Altshuler, Meckler, and Raskin (1983) points out that for children with developmental problems, the relationship between the Stanford-Binet L-M IQ and the WISC-R PIQ and VIQ is not always straightforward. Bloom et al. administered both the WISC-R and Stanford-Binet L-M to a sample of 121 children with developmental problems. Of these children, 23% had a difference of 12 or more points between their WISC-R FSIQ and their Stanford-Binet L-M IQ. In fact, the mean Stanford-Binet IQ was 94.2 while the mean WISC-R FSIQ was 79.2. This discrepancy is interesting in that the WISC-R VIQ was usually similar to the Stanford-Binet IQ, while the WISC-R PIQ was usually, though not always, lower. The implication is that children in special populations may have different levels of ability in different areas such as the Performance and Verbal realm. This difference is not seen when the Stanford-Binet L-M is used because it is a highly verbal test. The new Stanford-Binet IV, on the other hand, provides IQs for four major areas, and would hence appear to have potentially more importance for assessing children from special groups in that it can measure strengths or weaknesses in each of these four areas.

The Need for Tests Useful in Constructing Educational Plans

With the introduction of laws ensuring the right of public education for all persons including handicapped children, individualized educational plans have become mandatory. In the United States, Public Law 94-142 has had a profound effect. As quoted in Sattler (1982), part of this law states that educational agencies should insure that : "Tests and other evaluation materials include those tailored to assess specific areas of educational need and not merely those which are designed to provide a single general intelligence quotient" (p. 518). Thus, Public Law 94-142 in the United States makes in imperative that the Stanford-Binet L-M not be used in evaluating handicapped children because it is designed to provide only a single IQ score.

Furthermore, because trends in education had been moving towards individual educational plans, the global score provided by the Binet Form L-M does not provide enough specific information on areas such as language, comprehension, visual-motor ability, reasoning, memory, attention and concentration, and social awareness that are important for educators and clinicians. Despite attempts to adapt the Stanford-Binet to profile analysis (Covin & Sattler, 1985; Kaufman and Waterstreet, 1978; Sattler, 1965, 1982; Valett, 1964), the WISC-R has a distinct advantage in that its design easily lends itself to profile analysis. Given this limitation in the design of the Stanford-Binet L-M, its questionable and outdated norms, and factorial research done on it, the time was right for a change in the structure of the Stanford-Binet scale to make it more useful to educators and clinicians.

The ultimate decision of a test's utility must be judged on its usefulness in educating children: "Binet's test, like others used in education, must be judged in terms of its ability to facilitate constructive adaptions of educational programs for individuals. This is the challenge for the next 70 years" (Thorndike, 1975, p. 7). Clearly then, the Stanford-Binet's decline in popularity was largely a result of its structural design. Busy clinicians found the WISC-R a better researched and more easily interpretable instrument. Not surprisingly, the Stanford-Binet IV was destined to take on a new form--a form not unlike the WISC-R--one that is better suited to educational planning.

Criticisms of the Binet Scales and the Need For Tests with a Theoretical Basis

Dinnan (1977) gives a glowing analysis of Binet's contribution to the testing movement, and labels him the "grandfather of all testing utilized in psychoeducational evaluation" (p. 272). Dinnan's major complaint against the older Binet tests is that they were not based on a sound theory of psychological principals. For example, Binet did not claim that the items he used were unitary in nature; he was looking for a broad sampling of items that would solve a pragmatic problem. Later revisions of the Binet scales at Stanford University also tended to be more pragmatic than theoretical, though there was a strong commitment to measuring general intelligence. However, factor analyses on the Stanford-Binet scales provided information that group factors might also exist on these scales.

Another often quoted criticism of the Stanford-Binet is made by a reviewer in The Seventh Mental Measurements Yearbook: "The Stanford-Binet Intelligence Scale is an old, old vehicle. It has led a distinguished life as a pioneer in the bootstrap operation that is the assessment enterprise. Its time is just about over" (Friedes, 1972, p. 773). Despite advancement in theories of intelligence (e.g., Cattell, 1971; Gustafsson, 1984; Vernon, 1961), the Stanford-Binet L-M remained inert to the changing times, clinging as it were to a sinking ship--with occasional attempts to patch up major structural defects.

Davis and Rowland (1974) list several pragmatic criticisms of the Stanford-Binet Form L-M. First, they suggest that there are too many objects and too many kinds of items for examiners to handle conveniently. Second, because each item type occurs only once at each level, it is "taxing" for the examiner to maintain the correct sequence. Third, the layout of the manual with separate sections for administration and scoring is troublesome. Finally, children are often subjected to a large number of failures before a ceiling is reached.

Waddell (1980) provides two further criticisms regarding the norming and validity of the Stanford-Binet L-M. One is the lack of new reliabilities and standard errors of measurement for the 1972 norms. Another weakness is the lack of validity reports; that is, does the Stanford-Binet L-M measures what it intends to with a 1980's population. Waddell notes that there has been a lack of studies assessing the predictive validity of the 1972 norms edition. Referring to the technical data available for the 1972 norms, Waddell boldly states that "It is contended that the Stanford-Binet should not be used in important decision making processes unless this weakness is corrected" (p. 203). Holroyd and Bickley (1976) come to an almost identical conclusion: "Despite its superiority over the 1960 norms, the inadequacy of the 1972 restandardization throws into doubt the advisability of continuing to use the Stanford Binet at all" (p. 104). Briefly, despite attempts to patch up the Stanford-Binet L-M, clinicians have become wary of its usefulness.

According to Hollinger and Kosek (1986) tests like the Stanford-Binet L-M which provide only a global measure of intelligence are antiquated and undesirable. The findings from their study give some credence to the rationale behind Public Law 94-142: unitary tests may not adequately capture differences in potential on differing areas of intelligence. Hollinger and Kosek's sample included 26 children between the ages of 6 and 15 years. All of these children were being assessed for placement in programs for the gifted (IQ ≥ 130). Despite their high IQs, 96% of the children had at least one WISC-R subtest score that was in the normatively average range. Appropriately, Hollinger and Kosek's conclusion and warning was that the measurement of intelligence with tests designed to provide only a global score may results in a lack of ability to differentiate unique patterns of cognitive abilities. For example, the WISC-R PIQs were more likely than the VIQs to be in the normative range; that is, these gifted children were more likely to demonstrate superior levels of performance on subtests requiring verbal comprehension and expression (i.e., subtests of Vocabulary, Similarities, and Comprehension); conversely, they were more likely to demonstrate only average ability on subtests requiring simultaneous processing, visual orientation, and distinguishing of essential from nonessential elements. Thus, using the Stanford-Binet L-M on children with special ability may fail to adequately sample and identify strengths and weaknesses.

With the development of newer and more theoretical based tests, the antiquated Stanford-Binet L-M was losing out. A good explanation of why this is happening is offered by McCallum, Karnes, and Edwards (1984). McCallum et al. compared the K-ABC, WISC-R, and Stanford-Binet for the use with 41 gifted children ranging from 9.5 to 12.5 years of age. In accordance with the recent shift towards more "process-oriented" tests and those with a strong theoretical basis, McCallum et al. suggest that the Stanford-Binet L-M is not the most appropriate test for assessing children in special populations. They suggest that the reason for this is that: "Apparently test results based on a theoretical model are perceived as more conducive to subsequent educational intervention than are the more traditional product-oriented results (p. 57). Kieth (1987) also agrees that one of the pervasive problems in psychological assessment for schools has been the absence of theory guiding research design and interpretation of results. He applauds the development of the Stanford-Binet IV for its contribution towards this goal.

Changing trends in the development of tests has also been noted by Sternberg (1986). He suggests that there has been a move towards designing new tests that are more closely tied to theories of intelligence. Not only this, but he also notes that the present trend is for new tests to measure "a broader set of abilities underlying intelligence" (p. 20). However, he also warns that test designers needs to be careful in not broadening their definition of intelligence so much that they "include everything but the kitchen sink" (p. 20). Thus, the new structure of the Stanford-Binet IV follows current trends, measures four broad Areas, and flows from a well respected theoretical basis.

Hargreaves (1974) provides an overview of the psychometric tradition and this provides an excellent rationale for the designing of the Stanford-Binet IV around a theoretical model. He notes that the pioneering work in psychometrics was largely based upon "technological rather than theoretical advance" (p. 27). He suggests that even the work of Vernon and Burt was based on factor analysis of existing data rather than being guided by theoretical perspectives. To some extent this may have been true. When little is know about an area, it is best to work from the data and then to develop theories to fit the data. However, once the theories have been refined, they can begin to guide the construction of tests to a greater degree. This is, to some extent, what has happened on the Stanford-Binet IV. Cattell's and Vernon's theories of intelligence have been used to determine how the test should be structured, and what areas should be included.

In summary, researchers and clinicians have found the Stanford-Binet L-M to be antiquated and lacking a solid theoretical base by which results could be interpreted. Also, it has not proven to be directly amenable to assessment of differential abilities in children. Though the Stanford-Binet L-M was a formidable instrument in its time, the time has come for it to retire and let its offspring, a new and theoretically appealing Stanford-Binet IV, replace it.

E. Theoretical Models of Intelligence: Implications and Background For The Stanford-Binet IV

The following section reviews several pertinent theories of intelligence which provide justification for the structure of the Stanford-Binet IV. To understand and appreciate the structure of this new Binet scale a basic understanding of these theories is needed. No attempt will be made to provide a comprehensive review of all theories of intelligence; rather, the primary focus will be on theories of intelligence which have been derived or supported by the use of factor analysis.

In the last 80 years several theorist have made major contributions to the field of intelligence and intelligence testing: Spearman, Thurstone, Vernon, Cattell, Horn, Jensen, Guilford, and Sternberg to mention but a few. There are many different ways that these theorists can be grouped. One way is to group those that subscribe to a general intellectual factor as opposed to those that do not allow for such a factor. Another way is to examine whether the theorist has a hierarchical model of intelligence or not. Yet another way is to determine whether the theorist follows a psychometric tradition or a cognitive perspective. Although the purpose of this review is not to categorize the relevant theories, four implicit areas or categories will be examined. First, the hierarchical models proposed by Cattell and Horn, and by Vernon will be examined; next, the controversy over the existence of a general factor (*g*) will be examined via the work of Spearman, Thurstone, and others; third, the information processing theories of Guilford and Sternberg will be examined because they have implications for factorial models; finally, the HILI model proposed by Gustafsson (1984) will be examined as it represents a synthesis of factorial studies. These implicit categories will hopefully aid the reader in organizing his or her thoughts, and will allow for some meaningful comparisons.

In the opening chapter of the *Technical Manual* for the Stanford-Binet IV (Thorndike et al., 1986), it quickly becomes apparent that this new Binet scale places more emphasis upon theoretical models of intelligence than any of its predecessors. The theories of Cattell and Vernon have played an important role in shaping the structure of the

Stanford-Binet IV. To illustrate the contributions of these authors, the theoretical structure of the Stanford-Binet IV (Thorndike et al., 1986) is presented in Figure 1. Adequate understanding of the theoretical structure of the Stanford-Binet IV is an essential step in helping to interpret factorial results from this test. Hence, a relatively lengthy review of the work of Cattell and Vernon is needed and will be covered next.

Figure 1

Theoretical Model of Intelligence as Presented in the Fourth Edition of the Stanford-Binet

Figure 1 has been removed because of the unavailability of copyright permission.

* Figure 1 has been reproduced from the *Technical Manual* for the Stanford-Binet IV (Thorndike, Hagen, & Sattler, 1986).

Cattell and Horn's Theory of Fluid and Crystallized Intelligence

Although Cattell (1971) admits that the Gf-Gc[16] theory does not rest solely on the results of factor analysis, the latter is generally taken as the main support for the theory. It should be noted that the factorial work by Cattell (1971) and Cattell and Horn[17] (1966, 1982) used principal axes factoring with iteration for communalities, and oblique rotations that were done graphically. Through the years Cattell's theory of intelligence has been modified somewhat. However, the most commonly known version of the theory represents the earliest of his work, and postulates two types of intelligence, Fluid (Gf) and Crystallized (Gc). In the*Technical Manual* for the Stanford-Binet IV (Thorndike et al., 1986), the early version of Cattell's theory seems to be espoused--that is the Gf-Gc theory; later revisions by Cattell and Horn will be considered after the presentation of this simpler Gf-Gc model.

Horn and Cattell (1966) identify Fluid Intelligence as the innate-biological, or constitutional element of intelligence: "the major measurable outcome of the influence of biological factors on intellectual development--that is, heredity, injury to the central nervous system (CNS) or to basic sensory structures, etc." (p. 254). Typically, Fluid Intelligence is measured by any type of item that requires analytic ability, perception of relationships, or educing of correlates; this includes series, classifications, analogies, matrices, reasoning, and also items requiring immediate awareness and memory (Horn & Cattell, 1966). However, Gf is not only present in tasks requiring figural and symbolic materials, but can also be present in tasks with semantic content. For example, when verbal synonyms or

16 Depending onthe publication, Fluid Intelligence is alternately symbolized by Gf or g_f; similarly Crystallized Intelligence is alternately symbolized as Gc or g_c. For the purposes of this thesis, the uppercase symbolization will generally be used, though in certain cases the lower case subscripted symbols will be used. The uppercase symbolization generally is used in the American tradition, while the lower case symbolization comes from the British tradition following Spearman's designation of general intelligence as *g*.

17 Although Horn was an active participant in the work done by Cattell, his name is often omitted and Cattell is given the primary credit for developing this theory of intelligence. As such, when reference is made to Cattell's theory in this thesis, the contribution of Horn is also assumed.

analogies are presented and the words are chosen to be easily within the vocabulary of the group tested then Gf may be required (Cattell, 1971).

Whereas Fluid Intelligence is closely tied to innate potential, Crystallized Intelligence, on the other hand, relates to knowledge that has been acquired through learning. Horn and Cattell (1966) further clarify the nature of Gc: "Crystallized intelligence, representing similar processes of perceiving relations, educing correlates, etc., in speeded and unspeeded tasks involving various kinds of content, but tasks requiring considerable pretraining to acquire techniques representing the accumulated wisdom of a culture" (p. 268). Usually, Gc is demonstrated on task requiring verbal ability, numerical ability, and mechanical information and skills. However, Gc is also involved in areas where reasoning, judgment, and discrimination have been systematically taught or experienced (Cattell, 1971). In a more current paper, Cattell and Horn (1978) give a simplified definition of Gc that ties it fairly closely to achievement: "Crystallized intelligence (Gc) is similar to achievement in that it is the accumulated knowledge of an individual" (p. 140). Thus, any knowledge that is taught or learned generally represents Crystallized Intelligence rather than Fluid Intelligence.

Another way of conceiving of the difference between Gf and Gc has been suggested by Nicholls, Patashnick, and Mettetal (1986): "First, the concept of fluid intelligence appears to embody a more general, higher-level skill than does the concept of crystallized intelligence, in which the contents of memory play a larger role" (Nicholls, Patashnick, & Mettetal, 1986, p. 637). This implies that Gf requires ability to deal with more conceptual matters while Gc is found on tasks requiring more passive remembering of facts, or the application of learned knowledge.

Despite the distinction between Gf and Gc made by Cattell and Horn, there is a considerable overlap between these two areas of intelligence. Particularly interesting is the fact that although Crystallized Intelligence does not enter into tasks that require Fluid Intelligence, Fluid ability does enter into some tasks (though to a lesser degree than Gc) that are verbal and numerical in nature. Cattell (1971) suggest the following reason for this finding: "In other words *purely learned* judgmental skills are not enough, even in the *traditional* intelligence tests, to enable such problems to be solved. Some fluidity of relation eduction is needed and some adaptability to new situations is demanded, even when using acquired judgments" (p. 95, emphasis in the original). Furthermore, Cattell (1971) himself admits that Gf and Gc partially overlap. In fact, he reports that the correlations between the two generally run between .4 and .5 for all ages. To explain this finding he suggests that the acquisition of Crystallized Intelligence "depends *partly* on the level of insightful 'fluid' ability and *partly* on hours spent in school" (Cattell, 1971, p. 100).

The origin of Crystallized Intelligence is further elaborated by Cattell (1971). He suggests that "crystallized ability is a product over time of earlier fluid ability action" (p. 79). It is at this point that Cattell and Horn's earlier work (1966) begins to flounder. Because of the moderate correlations between Gf and Gc, Cattell has attempted to find higher-order factors. He asserts that the two main general factors (Gf andGc) do not reliably come together into a single factor on a third-stratum analysis; however, the third-stratum factor which he labels "$g_{f(h)}$" does emerge, and Gf loads more heavily on it (.80) than does Gc (.60) (Cattell, 1971, p. 117). In explaining $g_{f(h)}$ Cattell implies that it is a "historical" antecedent that one is born with. From this Gf and Gc emerge. In simplified form, this means that Gc is a result of earlier influences of Gf. In Cattell's (1971) own words,

> we are going to argue that the fluid ability factor typically found at the second stratum, and which can be estimated from the individual's present scores on the primary abilities, is his *present* fluid intelligence level, but that the single ability appearing at the third (or fourth) stratum and loading both second-stratum g_f and g_c is the fluid ability of yesteryear, which

fathered the present fluid ability directly and begot the present crystallized ability out of past experience (p. 119).

This finding of a higher-order factor tends to substantiate a general factor at the apex of the hierarchy, much like proposed by Vernon (1961).

Of interest to this thesis, Cattell and Horn (1978) suggested that the development of Gc follows a developmental trend. That is, one is born with an innate ability (whether it be Gf or $g_{f(h)}$) from which Gc develops because of the uniformity of the cultural learning environment in most of the English speaking world. This developmental trend is clearly elaborated by Horn and Cattell (1966): "It follows that the distinction between Gf and Gc will be difficult to draw in early childhood, will become gradually more clear-cut as development proceeds through childhood and will become most evident in adulthood" (1966, p.259). Furthermore, though perceptual abilities (Gf) tend to decline after age 25, verbal and numerical abilities (Gc) tend to show a slight but steady rise throughout most of life. Thus, based on the theory of Gf and Gc, it should be expected that the influence of a general factor will be greater for younger children, while older children will have had a greater influence from cultural-educational learning and hence Gc will be stronger than at the younger ages.

The preceding discussion of Cattell and Horn's theory in terms of Gf and Gc has proven to be a slight oversimplification. Although Fluid and Crystallized Intelligence are the most prominent factors that have been found in this stream of research, additional factors have emerged fairly consistently, depending on the types of tests that were included in the analyses. For example, Cattell (1971) suggested that besides Gf and Gc, other general factors exist: Visualization (Gv or sometimes labelled p_v), Speed of Cognitive Processing (Gs), and General Fluency / Retrieval From Memory Storage (Gr). Cattell's conclusion about the number of general factors is as follows: "To return to our metaphor, there is not one vast mountain range, or even two (g_f and g_c), but several great ranges in the domain of cognitive effectiveness" (p. 109). What started out as a possible theory of two general factors to replace Spearman's one had mushroomed into several "g's" which began to look like group factors rather than general factors (See Undheim's 1982a, 1982b work presented later).

Cattell calls the latest version of his theory a triadic one: "three classes of abilities (1) the primary abilities or agencies--the a's; (2) the provincial neural-experimental organizations, visual, auditory, etc.--the p's, and (3) the general capacities--the g's" (Cattell, 1971, p. 163). In essence, this "triadic" theory absorbed his earlier g_f - g_c theory with the former being more general than the later (Carroll, 1984). Unfortunately, the triadic theory is not totally clear as to which level of the stratum the p's as opposed to the g's should emerge. That is, sometimes the p's and g's are found at the same level of the stratum even though they presumably do not have the same degree of broadness.

Because of the difficulties involved in the triadic theory mentioned above, it is probably better to conceptualize Cattell and Horn's various factors as group factors. In Horn and Cattell's 1966 paper a revised theory including Fluid, Crystallized, and three further factors is outlined. Because these three additional factors (Gv, Gs, and Gr) are often not considered when the Cattell-Horn theory is described, they will be dealt with more thoroughly in the three following paragraphs. Because Horn and Cattell do an admirable job of describing the processes involved in these three factors, rather extensive quotations will be used so that the exact flavor of these factors can be reported.

First, General visualization (Gv) is suggested as a process involving visual acuity, depth perception, breadth and depth of the visual field, transformation of spatial patterns, and location of elements in a visual field. Horn and Cattell (1966) provide the following definition of the processes involved in this factor: "General visualization, representing processes of imagining the way objects may change as they move in space, maintaining orientation with respect to objects in space, keeping configurations in mind, finding the gestalt among disparate parts in a visual field, and maintaining a flexibility concerning other possible structuring of elements in space" (p. 268).

Second, General fluency (Gr) involves recall and recognition of cultural concepts usually requiring the use of conceptual labels. This factor is described by Horn and Cattell (1966) in the following manner: "Facility in the use of concept labels, an ability to quickly bring words (i.e., concept labels) from long-term memory into immediate awareness, a facility which the evidence of this study suggests is largely independent of comprehension of the subtlety of the concepts themselves, as indicated in Gc" (p. 269). As noted previously, Cattell (1971) alternately refers to this factor as Retrieval From Memory Storage.

Third, General Speediness (Gs) involves speed of performance as opposed to capacity to perceive relationships. Put in a different way, "General Speediness, an attribute measured in simple writing and checking tasks requiring little in the way of complex relation-perceiving, but an attribute producing variance in the measure of most intellectual factors" (Horn & Cattell, 1966, p. 268).

The existence of these three factors in addition to Gf and Gc has been recorded through the work of Undheim (1976, 1981a, 1981b). Because these factors are part of the fuller model proposed by Horn and Cattell, and because one or more of these factors may emerge in the factoring of the Stanford-Binet IV, Undheim's work will be briefly reviewed.

Undheim (1976) examined the abilities of 144 fourth-grade children in Norway. The twelve subtests from a preliminary version of the "Norwegian Wechsler Intelligence Scale for Children were included in the test battery along with 23 group-administered tests. Nine factors were extracted even though only six had eigenvalues greater than one. Principal axes factoring with communalities iterated to convergence was used. Rotations were done graphically as well as with the promax and varimax criteria. Unfortunately, Undheim's (1976) work is handicapped by the fact that he deleted several variables that he felt they would not clearly define primary factors, including many of the WISC-R subtests (leaving a total of 18 variables). This process of selecting tests based on theoretical and previous factor results may have strongly biased the factors that emerged. However, Undheim's (1976) results are of interest in that four of Horn and Cattell's (1966) five factors were confirmed. The four factors were interpreted by Undheim to be Broad Speediness (Gs), Broad Visualization (Gv), Fluid Intelligence (Gf), and a Crystallized / Verbal-Educational Intelligence (Gc).

Encouraged by his previous findings, Undheim (1981a) designed another study in which he hoped to distinguish whether the measurement and definition of Gf and Gc was distinct from other broad factors of intelligence like Broad Visualization, Cognitive Speed, and Broad Fluency. Twenty-one tests were administered to 148 eighth- and ninth-grade children in Norway. Five factors were found to have eigenvalues greater or equal to one. These five factors were interpreted as Fluid Intelligence (Gf), Crystallized Intelligence (Gc), Broad Visualization (Gv), Speediness (Gs), and Fluency (Gr).

In another publication in the same year, Undheim (1981b) looked at the hierarchical nature of Horn and Cattell's theory. Undheim proposes that Gf is really equivalent to Spearman's conception of *g* , whereas Gc probably represent a broad verbal-educational factor. To test this hypothesis, Undheim reanalysed several previously published works. Using a hierarchical analysis, he concluded that Gf is indeed similar to Spearman's *g* . Further, three group factors were present representing Crystallized Intelligence (or alternately called verbal-educational intelligence), Visualization, and Speediness. From this analysis Undheim concludes that the twin peaks of Gf and Gc as initially proposed by Cattell are not necessarily substantiated. Rather he suggests that the results are more consistent, and more parsimonious if explained with a "neo-Spearman structuring of broad intelligence factors" (p. 185). However, because of Undheim's (1981b) analysis was not based on a new study that was well designed, his findings and conclusions must be regarded as tentative.

In summary, Cattell and Horns theory of Fluid and Crystallized Intelligence has been a fairly popular depiction of intelligence. Initially, Cattell (1971) suggested that Spearman's concept of *g* was erroneous, and that there were actually two types of general intelligence (Gf and Gc). However, as early as 1966 Horn and Cattell found it necessary to revise their theory to include three additional factors: General Speediness (Gs), Broad Visualization (Gv), and General Fluency (Gr)--though Fluid and Crystallized Intelligence were usually the most stable and largest factors. A glance at Figure 1 illustrates how Thorndike et al. (1986) have incorporated Cattell and Horns Gf and Gc into a hierarchical model of intelligence. It is noteworthy that Thorndike et al (1986) have demoted Gf and Gc to the level of group factors. This has some merit considering Horn and Cattell's (1966) suggestion that Gf and Gc are possibly accompanied by Gv, Gs, and Gr at the same level in the hierarchy. The retention of a general factor (*g*) at the apex of the hierarchy corresponds partially to Cattell's (1971) conception of $g_{f(h)}$, but in reality, is more akin to Vernon's staunch adherence to the presence of *g* at the apex of the hierarchy or to Undheim's (1982b) work.

Vernon's Hierarchical Model of Intelligence

Like Cattell and Horn, Vernon maintains that intelligence is hierarchical, with a general ability at the top, with group factors under this, and with specific factors under the group factors. Following the British tradition, Vernon suggested that group factors not be given names, but rather be identified as symbols such as *v* for verbal, and *k* for spatial; he suggested that using symbols rather than names helps avoid misinterpretation. Vernon also warned that factors be regarded "primarily as categories for classifying mental or behavioural performances, rather than as entities in the mind or nervous system" (1961, p. 8). Despite these suggestions, names are usually applied to the factors because doing so helps clinicians to understand the nature of the factors.

Based on his work with Army and Navy recruits in the 1940s, Vernon gradually developed his model of intelligence. He found that in most analyses *g* accounted for more than twice as much variance as all group factors combined. These analyses led Vernon (1961) to the conclusion that intelligence is hierarchical: "After the removal of *g*, tests tend to fall into two main groups: the verbal-numerical-educational on the one hand (referred to as *v:ed* factor), and the practical-mechanical-spatial-physical on the other hand (referred to as *k:m* factor)" (p. 23).

Vernon goes on to specify that the *v:ed* and *k:m* factors can be further broken into minor group factors. Lest one be carried away, Vernon (1961) warns that "the strict hierarchical picture of mental structure is an over-simplification. For the results of any factor analysis depend largely on the composition of the population tested (e.g. its degree of selection), and on the number and kind of tests studied" (p. 25). However, based on the results from much research, Vernon (1961) suggested that his model of intelligence is fairly stable. This model is presented in Figure 2 so that it can be compared to the theoretical structure of the Stanford-Binet IV presented in Figure 1. Vernon's main contribution to the theoretical Structure of the Stanford-Binet IV can be seen in the hierarchical or three tiered conceptualization of intelligence with *g* at the apex.

In a later publication Vernon (1965) adeptly summarized his theory:

> After removing the general factor (whether by group-factor technique or by rotation of centroid factors), the positive residual correlations always fall into two main groups--the verbal-educational (*v:ed*) group and the spatial-practical-mechanical group. The *v:ed* factor usually yields additional minor fluency and divergent thinking abilities--scholastic and*n* or number of subfactors. Likewise, the *k:m* complex includes perceptual, physical, and psychomotor, as well as spatial and mechanical factors, which can be further subdivided by more detailed testing. In addition there seem to be various cross-links: For example clerical tests usually combine verbal

ability and perceptual speed, p ; likewise math and science depend both on number and spatial abilities, n and k. Sometimes an inductive reasoning ability (also very relevant to science) can be distinguished, though most of the common variance of reasoning tests is apt to be absorbed into g . At a still lower level in the hierarchy comes what are usually referred to as specific factors, though of course any specific can be turned into an additional narrow group factor by devising additional tests" (Vernon, 1965, p. 725).

Figure 2

Vernon's Hierarchical Structure of Human Abilities

Figure 2 has been removed because of the unavailability of copyright permission.

* Diagram taken from Vernon (1961)

Regarding group factors, Vernon (1961) stated that "nevertheless there is ample evidence to support the view that group factors are almost infinitely subdivisible, depending only on the degree of detail to which the analysis is carried" (p. 26). He goes on to boldly state that "the only truly specific element is the unreliability or error variance of the test. Thus in a complete factorial investigation the communality of each test should approximate to its reliability coefficient" (p. 26). He also warns that there is "no absolute distinction" between a general or specific factors (p. 26). Whether a factor emerges as more specific or more general may depend upon the other variables that are included in the analysis. Rather than talk about whether a latent construct is a group or specific factor, Vernon prefers to make a distinction between "broad" and "narrow" group factors. He suggests that to be considered a broad factor, a factor should account for at least 5 percent of the variance of some educational or occupational criterion; a narrow factor will contribute less than 5 percent to the variance.

Vernon further clarifies the relative importance that each level in the hierarchy plays. First, the hierarchical model suggests that the largest amount of variance in human abilities is attributable to the general factor at the apex of the hierarchy. Surprisingly, Vernon (1961) attributes little influence to the group factors:

> The hierarchical group-factor viewpoint implies that most of the variance of human abilities in daily life is attributable to *g* and to highly specific (or very small group) factors, and that the role of the broader group factors is rather meagre. If our diagram could be worked out completely to cover all human abilities, the *g* -variance might amount to about 40 per cent., the major and minor group factors to some 10 per cent. each, and the remaining 40 per cent. would consist of very narrow group factors and unreliability" (pp. 27-28).

The implication from this statement is not totally clear. However, it seems that Vernon is strongly advocating the use of general intelligence as a predictor of performance. In terms of group factors, these would seem to be of little importance as long as the variance from the general factor has been partialled out. As a sidenote, the North American tradition and preoccupation with the varimax criterion for rotation of factors (e.g., see the factorial studies on the WISC-R) does not follow with Vernon's suggestions above--because the variance from the general factor is spread out to the group factors when the varimax criterion is used, the group factors emerging in this manner may indeed not be "meager" but important factors for predicting performance. Differences in methodological and theoretical proceedings between the North American and British approaches need to be considered in the evaluation of factorial studies. For example, the method of factor analysis used on the Stanford-Binet IV (Thorndike et al., 1986) more closely follows the British tradition than the American tradition; that is, extraction of *g* first followed by factoring of the residuals to find group factors. The appropriate factoring procedure depends a great deal on the theoretical assumptions that are made about intelligence.

A further assumption of Vernon deals with the heritability of intelligence. He does not maintain that *g* is "purely an inherited" quality (Vernon, 1961, p.33) but rather that education, training, and occupation may be important in determining *g*. As such, Vernon (1961) suggests that intelligence is formulated through an interaction between hereditary and the environment: "factors over and above *g* arise, partly perhaps from hereditary influences, but mainly because an individual's upbringing and education imposes a certain grouping on his bonds" (p. 32). By bonds Vernon is referring to both inherited "reflexes" as well as acquired habits and associations. As such he suggests that the *v:ed* factor is strong because our society gives a fairly uniform education to all its members. This is interesting, because it parallels Cattell and Horn's suggestion that Crystallized Intelligence also results from uniformity of education.

In essence, the Stanford-Binet IV does measure most of these areas suggested by Vernon (1978), especially if the Perceptual-Organization and Psychomotor abilities are equivalent to the Stanford-Binet IV Abstract / Visual Reasoning Area. At the heart of Vernon's criticism of the Wechsler type tests is that they do not sufficiently measure the group factors in his hierarchical model of intelligence. In contrast the Stanford-Binet IV is more theoretically in agreement with his hierarchical model of intelligence. It is as if the authors of the Fourth Edition of the Stanford-Binet took careful note of Vernon's (1961) statement: "The Stanford-Binet, for example, does not give reliable diagnostic indications of verbal, numerical, memory, spatial or other abilities" (p. 64). These are the exact areas that Thorndike et al. (1986) have included on the Stanford-Binet IV.

Comparison of Gf and Gc with *v:ed* and *k:m*

At this point it is appropriate to consider the similarity between the Cattell-Horn Gf-Gc theory and Vernon's *v:ed-k:m* theory. This comparison is extremely relevant because the *Technical Manual* for the Stanford-Binet IV (Thorndike et al., 1986) suggests that the theoretical structure of the Fourth Edition is based on both Cattell's and Vernon's models. Although many researchers tend to equate Gc with *v:ed* and Gf with *k:m* , Cattell (1971) strongly denies the comparability of the two models. The primary distinction made by Cattell is that in Vernon's model *v:ed* and *k:m* are subservient to g, while in his model, Gf and Gc replace *g* . Methodologically, Vernon works from the top to the bottom of the hierarchy, first extracting g and then factoring the residuals to obtain group factors. In contrast, Cattell works from the bottom to the top of the hierarchy, first factoring the variables to find oblique group factors and then factoring the correlations between the group factors to find higher-order factors. Despite Cattell's assertion that his Gf and Gc are not comparable to *v:ed* and *k:m* , evidence suggests otherwise, especially if it is recognized that later research (Horn & Cattell, 1966; Undheim, 1976, 1981a) has found Gf and Gc to have counterparts of Gs, Gv, and Gr, with higher-order factors presumably more closely representing Spearman's *g* ; or put another way, Gf and Gc may be closer to group factors rather than general factors as initially claimed by Cattell. If Gf and Gc are tentatively considered as very broad group factors, then they seem to be quite comparable to *v:ed* and *k:m* .

Horn (1976) seems to have recognized the comparability between the Gf and Gc factors and Vernon's *v:ed* and *k:m* factors and further their similarity to the Verbal Comprehension and Perceptual Organization factors found on the Wechsler scales. He states that in much of the British work (e.g., Vernon) Gc has been labeled *v:ed* (verbal-educational intelligence). He describes this as "Awareness of concepts and terms pertaining to a broad variety of topics..." (p. 445). He goes on to say that it is found in tests measuring general information, vocabulary, as well as tests requiring a knowledge in science, mechanics, social studies, English literature, and mathematics. From the Wechsler scales, Horn says that the Information, Vocabulary, Comprehension, Similarities, and to some extent the Arithmetic subtests manifest Gc.

On the other hand, Horn (1976) suggests that Gf involves "Facility in reasoning, particularly in figural and non-word symbolic materials..." (p. 445). He suggests that the Block Design, Object Assembly, Picture Arrangement, Picture Completion, and the Mazes subtests from the Wechsler scales, as well as matrices from other tests, measure Gf. He further likens it to the British factor known as *k:m* (spatial-perceptual-practical intelligence).

Although Cattell explicitly states that his Gf and Gc factors are not the same as Vernon's *v:ed* and *k:m*, the two theories seem very similar. Sternberg (1980) in referring to Cattell and Horn's theory goes as far as saying "A very similar theory has been proposed by Vernon (1971), whose major group factors of practical-mechanical ability and verbal-educational ability seem to correspond closely, if not exactly, to Cattell and Horn's factors of fluid and crystallized ability" (p. 41). For all intents and purposes, the two theories will be treated as somewhat interchangeable; preference will be given to Vernon's methodology, as it fits better with that used on the Stanford-Binet IV, and also is slightly more straight-forward in that higher order factoring of oblique factors is not required.

Spearman's Theory of *g*

A major cornerstone of the Binet scales, including the Stanford-Binet IV, is the conception of general intelligence. The debate over general intelligence versus primary factors would not be complete if the work of Spearman and Thurstone was not covered. The theoretical work of these two authors has had a far-reaching influence, and their contributions to the field of intelligence testing provides some insights into the pitfalls that theorists and factor analysts can fall.

Spearman's early work (1904) was rather crude and his conclusions were not totally supported by his limited data. Nonetheless, his 1904 w[illegible] often [illegible]ed as a classic. Though the method of arriving at his conclusions was b[illegible], the conclusion itself seems to have held much merit. In h[illegible] work, Spearman suggested and measured four types of intelligence: (a) a rough in[illegible]ool performance, (b) rank in school ability vs rank in age, (c) rating of "brightness [illegible] school cleverness" by teacher, (d) rating of common sense by two peers (oldest children in the class). Besides these four measures of intelligence, he also measured three types of sensory discrimination: sight, touch, and hearing. The average correlations between the intellectual measures (a & b) and the sensory data was .38; between school cleverness (c) and common sense (d) the rankings intercorrelated .55; and finally, the three sensory measures had correlations of around .25 with each other (Fancher, 1985a).

Based on the above correlations Spearman (1904) arrived at a monumentous conclusion: "On the whole, then, we reach the profoundly important conclusion that *there really exists a something that we may provisionally term 'General Sensory Discrimination' and similarly a 'General Intelligence,' and further that the functional correspondence between these two is not appreciably less than absolute*" (p. 272, emphasis in the original). This suggestion that sensory discrimination and general intelligence correlate perfectly is puzzling, given our knowledge on intelligence. A noteworthy point is that Spearman (1904) was still attempting to measure intelligence by means of sensory discrimination.

Comparison of Spearman's measurements of intelligence with the method used by Binet and Simon (1905/1916) clearly illustrates the revolutionary advancement that Binet and Simon's work brought to the field of intelligence testing. Although the work of Spearman as well as Binet and Simon focused on global general ability, Spearman's work was mainly theoretical while Binet and Simon's work was pragmatic.

A careful reading of Spearman's 1904 work has often left readers with questions about the accuracy of Spearman's statements and calculations. To this end Fancher (1985a) reexamined Spearman's (1904) calculations and concluded that Spearman's findings had "a large number of errors of many different kinds, and results considerably less theoretically 'perfect' than he claimed" (p. 341). Using Spearman's raw data, Fancher (1985a) recalculated all of Spearman's major correlations. Fancher (1985) suggests that "In sum, Spearman's treatment of the Village School data seems marked from beginning to end by arithmetical mistakes, erroneous calculation of correlations, ambiguous or incomplete descriptions of method, and inconsistencies of reporting between his own text and tables of results" (Fancher, 1985a, p. 345). The general pattern of correlations remains the same as reported by Spearman (1904) except that the precision of the calculations is not as "perfect" as reported. Despite the errors in Spearman's 1904 article, his work has been an important contribution to the field because of the debate that it created over the existence of *g* .

In his 1927 book, Spearman expounds on what has come to be known as the "two factor" ability model. As suggested by the name, the main concept is that intelligence can be broken into two parts. First, a general factor (*g*) pervades all abilities, and second, specific factors (*s*) exist which are specific to given tasks. The relationship between *g* and *s* is expounded by Spearman (1927): "Although, however, both of these factors occur in every ability, they need not be equally influential in all" (p. 75). Furthermore, Spearman (1927) suggests that heredity has a large effect on the general factor , while training may have a greater impact on the specific abilities.

According to Spearman (1927), many different *s*'s exist, "one for each distinct kind of performance". However, he also recognized that these specific factors may overlap to some extent: "Obviously, the specific factors for any two performances can only be independent of each other when these performances are quite different" (1927, p. 80). From these specific factors, if enough exist close together, it is also possible to find group factors. This relationship between specific factors and group factors is further elaborated by Spearman (1927):

> Overlapping specific factors have since often been spoken of as 'group factors.' They may be defined as those which occur in more than one but less than all of any given set of abilities. Thus, they indicate no particular characters in any of the abilities themselves, but only some kinship between those which happen to be taken together in a set. Any element whatever in the specific factor of an ability will be turned into a group factor, if this ability is included in the same set with some other ability which also contains [illegible] element. The most that can be said is that some elements have a broader range than others, and therefore are more *likely* to play the part of group factors. (p. 82, emphasis in original)

Thus, although Spearman did not heartily endorse the idea of group factors, he did acknowledge their existence. Specifically he suggested that "Among the exceptional cases where, on the contrary, specific correlations and group factors do become of appreciable magnitude, the four most important have been in respect of what may be called the logical, mechanical, the psychological, and the arithmetical abilities (Spearman, 1927, pp. 241-242).

Defining *g* and *s* proved to be a rather difficult task for Spearman. Before his 1927 work Spearman referred to *g* as that which is common to all abilities; it was defined basically by the process by which it was measured. In his 1927 book he clarifies the nature of the general factor by using analogies. The general factor is likened to "mental energy", a force capable of being transferred from one type of mental activity to another. As such, general intelligence was described as the energy that powers the "specific engines" (Spearman, 1927, p. 133). That is, specific factors are empowered by the general ability that the person possesses.

Spearman's work was not without criticism and a few of these will be presented in order to identify some of the potential pitfalls in his methods. First, a major criticism, and probable cause of his not finding group factors, has to do with how he chose variables to include in his analyses. He insisted that no two tests in a battery be too similar--as this would create problems in that the two tests would correlate above and beyond the contribution made by *g*. By carefully selecting tests and ensuring that no two tests had much in common, group factors were not likely to emerge. A second criticism of Spearman's methodology was that his samples were quite small (Vernon, 1961). Because of these criticisms, it is perhaps safest to think of Spearman's contribution to the field as being theoretical rather than empirical.

Rather than criticize Spearman extensively, Cattell (1971) chooses to point out three major contributions that Spearman's work made to the field of intelligence: first, the suggestion that correlations between most measures of intelligence are positive; second, that tests have different loadings on the general factor depending on the nature of the tests (those requiring complex mathematical and abstract verbal abilities have the highest loadings, and those requiring motor skills and repetition of tasks have the lowest loadings; and third, that speed by which an item can be solved and level of intelligence have some relationship. Though these contributions are important, it was Spearman's insistence that a general intellectual factor (*g*) existed that made him famous. And it is this belief that has filtered its way down through the psychometric tradition and now is embodied in the Stanford-Binet IV.

Thurstone's Primary Mental Abilities

While Spearman championed the idea of general intelligence, it was Thurstone (1938) who propagated the idea of primary mental abilities. A primary ability has been defined as "*a functional unity that is strongly present in some tests and almost completely absent in many others*" (Thurstone & Thurstone, 1941, p. 9, emphasis in original). The object of Thurstone's early work (1938) was to construct tests that were heavily saturated with only one primary ability while simultaneously having minimal loadings on other primaries.

Thurstone's (1938) sample consisted of 240 highly select college students who had "exceptionally high" average mental endowments, and who also had volunteered for the study (Thurstone, 1938, p. 17). As pointed out by Eysenck (1979), the homogeneous nature of this group may have precluded the emergence of a strong general factor and this must be remembered when the results are considered. The battery included 56 separate tests. Results were analyzed using the centroid factoring method which Thurstone had perfected. As was also common in that era, the factors were rotated graphically .

Nine primary factors were identified: Verbal Relations, Perceptual, Inductive Reasoning, Number, Rote Memory, Word Fluency, Space or Visualization, Deductive Reasoning, and Restriction in Solution; the two latter factors being taken as tentative. The implication of Thurstone's 1938 work was that seven or eight distinct and well defined primary abilities might exist rather than a singular general intellectual ability. However, by 1941 Thurstone and Thurstone clearly indicated that it was possible to identify a second-order general factor from the primaries.

To demonstrate that a general factor was obtainable from Thurstone's work, Paden (1981) reanalysed Thurstone's 1938 data using oblique rotations. Thurstone's primary factors were taken as first-order factors. Paden found that three second-order factors (Verbal, Spatial, and Numerical) emerged with all three loading strongly on a third-order factor representing general intelligence. She concluded that a hierarchical analysis of

Thurstone's 1938 data results in a structure that is quite compatible with a hierarchical model such as Vernon's.

Eysenck (1939) also reanalysed Thurstone's 1938 work. His reanalysis is pertinent at this point because it demonstrates a very important point. That is, when factor analytic procedures are used, different theorists can draw seemingly contrary conclusions from the same data. Much depends on the rotational method and the type of analysis used to extract factors. The method used by Eysenck (1939) follows the British tradition used by Vernon, and is similar to that used in the factoring of the Stanford-Binet IV. This procedure follows a theoretical perspective that strongly supports the presence of a general intelligence factor--and guides the extraction of factors accordingly: first the general factor is extracted from the correlation matrix; then the remaining residual matrix is factored and group factors are identified. Using this method of factoring Eysenck (1939) found Thurstone's data had a strong general factor and also a number of sp[illegible]lity factors that were quite similar to Thurstone's primary abilities; *g* accounted for approximately 31% of the total variance while each additional group factors accounted for 2 to 6% of the total variance.

In 1941 Thurstone and Thurstone continued their work on primary mental abilities. However, this time instead of using college students, eighth grade children (14 year-olds) were used. A total of 60 tests were administered to 710 children. Once again the centroid factoring method was used. Because the orthogonal rotations failed to achieve simple structure, Thurstone and Thurstone chose to use oblique rotations in their 1941 work. Although ten primary factors were identified, only six were interpretable and well defined. These six primary factors were Number, Word Fluency, Space, Verbal Comprehension, Rote Memory, and Inductive or Reasoning.

Indeed, the Thurstones (1941) indicated that the second-order correlations between these six primaries was considerably higher than had been found for the adult population (Thurstone, 1938). In fact, they concluded that a single general factor accounted for most of the correlations between the primaries with the Verbal factor loading most heavily and the Rote Memory factor having the lowest loading. Thelma Thurstone (1941) also noted that the Inductive factor had a high loading on the second-order general factor. Of further interest, Thelma Thurstone pointed out that "Among the high correlations we note that the Number factor is correlated with the two Verbal factors. The Word Fluency factor has high correlation with the Verbal Comprehension factor and with Induction. The Rote Memory factor seems to be independent of the other factors." (Thurstone, 1941, p. 109). This finding is of theoretical importance in that numerical and verbal ability are generally found to load on Crystallized Intelligence as well as Vernon's *v:ed* factor. Further, the combination of Word Fluency, Verbal Comprehension, and Induction appear to approximate a Verbal factor of some type. In a major revelation, Thurstone and Thurstone (1941) arrived at the following conclusion: "Each of the primary factors can be regarded as a composite of an independent primary factor and a general factor which it shares with other primary factors" (p. 38).

Defending the Concept of General Intelligence (*g*)

Because the concept of general intelligence has been fundamental to the older Stanford-Binet scales, and because the Fourth Edition continues this tradition to a large degree, it is appropriate to look at some further information regarding *g*. The importance of general intelligence on the Stanford-Binet [illegible]ows the tradition established by Terman and Merrill; Thorndike et al. (1986) state that, "It is our strong belief that the best measure of *g* -- and consequently broadly effective prediction--will stem from a diverse set of cognitive tasks that call for relational thinking in a diversity of contexts" (p. 6). Thus, although the format of the older Binet scales was changed, and though group factors are theorized on the Fourth Edition, the test constructors still maintain that a general ability factor is an important entity, and is consequently reported as the "Composite IQ".

At least three major reasons for the retention of the concept of general intelligence can be marshalled. First, *g* is parsimonious. Second, there is a pervasive source of common or general variance associated with most tests of mental abilities. Third, clinical experience as well as day-to-day living suggest that people differ in intellectual ability. These three reasons are compelling and along with the empirical support have convinced many theoreticians to adopt a model of intelligence that includes a general factor.

Although the mechanisms underlying *g* have not been clearly identified by researchers, the product of *g* can be measured. It is entirely foreseeable that the identification of these processes will have to be relegated to information processing approaches (e.g., Sternberg, 1981a). Despite uncertainty about the exact mechanisms underlying general intelligence, it is maintained that general intelligence does exist and can be measured. The importance of the Stanford-Binet IV as an assessment tool is not it ability to explain the mechanism of *g* , but rather to assess the product or outcome of *g* ; that is, the Stanford-Binet IV measures the behavioral outcome (product) of the hypothetical construct of general intelligence[18] .

The concept of *g* maintained in this thesis is similar to that outlined by McNemar in 1964: "It has been the thesis of this paper that the concept of general intelligence, despite being maligned by a few, regarded as a second-order function by some, and discarded or ignored by others, still has a rightful place in the science of psychology and in the practical affairs of man. It has not been argued that the nature of general intelligence is well understood" (p. 880). Although mechanisms underlying general intelligence have not been thoroughly explained in the literature, it is assumed that the construct of *g* can be measured reliably, and statistically isolated through factorial studies[19] .

---

18 If the reader does not hold this philosophy, he or she would probably feel more comfortable with a tests like the K-ABC that purport to utilize an information processing approach.

19 The Stanford-Binet tests, for example, have effectively served as measuring tools for *g*

Gustafsson's Synthesis of Factorial Models of Intelligence

Gustafsson (1984) has written an interesting article that outlines his work with approximately 1,000 sixth-grade children in Sweden. He used a maximum likelihood method of confirmatory factor analysis (the LISREL program) and compared the models of Spearman, Thurstone, Guilford, Vernon, and Cattell-Horn. His test battery included 13 test of ability and 3 achievement tests. However, before looking at his findings, some serious methodological short-comings in his study need to be addressed.

Unfortunately, his choice of particular tests, and the limited number he included, may have biased his findings. As noted previously by Cattell (1971), with such a limited number of variables, and with the "factorial pyramid illusion" it is highly likely that the number of second- and third-order factors could have been severely limited by the number of tests at the first order of analysis. Another major problem encountered in Gustafsson's work is his splitting of subtests into halves so that he could substantiate factors; the result of this may have been an artificial confirmation of several factors. However, despite these problems, Gustafsson's research is interesting in that it attempts to demonstrate how several seemingly contrary theories may have more in common than first appears.

Gustafsson (1984) calls his final model the "HILI-model" which is short for "HIerarchical, LISREL-based model" (p. 193). This model has a third order general factor, second order factors of Gf, Gc, and Gv, and ten first order factors. As such, it is quite similar to the revised Horn and Cattell (1966) model. Gustafsson (1984) suggests that the HILI-model is a unifying one:

> The Spearman, Thurstone, and Cattell-Horn models may, in a structural sense at least, be viewed as subsets of the HILI-model: the Spearman model takes into account variance from the third-order factor; the Thurstone model takes into account first-order variance; and the Cattell-Horn model takes into account both first- and second-order variance. The Vernon model comes close to the proposed model: The *g* -factor is included in both

models, and at the second-order level v:ed closely corresponds to Gc', and k:m corresponds to Gv' (p. 193).

Although the Gustafsson model and methodology have inherent limitations resulting from the choice of initial variables and the progressive modification of models, his attempt to integrate diverse factorial models holds merit. Most noteworthy and applicable to the Stanford-Binet IV is the fact that his final model included a general factor and was hierarchical in nature. Also, he suggests that both the Cattell-Horn model as well as Vernon's model fit the structure fairly well--an encouraging finding for the Stanford-Binet IV which is based on these theoretical models. With a cautionary note, Gustafsson warns that the HILI-model is extremely tentative. This warning should be duly noted, for with more initial variables the number of levels in the hierarchy as well as the number of factors at the second-order might change--perhaps drifting closer to the Horn and Cattell model with five or so second-order factors.

Summary of the Theoretical Models

From this review of the theoretical models of intelligence which are pertinent to the Stanford-Binet IV, it is evident that general intelligence is a defensible construct. Furthermore, the work of both Vernon and Cattell substantiate a hierarchical structure of intelligence from which the Stanford-Binet IV has been constructed. A noteworthy point to consider is that several different methods of factoring exist, and that the method used will partially determine the findings that emerge. Therefore, it must be remembered that factor analysts use methods that tend to support their theoretical notions. Hence, in evaluating the Stanford-Binet IV, it is appropriate to use factoring methods that at least allow for the possibility of confirming the proposed theoretical structure.

F. Review of the Factor Analytic Studies of the WISC-R

Because the WISC-R is included in this thesis as an instrument against which the concurrent validity and factor structure of the Stanford-Binet IV can be judged, factor analytic studies done on the WISC-R are reviewed next. Although principal axes factoring followed by varimax rotations has been the most popular method used for examining the factor structure of the WISC-R, hierarchical analysis using the Wherry and Wherry method has also been done. The latter is of special interest because it attempts to deal with the general factor at the same time as the group factors. Because Kaufman's (1975) analysis of the WISC-R standardization sample has been the prototype for subsequent factoring of the scale, it will be given more priority than other studies. Also, because Kaufman's sample was a "normal" one, and because the current sample for this thesis is a clinical sample, a review of the factorial studies for clinical populations was included. Finally, the development of deviation quotients based on the factorial structure of the WISC-R will be considered. The reason for including the latter is that a similar type of procedure is anticipated for the Stanford-Binet IV as further factorial studies clarify the factorial structure of this new Binet test.

A Brief History of the Wechsler Scales

To begin this section on the WISC-R it is only fitting that a brief history of the Wechsler scales be undertaken. Given David Wechsler's contribution to the measurement of intelligence, his work deserves credit. Further, the conception behind his tests will hopefully help to clarify the factor structure that emerges when the WISC-R and the Stanford-Binet IV are factored jointly in this thesis. Once Wechsler's conception of intelligence is understood, then empirical findings in this thesis can be compared to not only the theoretical model as proposed on the Stanford-Binet IV but also with Wechsler's theoretical conception of intelligence. It is of considerable interest to see which theoretical model of intelligence best fits the factor structure that results when these two tests are factored jointly.

A family of Wechsler tests have sprung from the original Wechsler-Bellevue scales (Form I and Form II). This family has included the Wechsler Intelligence Scale For Children (WISC)[20], the Wechsler Adult Intelligence Scale (WAIS), the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), the Wechsler Intelligence Scale For Children-Revised (WISC-R) and finally the Wechsler Adult Intelligence Scale-Revised (WAIS-R). In fact, whereas the Binet scales once served as the standard by which other tests were judged, Hill, Reddon, and Jackson (1985) note that the Wechsler scales have become the new standard by which intelligence tests are judged. The WISC-R is, therefore, an excellent test against which the Stanford-Binet IV can be compared.

The WISC-R shares much with its predecessor the WISC. The popularity of the earlier WISC almost guaranteed that the WISC-R would thrive, especially with its improved reliability and organization. According to Wechsler the WISC-R remains structurally and contextually the same as the WISC, though on the revised test the order of administering the Verbal and Performance subtests was alternated. Indeed, the bulk of research comparing the WISC-R to the WISC tends to support the similarity of the two tests (for a review see Quattrocchi & Sherrets, 1980). An exception is that the age ranges for which the WISC-R is applicable (6-0 to 16-11 years) is slightly different from the WISC (5-0 through 15-11 years). In terms of content, a substantial amount of continuity exists between the two scales: 78% of the items on the WISC-R have come directly from the WISC; an additional 5.9% came from the WISC but were substantially modified; in contrast, 16.1% are entirely new (Swerdlik & Schweitzer, 1978).

The major theoretical underpinning behind the WISC-R is a commitment to the measurement of general intelligence. In the preface of the WISC-R manual (Wechsler, 1974), the opening statement is as follows: "The revised WISC, like the Scale it succeeds,

[20] The WISC was originally designed as a downward extension of the Wechsler-Bellevue.

has been designed and organized as a test of general intelligence. Its author believes that general intelligence exists; that it is possible to measure it objectively; and that by so doing, one can obtain a meaningful and useful index of a subject's mental capacity" (p. iii). Despite being arranged into subtests for convenience of administration, the WISC-R remains primarily a test of general intelligence. A look at Wechsler's definition of intelligence also confirms his belief in the idea of general intelligence: "Intelligence is the overall capacity of an individual to understand and cope with the world around him" (Wechsler, 1974, p. 5). As such, Wechsler's definition of intelligence is both practical and easy to grasp.

Not only does Wechsler stress that intelligence is a global entity, but he also maintains that it is multidetermined and multifaceted; as such, certain abilities (e.g., abstract reasoning) are not singled out as more important than other abilities. To illustrate the multifaceted nature of intelligence, Wechsler (1974) uses an analogy in which he likens the manifestation of intelligence to different languages:

> To the extent that tests are particular modes of communication, they may be regarded as different *languages*. These languages may be easier or harder for different subjects, but it cannot be assumed that one language is necessarily more valid than another. Intelligence can manifest itself in many forms, and an intelligence scale, to be effective as well as fair, must utilize as many different languages (tests) as possible (p. 5).

In order to measure these different "languages" the WISC-R was constructed to provide not only a Full Scale IQ (FSIQ) but also a Verbal Scale IQ (VIQ) and a Performance Scale IQ (PIQ). Although the Performance and Verbal scales each have six subtests, only five of each scale are mandatory. On the Verbal Scale, these are Information, Similarities, Arithmetic, Vocabulary, and Comprehension. On the Performance Scale, the mandatory subtests are Picture Completion, Picture Arrangement, Block Design, Object Assembly and Coding (or alternately Mazes as indicated by

Wechsler, 1974). The two supplemental tests are Digit Span on the Verbal Scale, and Mazes on the Performance Scale. An added benefit of the WISC-R is that the same subtests are administered to each child at each age. This makes it a useful tool for both longitudinal research as well as for comparison across different ages. Statistically, the three major scales have a mean of 100 and a standard deviation of 15; similarly, each subtest has a mean of 10 and a standard deviation of 3. The IQ's are deviation quotients, and thus allow for direct comparison of a child's score with those of other children.

A final point to consider about the WISC-R is the thorough norming that was done. The standardization sample consisted of 2,200 children--200 at each age level from 6 1/2 to 16 1/2. Furthermore, based on the 1970 U.S census data, the sample was stratified in terms of age, sex, race, geographic region, occupation of head of household, and urban-rural residence. Because of the care taken by Wechsler, the WISC-R standardization sample was quite representative of the U.S. population--something that the sample from the Stanford-Binet L-M was not.

Prevalence of Factor Analytic and Other Research on the WISC-R

Research publications on the WISC-R in the late 1970's and early 1980's were numerous. As noted by Kaufman (1981), the structure of the WISC-R is conducive not only to research but also to clinical utility. The ease of administration, reliability, and interpretability of the WISC-R resulted in "...the emergence of the WISC-R as the clear-cut instrument of choice for the assessment of intelligence in school age children, with the Stanford-Binet left wallowing in the WISC-R's wake" (Kaufman, 1981, p. 520).

The structure of the WISC-R makes it highly amenable to factorial work. Because each child is given the same ten mandatory subtests, it is usually assumed that the same type of intelligence is measured at each age. This conception is further affirmed by the fact that on both the WPPSI and WAIS-R have subtests that are similar to those on the WISC-R. A final advantage of the WISC-R, though not immediately evident, is that research studies are not plagued by the problem of missing data. This problem can be noted on tests like the Stanford-Binet IV which are administered according to an adaptive testing format. With this type of format not all children take the same tests, depending on the age and ability of the child. Consequently, researchers are often faced with the problem of missing data.

With this knowledge about the background of the WISC-R, it is now possible to begin a review of its factorial structure. As a preview, it is worth noting that most factorial studies done on the WISC-R have indicated that two robust factors emerge. These correspond roughly to the Verbal and Performance Scales, and are respectively called Verbal Comprehension and Perceptual Organization. For administrative purposes, when the phrase "the typical two factor solution" is used, reference is being made to the Verbal Comprehension and Perceptual Organization factors found by Kaufman (1975). Similarly, because many studies have found three factors rather than just two, when the phrase "the typical three factor solution" is used, the reference is to the above two factors as well as Kaufman's (1975) third factor which is known as Freedom From Distractibility.

Kaufman's 1975 Factor Analysis of the WISC-R Standardization Sample

Kaufman's (1975) factor analysis of the WISC-R is an extremely important article. It has been highly influential for two reasons. First, the correlations reported in the WISC-R manual for the standardization sample were used as the basis for the factor analysis. Second, the three factor solution and the pattern of loadings that he found have been typical of almost all subsequent factorial studies on the WISC-R. His analysis involved 200 children at each of 11 age levels. All twelve tests (the 10 mandatory and 2 optional ones) were used. The number of factors to retain at each age level was determined by performing principal components analyses and using the eigenvalue-one criteria. In the main analysis, however, principal axes factoring was used; squared multiple correlations were initially inserted into the diagonals as communality estimates and these were iterated.

From the principal components analysis, it appeared that three factor were indicated for most of the eleven age levels; the eigenvalues for the third factor ranged from .9 to 1.1. Even so, Kaufman decided to rotate 2, 3, 4, and 5 factors for each age level. He states that at all age levels, the two-factor rotated solutions indicated that "clear-cut" Verbal and Performance factors were present. To fit with previous literature, these were described as Verbal Comprehension and Perceptual Organization. When three-factor solutions were examined, Kaufman found that at all age levels except 6 1/2 and 14 1/2 years that the third factor turned out to be Freedom From Distractibility, with high loadings from Arithmetic and Digit Span; for these two exceptional age levels, Kaufman found that Freedom From Distractibility did emerge, but as the fourth rather than third factor. As a general rule Kaufman suggested that the four-, and five-factor solutions added no additional interpretability to any of the 11 age groups.

From Kaufman's results it is apparent that for all ages the Verbal Comprehension factor had its highest loadings from the Vocabulary, Information, Comprehension, and Similarities subtests. The loadings from the Arithmetic subtest on this factor were moderate with a median loading of .37. Somewhat unexpected, Picture Completion and Picture Arrangement also had moderate loadings on the Verbal Comprehension factor; Kaufman interpreted this to mean that both of these subtests probably involve a greater degree of verbal mediation than do other Performance subtests. Meanwhile, the Perceptual Organization factor was also quite consistent across all ages. It had high loadings from the Block Design, Object Assembly, and Picture Completion subtests; two further subtests, Picture Arrangement and Mazes had moderate loadings on this factor. Finally, at most age levels the Freedom From Distractibility factor had substantial loadings from the Arithmetic and Digit Span subtest--the Information and Coding subtests had moderate loadings on this third factor, but not at all age levels.

The factor loadings found by Kaufman (1975) on the three varimax rotated factors of Verbal Comprehension, Perceptual Organization, and Freedom From Distractibility have been summarized in Table 1. Because the current research will utilize three age groups (ages 6-0 to 8-11 years, 9-0 to 12-11 years, and 13-0 to 16-11 years) Kaufman's factor loadings for the 11 age levels have been regrouped according to the above three age groupings. The mean of the appropriate loadings was calculated for each subtest at each level. Collapsing across Kaufman's age levels was done in order to provide a reference structure for the appropriate age range for each of the three age groups used in this thesis.

Table 1

Three-Factor Varimax Rotated Solutions for the WISC-R Standardization Sample: A Regrouping of Kaufman's (1975) Findings into Three Age Groups

Kaufman's 1975 Analysis of theWISC-R Standardization Sample	Three Factor Varimax Rotated Principal Axes								
	Mean Loadings[a] For Ages 6 1/2 to 8 1/2			Mean Loadings[b] For Ages 9 1/2 to 12 1/2			Mean Loadings[c] For Ages 13 1/2 to 16 1/2		
Tests	I	II	III	I	II	III	I	II	III
Information	.50	.24	.46	.62	.30	.37	.67	.27	.35
Similarities	.56	.26	.31	.66	.38	.30	.66	.33	.28
Arithmetic	.34	.20	.56	.35	.27	.54	.43	.19	.54
Vocabulary	.67	.21	.28	.69	.32	.39	.77	.23	.37
Comprehension	.57	.29	.23	.64	.31	.25	.67	.30	.20
Digit Span	.45	.17	.53	.17	.11	.56	.23	.09	.54
Picture Completion	.30	.42	.17	.34	.60	.10	.35	.56	.09
Picture Arrangement	.38	.46	.21	.33	.48	.17	.31	.41	.11
Block Design	.28	.56	.28	.28	.64	.37	.26	.72	.26
Object Assembly	.21	.64	.21	.28	.64	.13	.19	.67	.09
Coding	.14	.20	.30	.14	.17	.46	.21	.20	.33
Mazes	.16	.56	.20	.12	.46	.21	.07	.44	.19

[a] The loadings reported for the three factors here are the mean of the loadings for the three ages 6 1/2, 7 1/2, and 8 1/2 reported in Kaufman's (1975) Tables 1, 2, and 3.

[b] The loadings reported for the three factors here are the mean of the loadings for the four ages 9 1/2, 10 1/2, 11 1/2 and 12 1/2 reported in Kaufman's (1975) Tables 1, 2, and 3.

[c] The loadings reported for the three factors here are the mean of the loadings for the four ages 13 1/2, 14 1/2, 15 1/2 and 16 1/2 reported in Kaufman's (1975) Tables 1, 2, and 3.

Kaufman (1975) examined the first unrotated factor at each level and found that it accounted for approximately 79-92% (Median = 82%) of the common factor variance. Most of the subtest loadings on this first unrotated factor ranged from .60 to .80 with the exception of Digit Span, Coding and Mazes which were in the .40 to .50 range. Kaufman interpreted this first unrotated principal factor as synonymous with general intellectual ability. In so doing he suggested that each subtest's loading on this first factor be taken as the degree to which the subtest measures general intelligence; by squaring these loadings on the first factor, one can obtain the proportion of each subtest's variance that may be attributed to general intelligence. Although this method of estimating *g* is acceptable, it becomes somewhat problematic if it is reported along with factors that have been rotated orthogonally by the varimax criterion--as was done by Kaufman (1975). When researchers report the general factor in this manner, the reader must remember that a large percentage of the variance from the general factor is duplicated in the rotated factors--the variances are not additive as they would be if the first unrotated factors was extracted and then the residuals factored (e.g., as done by Vernon (1961) or by Thorndike et al., (1986)).

Nonetheless, these loadings on the first unrotated factor found by Kaufman (1975) are typical of those reported by other researchers. For example, Silverstein (1980) compared Kaufman's (1975) reported values for the first unrotated factor with Wallbrown, Blaha, Wallbrown, and Engin's (1975) general factor that emerged from a hierarchical analysis of the WISC-R; this is a particularly appropriate comparison because both Wallbrown et al. (1975) and Kaufman (1975) used the standardization data from the WISC-R as a starting point. Despite the difference in these two methods, the general factor that emerged was nearly identical. Thus, Kaufman's (1975) suggestion that the first unrotated factor be considered synonomous with general factor seems meritous; however, it must be remembered that when factors are rotated according to analytic criteria that the variance from the first unrotated factor (general intelligence) is usually spread out onto other factors, and that to report the variance of the general factor along with these other factors may lead some to erroneously believe that the variance attributable to these other factor is independent of general intelligence. For example, in Kaufman's analysis he did not partial out the variance attributable to general intelligence but rather it was included in the varimax rotated factors he reported.

As a sidenote, before leaving the analysis of the standardization sample to look at the factorial structure of the WISC-R in clinical populations, it is worth noting that Reynolds and Gutkin (1980) reanalysed the standardization sample to compare the factor patterns for males and females. The three factors that they found for each sex were nearly identical to those found by Kaufman (1975). Similarly, Carlson, Reynolds, and Gutkin (1983) reanalysed the standardization sample by comparing the factor structure for lower and upper SES groups. They also found that the factor structure for the two groups was very similar and nearly identical with the structure found by Kaufman. These findings led the respective researchers to conclude that the factor structure of the WISC-R is invariant across sex, age, and socioeconomic status in terms of the magnitude of loadings and also the composition of the three factors that emerge.

Factorial Structure of the WISC-R for Clinically Referred Children

Several researchers have anlayzed the factor structure of the WISC-R in clinical populations. These populations have varied in terms of the type of problems that the children had as well as the IQ ranges that were present. Generally, the Verbal Comprehension and Perceptual Organization factors have emerged in the stable pattern found by Kaufman (1975). However, some variation in factor loadings on these factors does occur. Also, in these clinical populations, the subtests loading on the third factor, if it emerges, may change in magnitude or composition. (i.e., the size of the loadings and which subtests load on the factor). The purpose of this brief review of the factorial structure of clinically referred children is to provide some indication of how factor loadings on the three WISC-R factors vary as compared to the factor structure for "normal" children (Kaufman, 1975).

Unless reported otherwise, all the studies in this section have use the principal factoring method, squared multiple correlations in the diagonal, iterations to convergence, and varimax rotation of factors.

Lombard and Riedel (1978) looked at the WISC-R factor structure for 76 children (ages 8-0 to 16-11) who had been referred for assessment and placement in special education programs because of learning disabilities. The Full Scale IQs ranged from 59 to 138 with a mean of 94.7 and a standard deviation of 15.7. Unfortunately only the 10 mandatory subtests were used in the analysis. The pattern of loadings on the Verbal Comprehension and Perceptual Organization factors was that same as reported by Kaufman (1975) except that the loadings were much higher. Loadings on the Verbal Comprehension factor were as follows: Information (.81), Similarities (.74), Arithmetic (.74), Vocabulary (.86) and Comprehension (.86). Loadings on the Perceptual Organization factor included Picture Completion (.80), Picture Arrangement (.69), Block Design (.81), and Object Assembly (.74). The third factor was defined mainly by the Coding subtest which had a loading of .93; this is not surprising in that Digit Span was omitted from the analysis. The implication from this study is that for referred populations, the factor loadings may be slightly higher than for normal populations--perhaps because of the extreme spread in the IQ range. With the large range in FSIQs, one would expect the correlations between subtests to be slightly higher than normal, and thus the factor loadings may have been slightly higher than reported by Kaufman (1975).

Naglieri (1981) also examined the factor structure of 140 children between the ages of 6-2 and 14-8 who had been classified as learning disabled (Mean FSIQ = 94.8). Both principal components and principal factoring followed by varimax rotation of factors with eigenvalues greater than one was employed on eleven subtests (Mazes omitted). From the principal components analysis, the Verbal Comprehension factor consisted of Information, Similarities, Comprehension, and Vocabulary (with loadings between .74 and .88), as well as Arithmetic (.33). The Perceptual Organization factor consisted of Picture Completion, Block Design, and Object Assembly (with loadings between .76 and .82), as well as Picture Arrangement (.52). The Freedom from Distractibility factor consisted of Digit Span (.67), Coding (.75), Arithmetic (.23), and Picture Arrangement (.29).

From the Naglieri (1981) study, two points are worth emphasizing. First, the principal factoring and principal component methods produced nearly identical patterns of loadings though the component model does produce slightly higher loadings. Second, the loading of Picture Arrangement on the third principal component is somewhat unusual. To account for this finding, Naglieri suggested that the nature of the third factor may involve sequencing, or perhaps successive processing. However, because the Picture Arrangement loading was only in the low moderate range, this hypothesis should be regarded cautiously.

Petersen and Hart (1979) examined the factor structure of the WISC-R for three groups of second-, third- and fourth-graders who had been referred for assessment: those who were slow learners or were learning disabled (n=162); those who were emotionally handicapped (n=147); and those who had 'no significant problem' (n=248). All th WISC-R subtests except Mazes were included in the analysis and all factors with eigenvalues greater or equal to one were rotated. In all three groups, the Verbal Comprehension and Perceptual Organization factor were clearly evident. However, the factor loadings for the third factor were somewhat different from group to group, and also somewhat different than the third factor loadings reported by Kaufman (1975). For the 'no significant problem' group, factor loadings on the third factor included Information (.32), Arithmetic (.46), Digit Span (.38), Block Design (.55), Object Assembly (.48) and Coding (.25). For the emotionally handicapped group, only Arithmetic had a salient loading (.75) on the third factor. For the slow learner / learning disabled group, the third factor had loadings from Picture Completion (.37), Picture Arrangement (.39), and Coding (.48). Thus, Petersen and Hart (1979) concluded that although the first two factors on the WISC-R are stable across populations, the third factor may have genuine compositional differences depending on the nature of the sample.

Groff and Hubble (1982) analysed the factor structure for children who were being considered for placement in special education programs because of low intellectual ability (IQs in the retarded to borderline range). Two groups based on age level were established; the younger group ranging from 9 to 11 years (n=107), and the older group ranging from 14 to 16 years (n=78). All the WISC-R subtests with the exception of Mazes were included in the analysis. Although two-, three-, four-, and five-factors solutions were rotated, the three factor solution seemed most appropriate. Several findings are worth noting. For the younger age group, the Similarities and Comprehension subtests had only moderate loadings (.41 and .36) on the Verbal Comprehension factor; also, the Coding subtest loaded only moderately on the Freedom from Distractibility factor (.31) which was defined primarily by Digit Span (.79) and Arithmetic (.48). For the older group, Picture Arrangement had a moderate loading on the Verbal Comprehension factor (.37); also, in an unusual fashion, Digit Span had its highest loading on the Perceptual Organization factor (.41); the third factor had moderate loadings from only Arithmetic (.58) and Coding (.42). Overall, the three factor solution was supported for both age levels even though some unusual factor loadings on some of the factors were present.

Swerdlik and Schweitzer (1978) also examined the WISC-R factor structure for children suspected of low intellectual ability. Their sample consisted of 164 children ranging in age from 7 to 15 years. All twelve subtests were used in a principal components analysis followed by varimax rotation of factors. When two factors were rotated the typical pattern emerged. However, when three factors were rotated the composition of the third factor was somewhat unusual; as typically found, Arithmetic, Information, and Digit Span had high loadings; Coding, on the other hand did not load on the third factor (.19) but Picture Arrangement did have a moderate loading (.43).

Van Hagen and Kaufman (1975) examined the factor structure of 80 children who were classified as mentally retarded. These children ranged in age from 6-3 to 16-9 years and had FSIQs ranging from 40 to 79. The principal factoring method was used with factors being rotated by the varimax, oblimax, and biquartimin criterion. Only two factors had eigenvalues greater than or equal to one, and these corresponded to the Verbal Comprehension and Perceptual Organization factors found by Kaufman (1975). The Verbal Comprehension factor had loadings from all of the Verbal Scale subtests as well as from Picture Arrangement (.46); the Perceptual Organization factor had loadings from all of the Performance Scale subtest as well as from Comprehension (.48). Further, the first unrotated factor had loadings of .44 to .74 from the subtests and accounted for 74% of the common variance. When a three-factor solution was rotated, the Arithmetic (.53), Picture Arrangement (.45) and Coding (.43) subtests had salient loadings on the third factor. As had been found by Kaufman (1975), the orthogonal (varimax) and the oblique (oblimax and biquartimin) rotations produced essentially the same loadings on each factor. Van Hagen and Kaufman (1975) concluded that the WISC-R factor structure is essentially the same for children with low intellectual ability as it is for children with normal intellectual ability.

Johnston and Bolen (1984) compared the factor structure for a clinical sample of black and white children who had been referred because of poor academic progress, reading difficulties, suspected mental retardation, or learning disabilities. Their sample consisted of 274 black children and 430 white children, all with FSIQs in the 70 to 100 range. Only the 10 mandatory tests were included in the analysis. Although no set criteria was used to decide on the number of factors to accept, Johnston and Bolen report that the usual three factor solution emerged for both blacks and whites. However, with the Digit Span test not being included in the analysis, the third factor was defined almost solely by a singular high loading (.98) from the Arithmetic subtest.

Karnes and Brown (1980) examined the WISC-R factor structure in a sample of 946 children with high IQs (Mean FSIQ = 126.4) who were in gifted programs. Unfortunately only the ten mandatory subtests were used in the analysis. When two factors were extracted the typical two factor pattern emerged with the exception that the

Coding subtest had its highest loading on the Verbal Comprehension factor (.36). When three factors were rotated, the high loadings on the third factor included Arithmetic (.81), Picture Completion (.55), and Coding (.27).

Sapp, Chissom, and Graham (1985) also looked at the WISC-R factor structure for children with high levels of intelligence (n = 371, Mean FSIQ = 128.7, SD = 9.4). The ten mandatory WISC-R subtests were used in the anlaysis. For the three factor solution, the first two factors had typical loadings as suggested by Kaufman's (1975) results; however, the third factor had loadings from the Information (.68), Arithmetic (.39), Vocabulary (.43), and Block Design (.39) subtests. Of interest, the Coding subtest failed to load on any of the three factors, with its highest loading being .05.

The unusual pattern of loadings on the third factor found by Sapp et al. (1985) may have resulted from the omission of the Digit Span subtest from the analysis. Because Digit Span is usually a strong marker variable for the third factor, its absence may have prevented the emergence of the typical third factor. The composition of Sapp et al.'s third factor seemed to fit with the category that Bannatyne labelled "Acquired Knowledge"; that is, a dimension that develops through learning experiences (Arithmetic, Information, and Vocabulary). Given the unusual composition of the third factor, Sapp et al.'s conclusion was that the third factor may have a slightly different meaning for gifted children than for "normal" children; however, this suggestion must be strongly qualified, as Digit Span and Mazes were not included in the analysis.

Schooler, Beebe, and Koepke (1978) factored the 10 mandatory subtests from the WISC-R in a sample of 799 children who had been placed in one of three groups: learning disabled, educable mentally impaired, or emotionally impaired. Principal components with eigenvalues greater or equal to one were subjected to varimax rotations for each group. With this criteria, only two principal components emerged. Similar to Kaufman's findings, these were remarkably similar to the Verbal and Performance scales. Interestingly, Arithmetic loaded on both factors. The reason for this is not totally clear, though several possible explanations are possible. First, because Digit Span was not included in the analysis, the third factor may not have been adequately defined in order for the eigenvalue criteria to be met; if a third factor had been rotated, the variance of the Arithmetic subtest may have been shifted to the third factor rather than being spread out on the first two factors. Second, as Schooler et al. suggest, Arithmetic may be a complex subtest, and therefore load on both the first and second factors. The overall conclusion reached by Schooler et al. was that the factor structure of the WISC-R was stable across the different groups in their sample.

Hodges (1982) examined the WISC-R factor structure in a psychiatric population of 240 children ages 6 to 16 who were "outpatients" at a mental health facility. The results fit very closely with Kaufman's (1975) three factor solution though some slight deviations were found: Picture Arrangement had a moderate loading (.31) on the Verbal Comprehension factor; Similarities had a moderate loading on the Perceptual Organization factor (.34); and the Freedom from Distractibility had moderate loadings from Arithmetic (.60), Coding (.58), Digit Span (.38), and Block Design (.37).

In summary, the factor structure of the WISC-R for clinically referred children is somewhat similar to the factor structure for normal children (Kaufman, 1975). However, though the Verbal Comprehension and Perceptual Organization factors appear to be stable across populations, the third factor (Freedom From Distractibility) does not emerge in all clinical samples; when it does emerge, the pattern of loadings sometimes is different than those found for normal children. Overall, though, the factor structure for learning disabled, retarded, gifted, and emotionally disturbed children is similar enough to that of normal children that most researchers suggest that no qualitative differences exist in the comparative factor structures.

The Number of Factors Problem Concerning the WISC-R

In most factor analytic studies of the WISC-R, the number of factors to rotate has been taken as the number of principal components that have eigenvalues greater than or equal to one. Following the lead of Kaufman (1975) many researchers have also rotated two-, three-, four- and five-factor solutions and retained only factors that are found to be psychologically meaningful--usually three which coincidentally happens to be the number of factors with eigenvalues greater or equal to one. Hill, Reddon, and Jackson (1985) address the issue of how many factor exist and/or can be interpreted on the WISC-R. They suggests that "A three factor solution has been the one typically reported, and seems to be the preferred number of factors for clinical applications" (p. 296). However, they also suggest that in terms of several "number-of-factor rules" that one or two factor solutions may be preferable. Further complicating the number of factors problem on the WISC-R is the fact that differing numbers of subtests are included in different analyses. In general, when using orthogonal rotations (primarily varimax) the two-factor solution appears to be well substantiated in terms of clinical utility; the three-factor solutions also appears to be clinically useful, though somewhat more tentative in certain populations.

Summary of the Three Factors and Their Loadings

The previous literature review has focused on the WISC-R factor structure in normal children (the standardization sample) and in clinically referred children, and in different ethnic groups. The most general finding is that the Verbal Comprehension and Perceptual Organization factors are very robust and emerge regardless of the factor analytic technique used, the age of the subjects, or the nature of the sample (e.g, psychiatric, learning disabled, mentally retarded, gifted, or normal). In contrast, the third factor, known by the name of Freedom From Distractibility, is less robust and may be limited to white / anglo children. Nevertheless, in clinical populations the third factor has proven to be useful and therefore substantiated to some degree. To help clarify the nature and variability of these three factors, a brief summary of each is given.

Verbal Comprehension

Four of the five regularly administer Verbal Scale subtests (Information, Similarities, Vocabulary, and Comprehension) almost invariably have strong loadings on the Verbal Comprehension factor. In contrast, the Arithmetic and Digit Span subtests sometimes load on this factor, but often load on the third factor more strongly. Of the Performance Scale subtests, Picture Completion and Picture Arrangement occasionally have small but apparently significant loadings on this factor.

Perceptual Organization

Similarly, most of the subtests from the Performance Scale tend to load substantially on the Perceptual Organization factor. Those with the highest loadings are usually Block Design, Object Assembly, Picture Completion, and Mazes. Picture Arrangement also usually loads on this factor though not always as strongly as the four previously mentioned subtests. The only Performance Scale subtest that usually does not load on the Perceptual Organization factor is Coding, which usually loads most heavily on the third factor.

Freedom From Distractibility

This factor is the most complex of the three typical factors found on the WISC-R. When the third factor emerges, it is most likely to have high loadings from the Digit Span, Arithmetic, and Coding subtests. However, studies have demonstrated that other subtests may have moderate loadings on this factor, especially in clinical samples. A list of these subtests and the studies finding them loading on the third factors follows: Information (Kaufman, 1975), Picture Arrangement (Naglieri, 1981; Swerdlik & Schweitzer, 1978), Picture Completion (Karnes & Brown, 1980; Petersen & Hart, 1979), and Block Design (Hodges, 1982; Kaufman, 1975; Petersen & Hart, 1979). Overall, though, it is still the Digit Span, Arithmetic, and Coding subtests that define the third factor.

Given the complexity of this third factor, it is not surprising that Kaufman (1975) suggests that its interpretation is a "thorny problem" (p. 139). At an pragmatic level, Kaufman (1979) warns that clinicians doing ipsative comparisons not consider a child 'distractible' unless that child has similar loadings--all low in comparison to other subtests--on the Digit Span, Arithmetic, and Coding subtests (and possibly on the Information subtest).

A further problem with the third factor is that it can be conceived of in many different ways. For example, Kaufman (1975) admits that this factor can and has been alternately conceived of as Freedom From Distractibility, as a Memory factor, and as a Numerical factor. Although he does not discount the possibility of the third factor being a Memory or Quantitative factor, Kaufman suggests that tentatively it be called Freedom From Distractibility in that each of the above subtests would seem susceptible to the effects of distractibility.

Development of Deviation Quotients For the Three WISC-R Factors

Based on the factorial work done with the WISC-R, several attempts have been made to develop factor scores as well as deviation quotients for each of the three WISC-R factors (Clampit, Adair, & Strenio, 1983; Clampit & Silver, 1986[21]; Gutkin, 1978, 1979, 1982; Sobotka & Black, 1978) These attempts have important implications for those using the WISC-R. Using factor scores or factorially derived deviation quotients rather than relying on the Verbal or Performance Scale IQs allows for more pure measures of different types of intellectual abilities (i.e., different group factors). Although these factorially derived deviation quotients (DQ) are not expected to replace Wechsler's VIQ or PIQ, a clinician can profitably use these new deviation quotients to help understand and guide profile analysis on the WISC-R.

[21] Clampit & Silver use unit weights rather than differential weights.

Of further anticipated importance, the methodology used for deriving factorially pure deviation quotients for the WISC-R has much merit for use on the Stanford-Binet IV. As more research is accumulated on the Stanford-Binet IV, factorially derived deviation quotients may be needed to replace the four Standard Area Scores; this would be a positive step to take if research consistently indicates that the subtests on the Stanford-Binet IV do not load on on their respective Areas as outlined in the structure of the test If such a finding emerges, it will behoove researchers to calculate alternate deviation quotients that are more compatible with the factor structure of the Stanford-Binet IV. Therefore, it is extremely worthwhile to note how factorially pure deviation quotients have been devised for the WISC-R. This is especially true given that Thorndike et al. (1986) found that their confirmatory factor analysis did not substantiate the arrangement of several subtests onto hypothesized factors.

Hierarchical Factor Analyses of the WISC-R

A look at the hierarchical factoring of the WISC-R is worthwhile, for it demonstrates how different methods of factoring can result in slightly different pictures of mental ability. It should be noted that all of the hierarchical analyses of the WISC-R have been done by a small number of researchers (primarily Vance, Wallbrown, and Blaha). Further, they proceed from Vernon's model of intelligence and subsequently talk of the *v:ed* (Verbal Educational factor) and the *k:m* (spatial-mechanical or spatial-perceptual factor) factors rather than the Verbal Comprehension and Perceptual Organization factors as named by Kaufman (1975). Nevertheless, despite the difference in names, the factors appear to be basically the same as those found using varimax rotations, though the general factor is included in the solution.

The method of factoring in these hierarchical analyses is the Wherry and Wherry Hierarchical method (see Wallbrown, Blaha, Wallbrown, & Engin, 1975, for a description of the Wherry and Wherry Method). Of some interest, the Wherry and Wherry method is not akin to other hierarchical factoring methods that use oblique rotations and then second-and third-order factoring of correlations between first-order factors. Rather, the Wherry and Wherry method maintains orthogonality among factors at all levels of the hierarchy (Wallbrown et al., 1975). As explained by Vance and Wallbrown (1978) the Wherry and Wherry method employs a principal factor analysis with use of Minres "cleanup" (ignores the diagonal of the correlation matrix) in order to eliminate problems of obtaining communality estimates; the resulting matrix is rotated by the varimax criterion in order to assign variables to clusters. A theoretical correlation matrix is then constructed with corrected communalities in the diagonal, and a multigroup analysis is done. Though this method is more complex than other factor analyses done on the WISC-R, it does have some theoretical merit.

Wallbrown et al. (1975) analysed the standardization sample for the WISC-R at each of the 11 age levels using the Wherry and Wherry Hierarchical factoring method. Specifications were preset so that only two group factors (*v:ed* and *k:m*) would be extracted along with the general factor. All of the subtests loaded on the *g* factor, ranging from a low of .44 to a high of .74, with a mean of .58. Further, the magnitude of the *g* - loadings was relatively stable across all 11 age groups. It should be noted that with this method of factoring *g* is defined as the "pervasive overlap among diverse intelligence assessors" (Vance & Wallbrown, 1977, p. 700). At the second level in the hierarchy (group factors), there appeared to be a clear distinction between two factors based primarily on the distinction between Verbal and Performance subtests.

Wallbrown et al. (1975) suggested that the verbal factor is akin to Vernon's *v:ed* factor while the performance factor is akin to Vernon's *k:m* factor. In general, all of the Verbal Scale subtests tended to load on the *v:ed* factor (mean of .32) with a few exceptions: first, the Digit Span subtests did not load on the *v:ed* factor for the 8.5 or 10.5 year old group; second, the Arithmetic subtest did not load on the *v:ed* factor for the 8.5 year old group; third, the Information and Comprehension subtests did not load on the *v:ed* factor for the 12.5 year old group. Further, the Coding subtest had small loadings on the *v:ed* factor for the age groups of 11.5, 12.5, and 16.5 years.

Meanwhile, Wallbrown et al. (1975) found that all of the Performance Scale subtests had loadings on the *k:m* factor (mean of .32) at each of the 11 age levels; Coding was the only exception. Table 3 provides the mean loadings for the general factor as well as the *v:ed* and the *k:m* factors collapsed across the 11 age levels. Because Wallbrown et al. (1975) used the standardization sample for the WISC-R as their sample, the results presented in Table 3 are more-or-less comparable to those presented in Table 1 which gives Kaufman's varimax rotated factors for the same sample.

From Table 3, the percentage of variance attributable to each of the three factors is given: about 36% of the total variance is attributable to *g* while 6% is attributable to *v:ed* and 5% is attributable to *k:m*. Interestingly, the specificity of each WISC-R subtest in the hierarchical model, ranges from .28 to .34 with a mean of .31. Thus, each subtest would appear to have adequate specificity in order to be interpreted individually (as per Kaufman's 1975 and 1979 suggestion of .25 as a minimum criteria). Thus, one can see that approximately 47% of the total variance is attributable to the three factors (*g*, *v:ed*, and *k:m*) while 31% is attributable to specific variance, and about 22% to error variance. From the hierarchical model, profile analysis appears to have some justification. However, at the end of this section, a pragmatic approach to interpretation of the WISC-R as given by Blaha and Wallbrown (1984) will be reviewed

Table 3

Mean Factor Loadings on *g, v:ed, and k:m* in the Hierarchical Model of Wallbrown, Blaha, Wallbrown, and Engin (1975) Based on the Analysis of the WISC-R Standardization Sample

Wallbrown et al. (1975) Hierarchical Factor Structure of the WISC-R Tests	Mean* Loadings For *g*	Mean* Loadings For *v:ed*	Mean* Loadings For *k:m*
Information	.70	.35	.02
Similarities	.70	.30	.07
Arithmetic	.59	.30	.01
Vocabulary	.74	.41	-.03
Comprehension	.65	.30	.05
Digit Span	.43	.23	.00
Picture Completion	.58	.01	.31
Picture Arrangement	.55	.06	.23
Block Design	.70	-.02	.40
Object Assembly	.59	-.09	.40
Coding	.39	.13	.09
Mazes	.44	-.04	.28
Percent of Total Variance Accounted for by Mean Loadings	36.0	6.0	5.0

* The loadings reported here are the mean across all 11 age groups for each variable This Table has been adapted from information given in Wallbrown et al.'s (1975) Tables 1, 2, & 3.

Wallbrown et al. (1975) conclude that the WISC-R has a substantial amount of construct validity: the large *g* factor supports the use of the Full Scale IQ; the *v:ed* and *k:m* factors support the use of the Verbal and Performance IQs. However, it should be noted that in this hierarchical model the specificity of the subtests accounts for almost as much variance as the general factor does.

Other hierarchical analyses have been done on several clinically referred samples. For example, Vance and Wallbrown (1977) did a hierarchical analysis of the ten mandatory subtests of the WISC-R in a sample of 169 children and adolescents who had been referred to a community agency for intellectual assessment. These children ranged in age from 6-3 to 15-11 years and had a mean Full Scale IQ of 75.5 with a range from 40 to 119. The procedure used was similar to that used by Wallbrown et al. (1975); the Wherry and Wherry factoring method was used and the number of primary factors to extract was preset to two. However, unlike Wallbrown et al. (1975), Vance and Wallbrown (1977) found that the general factor accounted for 55% of the total variance of the subtests. On the other hand, the two group factors of *v:ed* and *k:m* were again distinct and defined by the Verbal Scale and Performance Scale subtest respectively, and accounted for a similar amount of variance as had been found by Wallbrown et al. (1975): *v:ed* accounting for about 5% of the total variance and *k:m* accounting for about another 5% of the total variance.

This discrepancy between the the percentage of variance accounted for by *g* as found by Vance and Wallbrown (1977) as opposed to Wallbrown et al. (1975) is of some interest. Despite using the same factorial procedure, the loadings on *g* for each of the subtests was higher in the Vance and Wallbrown study. The most likely reason for this has to do with the extreme range of IQs (FSIQ ranging from 40 to 119) in the sample used by Vance and Wallbrown (1977). Because of this extreme range, the correlations between subtests were probably higher in this clinical sample than in the standardization sample; as such, the factor loadings for the general factor could be expected to be higher.

Another hierarchical analysis of the WISC-R was done on a sample of 85 learning disabled children with ages varying form 6-3 to 14-1 years (Blaha &Vance, 1979). All 12 WISC-R subtests were used in this analysis, and the range of Full Scale IQs was from 82 to 123 with a mean of 95. The number of factors to extract at the second level was not preset to two, but rather left open--eventually allowing three factors to emerge. In an interesting twist, the general factor only accounted for 20% of the total variance. In contrast, the *k:m* factor accounted for about 11% of the total variance, the Verbal Comprehension factor accounted for 7%, and the Freedom From Distractibility accounted for 6%. In typical fashion, the Freedom From Distractibility factor was defined by the Arithmetic (.48), Digit Span (.48), and Mazes (.33) subtests.

The puzzling nature of the general factor was further complicated by the results from a study by Vance and Wallbrown (1978) who used a clinical sample of 150 black children. In their hierarchical analysis of the factor structure of the WISC-R they found that the general factor only accounted for about 6% of the total variance. The loadings on this general factor were weak ranging from .16 to .35 with a mean of .24. Conversely, the *v:ed* and *k:m* factors showed a corresponding increase in the amount of total variance with the *v:ed* factor accounting for 22% of the subtest variance and the *k:m* factor accounting for 16%. Vance and Wallbrown (1978) attempted to explain the weak general factor by suggesting that the sample was very homogeneous in nature. These fluctuations on the general factor are rather disturbing; though they could be due to sampling fluctuations or actual differences, they may also have resulted from problems in the Wherry and Wherry factoring method. As such, the findings from the Wherry and Wherry factoring method should probably be interpreted cautiously and with some hesitation.

In yet another hierarchical study, Wallbrown and Blaha (1979) look at the WISC-R factor structure in a sample of 112 reading disabled children ranging in age from 9-2 to 13-7 years. They had a mean Full Scale IQ of 101.6 with a range from 86 to 129. The general factor only accounted for 17% of the total variance, and the mean factor loading on the general factor was .39. To complicate matters, the factors from the second level of the hierarchy were not sufficient to account for all the subtest variance and two "subgeneral" factors were needed. The *v:ed* factor had moderate loadings from all the Verbal Scale subtests (mean of .35) and accounted for 7% of the total variance; the *k:m* factor accounted for another 6% of the total varianee and had moderate loadings from all of the

Performance Scale subtests (mean of .27) except Coding which had a loading of -.05. Further, four factors at a lower level of the hierarchy were found. These were interpreted by Wallbrown and Blaha (1979) as follows: a Verbal Comprehension factor (Information, Similarities, Vocabulary, and Comprehension) accounting for 8% of the total variance; a Freedom from Distractibility factor (Digit Span and Arithmetic) accounting for 4% of the total variance; a Spatial factor (Block Design, Object Assembly, Picture Completion, and Mazes) accounting for 3% of the total variance; and finally, a Quasi-Specific factor (Picture Arrangement and Coding) accounting for 4% of the total variance.

Wallbrown and Blaha (1979) concluded that although the factor structure for the first two levels of the hierarchy for this sample was comparable to the hierarchy found for other groups, the full hierarchical structure of this reading disabled group was more complex than it was for normal children (i.e., the need to extract 4 primary, or third level factors, to explain the variance). Thus, with learning disabled children, the variance from the general factor appeared to be spread out more to the second and third levels of the hierarchy (if the Wherry and Wherry factoring method is indeed correct).

Because of the apparent confusion involved in comparing the discrepant findings of different hierarchical studies, Blaha and Wallbrown (1984) did a review of all the hierarchical studies of the WISC and WISC-R. As a guiding principal, they interpreted each study through the model proposed by Vernon. Despite contradictions found from various studies, Blaha and Wallbrown endorsed Vernon's suggestion that in heterogeneous samples the general factor (*g*) accounts for approximately 40% of the total variance while the two group factors (*v:ed* and *k:m*) account for about another 10%, and that minor factors (similar to Thurstone's primaries) account for an additional 10% of the total variance. The following diagram (Figure 3) has been taken from Blaha and Wallbrown (1984) and is their depiction of the hierarchical organization of the WISC-R as synthesized from their review of 13 hierarchical studies of the WISC and WISC-R.

Figure 3

The Hierarchical Structure of the WISC-R as Conceived of by Blaha and Wallbrown (1984) Using Vernon's Model of Intelligence

Figure 3 has been removed because of the unavailability of copyright permission.

* The above diagram has been modified and reproduced from Blaha & Wallbrown (1984).

So what we see in a hierarchical analysis is a large general factor (questionable in clinical samples), and usually the emergence of at least a verbal factor and a performance / spatial factor; however, the loadings on the two latter factors are substantially lower than in varimax rotated solutions because the heavy contribution of *g* has been removed prior to the emergence of the Verbal and Performance factors. Thus, especially with special populations, it may be possible that the hierarchical structure is not as stable as the structure that emerges from the principal axes factoring followed by varimax rotations. Overall though, the hierarchical analysis supports Wechsler's (1974) contention that the WISC-R measures general intelligence, and also his division of the subtest into the Verbal and Performance Scales.

Joint Factor Analysis of the WISC-R with Other Tests

A few studies have been conducted where the WISC-R has been jointly factored with other types of intelligence and achievement tests. The importance of such a procedure may not be immediately obvious to some. One proponent of joint factoring of new tests with older more established ones is Kieth (1987). He suggests that factor analysis of new instruments in combination with existing, well understood instruments, should become more prevalent; the reason for this being that such a procedure leads to a better understanding of what the new instrument measures. Put more simply, when a new test like the Stanford-Binet IV is jointly factored with an existing test that has a known factor structure (like the WISC-R), the resulting factor structure can be compared to the separate factor structures. Furthermore, such a procedure gives a better indication of which subtests from each of the tests are measuring similar constructs (factors). Because the WISC-R and Stanford-Binet IV will be jointly factored in this thesis, a brief review of joint factor analyses of the WISC-R and other tests is appropriate. Such a review will clarify what typically happens to the WISC-R factor structure when other tests are added.

<u>Factoring the WISC-R with Other Intelligence Tests</u>

DeHorn and Klinge (1978) performed a joint analysis of the WISC-R and the Peabody Picture Vocabulary Test (single score) in a psychiatric population. Using a principal components analysis three factors were rotated. The emergent factors were the same as the typical three-factor WISC-R structure except that loadings were a little higher than those found by Kaufman (1975) because principal components as opposed to principal factoring had been used. The Peabody Picture Vocabulary Test loaded (.75) on the Verbal Comprehension factor, suggesting that it has a verbal content of some sort.

Sutter and Bishop (1986) did a joint factor analysis of the WISC-R and the Illinois Test of Psycholinguistic Abilities (ITPA) in a sample with three types of children: learning disabled, emotionally disturbed, and a control group. Although the eigenvalue-one criterion suggested six factors, only three were rotated. Under these constraints, the familiar WISC-R pattern did not emerge. Only the Verbal Comprehension factor remained intact. The Block Design and Object Assembly subtests did not load on any of the three factors. Taken as a whole, the authors suggested that the three factors represent linguistic, auditory memory, and visual-motor ability. As it stands, this study failed to find the core factors from the WISC-R. However, in a serious methodological short-coming, only three of the factors that had eigenvalues greater than one were rotated. Quite possibly, then, the under-rotated matrix may have resulted in the distortion of the three typical factors usually found on the WISC-R. Therefore, Sutter and Bishop's (1986) results need to be replicated before any conclusions can be drawn from their methodologically flawed study.

Two final studies involving the joint analysis of the WISC-R with Kaufman Assessment Battery for Children (K-ABC) will be reviewed. These two studies are perhaps the most informative of those dealing with the joint analysis of the WISC-R with other intelligence tests. Before moving directly to these two studies, a little preliminary background is need on the K-ABC. This brief discussion of the K-ABC is not complete, and is intended only to provide an illustration of how important it can be to jointly factor new tests along with the WISC-R.

The K-ABC is designed to measure Achievement as well as Sequential and Simultaneous processing skills. Several investigators have looked at the factor structure of the K-ABC (e.g., Kieth, 1986; Valencia & Rankin, 1986). Although the factor structure does generally seem to substantiate the three major scales, Kieth (1985) has suggested that the K-ABC may provide measures of verbal reasoning, nonverbal reasoning, and verbal memory rather than of achievement, and simultaneous and sequential mental processing. Kaufman and McLean (1987) also acknowledge that such renaming of the K-ABC factors may be equally justifiable depending on one's theoretical orientation. From the joint analysis of the WISC-R and K-ABC, much has been learned, especially that both tests may in essence be measuring the same three constructs.

Kaufman and McLean (1987) did a joint factor analysis of the WISC-R and K-ABC on a sample of 212 "normal" children. All eleven of the WISC-R subtests (excluding Mazes) ar 3 of the K-ABC subtests were included in the analysis. Principal factor analysis was conducted with squared multiple correlations as initial estimates of communalities--these were then iterated to convergence. Furthermore, varimax rotation of factors was used. Using both the eigenvalue-one, and the scree test criteria, three factors were indicated for the joint analysis. The results indicated that the three typical factors found on the WISC-R and the three found on the K-ABC had a great deal in common. The following statement by Kaufman and McLean (1987) indicates the relationship between the WISC-R and K-ABC factors:

> The first factor represents a merger of the WISC-R Verbal Comprehension and K-ABC Achievement, the second a blend of Perceptual Organization and Simultaneous Processing, and the third a combination of Freedom From Distractibility and Sequential Processing. Only Coding did not load as predicted, loading below .40 on all three factors and loading about equally well on Factor 2 and 3 (pp. 112-113).

As noted by Kaufman and McLean (1987), the congruence between the WISC-R factors and the K-ABC factors does not answer the question about the specific construct that are being measured by each factor--this is determined largely by the researchers theoretical orientation. A further warning by Kaufman and McLean is that although the factors from the two test merge, this convergence does not always extend to the subtest level even though the underlying constructs are the same.

Another joint factor analysis of the WISC-R and K-ABC was done by Kaufman and McLean (1986). This time, however, the sample was a group of 198 children who had been referred for or diagnosed as having learning disabilities. Along with the full battery of the K-ABC, all of the WISC-R subtest except Mazes were given. Principal factor analysis was conducted with squared multiple correlations initially inserted in the diagonals--followed by iteration to convergence. The scree test again indicate that three factors were appropriate, and these were rotated by the varimax criterion. Once again the K-ABC Achievement factor and the WISC-R Verbal Comprehension factor merged into the first factor. The second factor was a combination of the WISC-R Perceptual Organization factor and the K-ABC Simultaneous factor. Finally, the third factor involved the K-ABC Sequential factor and the WISC-R Digit Span subtest.

The key thing to note from Kaufman and McLean's work (1986, 1987) is that despite differences in theoretical construction, the joint analysis of the WISC-R and K-ABC resulted in the merging of the factors from the two tests. That is, the core factors from the WISC-R were not changed with the addition of subtests from the K-ABC. This finding is encouraging, in that the WISC-R factor structure appears to be robust even when it is factored jointly with other tests. This has important implications for the joint analysis of the Stanford-Binet IV and the WISC-R as planned in this thesis. If additional fac beyond the three normally found on the WISC-R emerge, some support for a revised Cattell-Horn model, or modified version of Vernon's model as proposed in the Stanford-Binet IV may be validated.

G. Construction of the Stanford-Binet IV (1986)

In summarizing the impetus for the development of the Stanford-Binet IV, Thorndike et al. (1986) cite three major influences: (a) social and cultural changes, (b) research in cognitive psychology, and (c) difficulty in arriving at good diagnostic analysis when only a general intellectual factor is reported. Furthermore, intelligence tests are no longer being used only for classifying children in terms of ability. Rather, as seen in the opening chapter of the *Technical Manual* for the Stanford-Binet IV, Thorndike et al. (1986) acknowledge that intelligence tests are being used more for prediction as well as for providing guidelines for how to adapt instruction to the characteristics of the individual learner; in adapting instruction to individuals (individual educational plans) it is most often believed that measurement of specialized abilities provides a basis on which educational plans can be constructed to correct weak areas while simultaneously using strong areas[22]. Although Thorndike et al. suggest that measurement of specialized abilities has not been highly successful, they do suggest continued attempts: "It is important that we continue to seek such interactions and develop the measuring devices that will make such a search possible" (1986, p. 6). Thus, the Stanford-Binet IV is an attempt to make the out-dated Binet type scales into an instrument on which specialized abilities are included--abilities that are useful in developing individual educational plans.

Structure of the Stanford-Binet IV

Thorndike and his colleagues explicitly outline the theoretical model (see Figure 1) that helped guide the construction of the Stanford-Binet IV. Close examination of Figure 1 reveals that both Cattell's and Vernon's models have been synthesized into a single model. The general theoretical structure of the Stanford-Binet IV envisions a three-level hierarchical model. At the apex of the hierarchy is *g* , described as a general reasoning

[22] Thorndike et al. (1986) are largely referring to the use of profile analysis in order to construct educational plans that are suited to the individual child.

factor. The second level consists of three "broad" factors: Crystallized Intelligence, Fluid Intelligence, and Short-Term Memory. On the third level one finds factors that are more "specific" than on the second level: Verbal Reasoning and Quantitative Reasoning stemming from Crystallized Intelligence, and Abstract/Visual Reasoning stemming from Fluid Intelligence. It is these three "specific" factors along with Short-Term Memory that the Stanford-Binet IV is purported to measure.

Despite the theoretical three-level hierarchical model that is presumed for the Stanford-Binet IV, its structural layout suggests that it measures (a) general intelligence--the Composite score, and (b) four "Areas" or group factors--Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning, and Short-Term Memory. The second level of the hierarchy as depicted in Figure 1, is "invisible" and no score on either Crystallized or Fluid Intelligence is reported. However, the actual structure of the Stanford-Binet IV does have a third level which is made up by the subtests. Altogether there are 15 subtests on the Fourth Edition. The Vocabulary, Comprehension, Absurdities, and Verbal Relations subtests comprise the Verbal Reasoning Area; the Pattern Analysis, Copying, Matrices, and Paper Folding & Cutting subtests comprise the Abstract/Visual Reasoning Area; the Quantitative, Number Series, and Equation Building subtests comprise the Quantitative Reasoning Area; and the Bead Memory, Memory For Sentences, Memory For Digits, and Memory For Objects subtests comprise the Short-Term Memory Area.

The choice of areas to measure on the Fourth Edition was not based solely on factor analytic studies or on an existing model of intelligence. Rather, as in previous Binet scales, pragmatic concerns helped guide the selection of areas to include on the new test (Thorndike et al., 1986):

> To some extent, the selection of the theoretical model was influenced by consideration of the ways that clinicians and educators have used previous editions of the Stanford-Binet. They have used it most frequently to

> identify gifted students, to assess the cognitive abilities of mainstream students who were having difficulty learning, and to identify the mentally retarded. Since the results of the test are most frequently used, together with other information, to make recommendations for educational intervention, the authors wanted to assess the kinds of cognitive abilities that years of research have shown are correlated with school progress (p. 9).

The structure and organization of the Stanford-Binet IV is more like the Wechsler scales than like the Stanford-Binet Form L-M. That is, like the Wechsler scales, the Fourth Edition collects all items of a similar type onto a subtest rather than placing six items at each age level as on the L-M. Does this mean, as suggested by Vernon (1987), that the Fourth Edition is no longer a "Binet type" or "Stanford-Binet type" test? Probably not. What it means is that the cumbersome problematic age scale seen in the older Stanford-Binet tests has been updated to bring the Fourth Edition more in line with the needs of educators. What about the inclusion of group factors on the Stanford-Binet IV-- does this suggest that it is not a "Stanford-Binet type" scale? Indeed it does not. Rather, what the new Binet scale has done is make explicit the formerly implicit group factors. Those in need of convincing can review the numerous factor analytic studies conducted on the Stanford-Binet scales; the four Areas included on the Stanford-Binet have all been found as group factors in earlier factor analyses, though they may have had different names: Verbal, Nonverbal/Visual-Spatial, Number, and Memory. Further, as was seen in the review of the 1937 Forms L and Form M, as well as in the 1960 Form L-M, Terman and Merrill (1973) did admit to the possibility of group factors on the Stanford-Binet. Thus, what the Fourth Edition has done is make explicit the implicit, and done away with the pretense of hiding behind a unitary measure of general intelligence.

Although the inclusion of four Areas or group factors on the Stanford-Binet IV has considerable support from a variety of sources (e.g., factor analyses on the Stanford-Binet scales and Terman & Merrill's admission of their possible existence), this is a revolutionary step towards bringing the Stanford-Binet scales into a modern world. The inclusion of group factors on the Fourth Edition indicates that its authors recognized that the previous Binet Scales had minimized intra-individual differences in abilities. Conversely, current use of intelligence tests seeks maximization of intra-individual differences on which educational plans can be built. Further, the inclusion of four reasoning areas was made in response to criticisms that earlier Binet scales were too verbal, and therefore often underestimated the potential of children with language handicaps or stronger nonverbal than verbal abilities.

Yet, to be truly a Binet-type scale, the Fourth Edition must remain committed to the measurement of general intelligence, as this has been the hallmark of the Binet-type scales. Once again, the Fourth Edition does maintain a strong commitment to general intelligence. As outlined by Thorndike et al. (1986) they conceive of *g* in a manner similar to previous Stanford-Binet constructors (i.e., Terman & Merrill, 1937, 1973). According to Thorndike et al. (1986), "It is our strong belief that the best measure of *g* --and consequently broadly effective prediction--will stem from a diverse set of cognitive tasks that call for relational thinking in a diversity of contexts" (p. 6). Furthermore, Thorndike et al. emphasize the importance of *g* in that it is usually the best predictor of performance in both educational and vocational settings. They conclude their first chapter with the following statement: "Still, the general ability factor, *g* , refuses to die. Like a phoenix, it keeps rising from its ashes and will no doubt continue to be an enduring part of our psychometric theory and psychometric practice" (Thorndike et al., 1986, p. 6). Clearly, the theory behind the Stanford-Binet IV attributes the same prominence to *g* as previous Stanford-Binet scales.

Continuity and Changes From Earlier Stanford-Binet Scales

It has already been noted that the concept of general intelligence has retained the same prominence on the Fourth Edition as on earlier Stanford-Binet scales; also, it was noted that the inclusion of explicit group factors is not really a new idea on the Stanford-Binet scales, but rather a bringing of the implicit group factors found in factor analyses of the older scales out into the open. The only real change that has been mentioned so far is the abandoning of the problematic age-scale format of previous Stanford-Binet scales in favor of grouping items according to type; items of the same type (e.g., Vocabulary) have been grouped into a subtest "so that an individual's cognitive functioning could be more efficiently evaluated" (Thorndike et al., 1986, p. 8). The use of subtests as opposed to the older age scale format is a vast improvement. Children who are strong in one area do not need to suffer through repeated failure on other areas; or conversely, there is less fumbling than using an unorthodox procedure (as suggested by Vernon, 1987) of administering all similar type items from Form L-M at the same time. Further, by grouping similar items onto subtests, the reliability of the subtest is greatly improved over individual item reliability.

Two further items that indicate a continuation of the earlier Stanford-Binet scales deserve mention. First, the Fourth Edition has maintained the adaptive testing format that was used on previous Stanford-Binet scales. Second, like its predecessors, the new Stanford-Binet is designed to be applicable for measuring cognitive abilities in examinees from age two through to adult; however, as noted by Vernon (1987) the use of the Fourth Edition may be questionable beyond 18 years of age in that no norming was done.

Preparing Subtests for the Stanford-Binet IV

As pointed out by Thorndike et al. (1986) the decision of which item types to retain from the older Binet scales was made using several criteria: (a) the item had to measure either Verbal Reasoning, Quantitative Reasoning, Abstract/Visual Reasoning, or Short-Term Memory; (b) items could be scored reliably; (c) items were to be free from ethnic or gender bias; (d) the items would function over a wide age range. The item types retained from Form L-M included Vocabulary, Comprehension, Picture Absurdities, Opposite Analogies, Paper Folding and Cutting, Copying, Repeating of Digits, Memory for Sentences, Copying a Bead Chain from Memory, Similarities, Identifying Parts of the Body, Form Board, and several quantitative types of items (Thorndike et al., 1986). Thus, according to Thorndike et al. (1986) 9 of the 15 subtests on the Stanford-Binet IV have evolved from Form L-M. Furthermore, Picture Memory, Memory for Objects, and Arithmetic, were also to be found on the 1937 Stanford-Binet scales (Forms L or M). Stretching things, it seems that possibly up to 11 of the 15 subtest have some similar item type on some revision of the Binet scales; this leaves Pattern Analysis, Matrices, Number Series, and Equation Building as relatively new additions to the Binet family of tests.

Because many of the item types were not available for use with a wide age range, the authors had to construct similar item types for tryout. Besides the item types taken from earlier Stanford-Binet scales, Thorndike et al. (1986) report that they reviewed the research on the measurement of cognitive abilities in order to find new item types that would be useful in measuring the Abstract/Visual, Quantitative, and Short-Term Memory Areas (and hence reduce the verbal nature of the Fourth Edition).

Several stages of field testing were done. Initially 29 item types were included as possibilities; this was subsequently reduced to 23 and then to 17 types. After the second field trial the number of item types was reduced to 15, and these were retained on the final version of the Fourth Edition. Because of the nature of the item types, not all subtests are applicable for each age level. Therefore, the subtests are arranged in levels (known as "entry levels") which are designated by letters. The number of levels for each subtest varies from 7 to 20; those with fewer levels being applicable to a narrower age range and those with more levels being applicable to a wider age range.

ı L-M included Vocabulary, Comprehension, Picture Absurdities, Opposite
, Paper Folding and Cutting, Copying, Repeating of Digits, Memory for
Copying a Bead Chain from Memory, Similarities, Identifying Parts of the
m Board, and several quantitative types of items (Thorndike et al., 1986).
ırding to Thorndike et al. (1986) 9 of the 15 subtests on the Stanford-Binet IV
ʼed from Form L-M. Furthermore, Picture Memory, Memory for Objects, and
:, were also to be found on the 1937 Stanford-Binet scales (Forms L or M).
things, it seems that possibly up to 11 of the 15 subtest have some similar item
me revision of the Binet scales; this leaves Pattern Analysis, Matrices, Number
l Equation Building as relatively new additions to the Binet family of tests.
cause many of the item types were not available for use with a wide age range,
; had to construct similar item types for tryout. Besides the item types taken
ːr Stanford-Binet scales, Thorndike et al. (1986) report that they reviewed the
ı the measurement of cognitive abilities in order to find new item types that
ıseful in measuring the Abstract/Visual, Quantitative, and Short-Term Memory
hence reduce the verbal nature of the Fourth Edition).
ʼeral stages of field testing were done. Initially 29 item types were included as
:s; this was subsequently reduced to 23 and then to 17 types. After the second
he number of item types was reduced to 15, and these were retained on the final
the Fourth Edition. Because of the nature of the item types, not all subtests are
for each age level. Therefore, the subtests are arranged in levels (known as
ls") which are designated by letters. The number of levels for each subtest
ı 7 to 20; those with fewer levels being applicable to a narrower age range and
more levels being applicable to a wider age range.

underestimated while others were highly overestimated. The most under-represented groups were children whose parents had less than a high school graduation, children whose parents were operators/fabricators, and children whose parents were precision production workers; the most overestimated were children whose parents were college graduates and whose parents were in managerial/professional roles.

Authors like Slate (1986) have severely criticized the standardization sample of the Stanford-Binet IV as being unrepresentative of the U.S. population. Osberg (1986), however, attempted to rebut this criticism by pointing out that the standardization sample for the Fourth Edition is the largest that has ever been used on any individual intelligence test. Even so, this does not negate the fact that the standardization sample was not representative of the U.S. population. Even though a proportional weighting procedure was used to correct imbalances, the restricted number of individuals in some categories may not have ensured a good representation of examinees in these categories; by applying differential weighting, this lack of representativeness could have been further compounded. Although differential weighting is probably the best way to deal with the problem, it does not ensure that the norms will be representative of the U.S population. Vernon (1987) also addresses the problem of using differential weighting. In an unexpected fashion he suggests that "corrections were attempted for the much too large proportion of college graduates in the later group; but I would prefer myself to discard it" (p. 255). However, discarding the weighting procedure would seem unadvisable in that it would guarantee that the sample had serious problems in terms of representativeness of the U.S. population.

On a more pragmatic level, Vernon notes that approximately 1/4 of the standardization sample was non-white. As such, he suggests that for English-speaking Canadians, the mean would probably be 104 rather than 100. This consideration is one to keep in mind when working with Canadian children; a Composite score of 100 may be slightly below the mean for Canadian children.

Deviation Scores on the Stanford-Binet IV

For the individual subtests, Standard Age Scores (SAS) with a mean of 50 and standard deviation of 8 are used. Actual means and standard deviations in the standardization sample come close to these theoretical values: the means for the subtests range from 49.5 to 50.5 while the standard deviations range from 7.3 to 8.5. For the Area and Composite scores the SAS have a mean of 100 and standard deviation of 16. Slate (1986) has criticized the retention of a standard deviation of 16 on the Fourth Edition; he suggests that the authors should have used a standard deviation of 15 like that used on the Wechsler scales. However, a standard deviation of 16 has been used on the older Stanford-Binet tests, and the choice of retaining this figure can be equally justified in terms of continuity with older Stanford-Binet scales. Furthermore, Thorndike et al. (1986) have provided a table by which deviation scores with a standard deviation of 16 can be converted to scores with a standard-deviation of 15.

Reliabilities

As a general trend, Thorndike et al. (1986) note that the reliabilities tend to be slightly higher at the older age levels. Looking at the reliabilities across the different ages and subtests, it can be seen that there are considerable variations at different age levels. For example, on the Vocabulary subtest, the lowest reliability is for six-year-old children (.78) while the highest reliability is for the 18-23 year-old group (.94). In general, most of the reliabilities are in the .80s and .90s with relatively few in the mid or high .70s (with the exception of Memory for Objects which has most of its reliabilities, regardless of age level, in the .70s). Excluding Memory for Objects (which has a median reliability across ages of .73) the median reliabilities across ages 2 through 17 for each subtest are in the mid .80s to low .90s (ranging from .83 to .94). Thus, for most subtests, the Stanford-Binet IV appears to be fairly reliable across appropriate ages.

The reliabilities of the Area Scores are dependent on the number of subtests taken by the examinee in each of the Areas; the more subtests taken, the higher the reliabilities. In general, these reliabilities are in the high .80s and the .90s. The clearest trend of reliabilities is seen in the Composite SAS which has reliabilities ranging from .95 to .99 depending on the age level; once again the pattern is for older ages to have more reliable scores.

Confirmatory Factor Analysis on the Standardization Sample

After constructing the Stanford-Binet IV and administering it to the standardization sample, the test scores were factor analyzed with "a variant of confirmatory factor analysis" (Thorndike et al., 1986, p. 52). In this procedure, a general factor was extracted first, and then group factors were extracted "that generally accounted for the residual correlation among tests assigned to a given content area" (Thorndike et al., 1986, p. 52). According to Thorndike et al. the extraction of group factors was done in the following order: Verbal, Memory, Quantitative, and Abstract/Visual. Unfortunately, Thorndike et al. did not specify the criterion they used to determine the number of factors to extract.

In a personal written communication with Dr. R. L. Thorndike (June 2, 1987), he elaborated on some of the finer points of the confirmatory factor analysis used on the Stanford-Binet IV. He admits that his procedures were "not mathematically elegant" and that no indices of goodness of fit between the data and the hypothetical model (Figure 1) were available. In his comments, he indicated that his background in factor analysis dates back 50 years, and as such, he used an "old-style Thurstone centroid analysis". He makes the following statement concerning the way the factoring was done:

> I'm afraid my procedure of analysis was slightly idiosyncratic. Basically, the variables were assigned to their pre-specified clusters. Then by a series of iterations the G-factor weights were determined such that the cross-cluster correlations were minimized. That is, for each variable the sum of

> its correlations with all of the variables from the other clusters was made as nearly zero as possible.
>
> Subsequent factors were fitted to the matrix of residuals, after the correlations due to the G-factor had been removed. This was done one factor at a time, starting in each instance with the Verbal cluster of variables (R. L. Thorndike, personal communication, June 2, 1987).

Basically, the general factor was extracted first and then the residual matrix was factored to yield consecutive factors. Termination of factoring was done in an arbitrary manner by inspecting the size of the residual matrix left over after the previous factor was removed.

A key point to consider about Thorndike et al.'s (1986) confirmatory factor analysis is that the median correlations which were used were obtained from pooled age groups. Dr. Thorndike indicates that the decision to pool the correlation data from several age groups was done in order to obtain a larger sample size and therefore more stable correlations. Further, "The variables used in the analyses for a particular range of ages were those for which data existed for all or part of the persons in that age range" (R. L. Thorndike, personal communication June 2, 1987). This method of using median correlations did away with the problems of missing data. For example, looking at the correlations for each age group (reported in the *Technical Manual* pp. 110-126) one can see that a different number of children took different subtests. By taking the median intercorrelation across a range of ages, the result is that the correlations between any two subtests are based on different number of subjects.

Although Thorndike et al. (1986) report the results of a factor analysis on the median correlations across all age levels, such a procedure is severely lacking in that not all children took the same subtests; a possible confounding variable is developmental change in the factor structure with changing age (see Kieth, 1987). Given that changes do occur in terms of the content of subtests at different ages (e.g., the Copying subtest changes from stacking of blocks at younger ages to copying of figures with a pencil and paper at Level

G) as well as the number of subtest for a given Area, it would not be unexpected for some variation in factors to occur depending on the age group studied. Perhaps more importantly, Thorndike et al.'s (1986) choice of age groupings does not match the entry levels assigned to given subtests. For example, children age 7 fall below entry Level I which is required for the Matrices and Number Series subtests--yet the 7 year-old children have been placed in an age group for which the Matrices and Number Series subtests were included in the analysis.

Because of possible confounding of developmental trends, only the factor analyses that Thorndike et al. (1986) carried out on specific age groups will be discussed. Thorndike et al. report separate analyses for three age groups: 2 through 6 years, 7 through 11, and 12 through 18-23 years. The result from each of these groups will be discussed separately; to facilitate comparison with the results from this thesis, the three tables representing Thorndike et al.'s findings have been reproduced. Only subtests with loadings of .30 or greater will be considered in the naming and identification of the factors found by Thorndike et al., though subtests with loadings between .20 and .30 will be discussed if they were expected to load on a given factor. One must bear in mind, however, that the age groupings used by Thorndike et al. may have been rather poorly chosen and thus introduce some additional error into the analyses.

Age Group 2 Through 6 Years

Table 4
Confirmatory Factor Structure for the Stanford-Binet IV for Age Group 2 Through 6 in the Standardization Sample (Thorndike, Hagen, and Sattler, 1986)

Table 4 has been removed because of the unavailability of copyright permission.

* The above table has been reproduced from the *Technical Manual* from the Stanford-Binet IV (Thorndike, Hagen, & Sattler, 1986, p. 55).

As can be seen from Table 4, for the age group 2 through 6, Thorndike et al. (1986) only extracted two factors beyond *g* . Worth noting is the fact that all subtests at this age have fairly consistent loadings on the general factor which accounts for 42% of the total subtest variance. The first group factor is Verbal in nature and has loadings from Vocabulary, Comprehension, and Memory for Sentences. Given that Memory for Sentences has verbal content, and with the absence of Memory for Digits at this age level, it is not surprising that a Memory factor does not emerge and consequently that Memory for Sentences loads on the Verbal factor. In contrast, Absurdities--which is on the Verbal Reasoning Area--has only a slight loading on the Verbal factor (.26); perhaps this occurs because Absurdities has a strong loading on the general factor.

On the second factor, which Thorndike et al. have called Abstract/Visual, it is interesting that Bead Memory has the highest loading of any subtest. In discussing this subtest, Thorndike et al suggest that Bead Memory deals with abstract/visual materials that are presented simultaneously; however, this interpretation is specific only to the age 2 through 6 group as Bead Memory did not load on any group factor for the other two age groups. Further, in a personal communication with Dr. R. L. Thorndike (June 2, 1987) he indicated that Bead Memory not only seems to require simultaneous processing, but also the use of strategies for perceiving and organizing materials. The only other loading on the Abstract/Visual factor that is higher than .30 is Copying. It is of further interest that the Quantitative subtest has a slight loading on the second factor; Thorndike et al. imply that this occurred because there were not sufficient similar subtests for a Quantitative factor to emerge. Finally it is puzzling that the Pattern Analysis subtest does not load significantly on the Abstract/Visual factor. Because Pattern Analysis does not load substantially on the second factor, the "Abstract" designation is less substantiated at this age. Taken together, the pattern of loadings on the second factor clearly indicates that it has a visual nature, and may be akin to Broad Visualization (Gv) as outlined by Horn and Cattell (1966)[23].

From the confirmatory factor analysis conducted by Thorndike et al. (1986) it appears that the grouping of subtests onto four Areas for young children (ages 2 through 6) is not supported. Rather, at best there appears to be only two group factors.

23 Technically the term "Broad" comes from Undheim (1976). Horn and Cattell (1966) use the term "General Visualization".

Age Group 7 Through 11 Years

Table 5
Confirmatory Factor Structure for the Stanford-Binet IV for Age Group 7 Through 11 in the Standardization Sample (Thorndike, Hagen, and Sattler, 1986)

Table 5 has been removed because of the unavailability of copyright permission.

* The above table has been reproduced from the *Technical Manual* from the Stanford-Binet IV (Thorndike, Hagen, & Sattler, 1986, p. 56).

For the age group 7 through 11, Thorndike et al.'s (1986) confirmatory factor analysis yielded a general factor and three group factors (Verbal, Memory, and Abstract/Visual). The general factor again accounts for 42% of the total subtest variance, just as in the age group 2 through 6. However, the loadings on the general factor are not as consistent as they were in the younger age group. Notable deviations include Memory for

Digits, Memory for Objects, and Copying which have the lowest loadings on *g* ranging from .54 to .57; the remaining subtest have loadings ranging from .65 to .72.

The first group factor is Verbal in nature and has loadings from the Vocabulary and Comprehension subtests. Once again the Absurdities subtest fails to load substantially on the Verbal factor. The second factor is clearly a Memory factor, with loadings from Memory for Sentences, Memory for Digits, and Memory for Objects. This factor contains most of the subtests from the Short-Term Memory Area, with the exception of Bead Memory which does not load on any group factor in this age group. Thorndike et al. (1986) named the third factor as Abstract/Visual. It should be duly noted that only one subtest (Pattern Analysis) has a substantial loading on this factor. Absurdities does have a weak loading on this factor (.20) but not strong enough to warrant inclusion. As such, this factor is a singlet and probably should not have been extracted (see Zwick & Velicer, 1986).

A potential problem for this age grouping as used by Thorndike et al. (1986) is that several of the subtests are not totally applicable for all ages included in the group. For example, Matrices and Number Series are most applicable for the average child of 8 years or greater; conversely, Copying is most applicable for the average child up to 9-11 years (see the methodology section of this thesis for a more detailed explanation). Though these subtests can be given to all children between 7 and 11 years, they are not always the most applicable; this is a pragmatic problem which may result in some distortion in the factor structure for this age grouping as used by Thorndike et al.. The inappropriateness of these subtests for some ages can also be seen in Thorndike et al.'s reported intercorrelations for each specific age (i.e., Tables B.7-B.10 in the *Technical Manual*). For example from the 415 eight-year-old children, only 176 took the Matrices subtest and 162 took the Number Series subtest. Thus, by taking the median correlation from the different ages combined into a single group (e.g., the median for 7, 8, 9, 10, and 11 year-olds) and factoring this median, Thorndike et al. have introduced unnecessary error. This could have easily been avoided by more appropriate grouping of ages based on the which tests were taken at each age.

To this point the issue of variance accounted for by each of Thorndike et al'.s (1986) group factors for this age group has not been mentioned. With the Verbal, Memory, and Abstract/Visual factors respectively accounting for 5%, 4%, and 3.5% of the total subtest variance, one can question whether they should have been interpreted at all. These group factors simply do not account for enough of the total subtest variance to warrant interpretation; hence, there appears to be limited support for the grouping of subtests onto Areas at this age level. On the other hand, the specific variance is high and suggests that profile analysis may be worthwhile, as each subtest seem to be measuring something unique and separate from *g* and the group factors.

Age Group 12 Through 18-23 Years

Table 6
Confirmatory Factor Structure for the Stanford-Binet IV
for Age Group 12 Through 18-23 in the Standardization Sample
(Thorndike, Hagen, and Sattler, 1986)

Table 6 has been removed because of the unavailability of copyright permission.

* The above table has been reproduced from the *Technical Manual* from the Stanford-Binet IV (Thorndike, Hagen, & Sattler, 1986, p. 57).

For the age group of 12 through 18-23, the confirmatory factor analysis lead Thorndike et al. (1986) to mention that "loadings on the general factor were noticeably larger than at the earlier ages..." (p. 57). This is interesting in that the work of Horn and Cattell (1966) suggests that Fluid Intelligence (the "more innate" aspect of intelligence) tends to have less of an impact as age increases while Crystallized Intelligence tends to increase with age; hence, from a developmental perspective, one would possibly expect the amount of variance accounted for by *g* to decrease as age increases, unless the test is measuring learned information. Many explanations for an increase in the variance accounted for by *g* can be postulated. For example, the Matrices and Number Series subtests are more appropriate for older children and therefore the battery may be measuring slightly different areas than at younger ages--more nonverbal reasoning type skills being included. Second, as suggested by Cattell (1971) most intelligence tests may be heavily weighted to measuring Crystallized Intelligence, and hence one would expect Gc to increase in magnitude with age--the assumption then being that *g* as measured on the Stanford-Binet IV is really closer to Gc than to either Gf (Cattell, 1971) or *g* as thought of by Spearman. Neither of these two explanations is clearly supported by the data; the high loadings on Matrices and Number Series support the former explanation while the high loadings on Vocabulary, Comprehension, and Quantitative partially support the latter explanation.

Looking at the Verbal factor for the 12 through 18-23 age group, Vocabulary, Comprehension, and Verbal Relations load as expected, supporting to some degree their inclusion on the Verbal Reasoning Area at this age. Once again the Memory for Sentences subtest has a small loading (.24) on the Verbal factor; this provides additional confirmation of this subtest's relationship to the Verbal factor.

On the Memory factor, the Memory for Digits and Memory for Objects subtests have loadings of .45 and .38 respectively. However, the loading from Memory for Sentences is less clear (.29). In fact, though Memory for Sentences does load on the Memory factor, the loading is not much higher than its loading on the Verbal factor. Bead Memory again does not load on the Memory factor. Thus, the Short-Term Memory Area is only partially substantiated at this age level.

The third group factor at this age level is Abstract/Visual. It has substantial loadings from the Pattern Analysis and Paper Folding and Cutting subtests. The Matrices subtest fails to load on this factor despite its inclusion on the Abstract/Visual Reasoning

Area; the reason for this probably being Matrices' strong loading on the general factor. Therefore, the inclusion of Matrices on the Abstract/Visual Reasoning area at this age is questionable in terms of factorial purity.

The fourth factor extracted for this age level was labelled Quantitative. This factor was a singlet with only Equation Building loading on it. Neither the Quantitative or Number Series subtests loaded on this factor though both had extremely high loadings on the general factor. As such this factor probably should not have been extracted. Further, the Quantitative Reasoning Area was not substantiated at this age level.

To conclude the review of the confirmatory factor analysis at the 12 through 18-23 age group it is appropriate to note that all of the group factors were extremely small. Hence, from a factorial viewpoint the inclusion of the four Areas on the Stanford-Binet IV is not strongly supported for this age group.

Summary of the Confirmatory Factor Analyses

What can be concluded by comparing the factor analyses done on each of the three age groups (Thorndike et al., 1986)? First, it is apparent that several of the subtests do not load on the factors that Thorndike et al. (1986) suggested they would. Lacking support from the confirmatory factor analyses, several subtests do not seem to be justifiably included on the Reasoning Areas to which they belong.

As was seen, the Absurdities subtest does not load substantially on the Verbal factor as hypothesized by Thorndike et al.; rather it has a high degree of specificity (.54-.57) for the two age groups on which it was included in the analysis. As such, the Absurdities subtest does not seem to belong on the Verbal Reasoning Area.

Similarly, the Bead Memory subtest loads on a visual factor for the youngest age group and does not load on a memory factor (or any other factor) for the two older age groups. As such, the exact relationship of Bead Memory to the rest of the battery is not clear; Slate (1986) has also suggested that the inclusion of Bead Memory on the Short-Term Memory Area is not substantiated by Thorndike et al.'s confirmatory factor analysis.

In a personal communication, Dr. R. L. Thorndike (June 2, 1987) acknowledged that Bead Memory does not "hang" together well with the other memory tests, but rather seems to involve simultaneous processing as well as use of strategies of perceiving and organizing materials. Thus, the inclusion of Bead Memory on the Short-Term Memory Area is not justified by the confirmatory factor analysis.

Another problematic subtest is Memory for Sentences. For the age group of 2 through 6, and to a lesser degree for the age group 12 through 18-23 it loaded on the Verbal factor. Thus, though it may have some relationship to Short-Term Memory, it also seems to have enough verbal content to warrant its inclusion on a Verbal factor, at least for the youngest age group where no Memory factor emerged. When the loadings from Bead Memory and Memory for Sentences are considered in conjunction, they provide a formidable basis for questioning the Short-Term Memory Area on the Stanford-Binet IV, especially for young children.

Next, the Number Series and Matrices subtests do not load on the Abstract/Visual factor but rather appear to be very strong measures of *g* , especially in the oldest group. This is not surprising in that both measure nonverbal reasoning, and have been reported as good measures of general intelligence (Eysenck, 1979); once the variance attributable to *g* is extracted, these two subtests do not seem to load strongly on the Abstract/Visual factor.

Somewhat surprising is the fact that the Quantitative subtest does not load substantially on any group factor at any age level. Instead it loads only on the general factor, with an especially high loading on *g* for the oldest age group. The reason for this pattern of loadings is not clear, unless the general factor is closer to Gc than to Gf. Given that neither the Quantitative nor Number Series subtests load on the Quantitative factor, the Quantitative Reasoning Area seems to be totally unsupported.

Finally, for all three age levels the size of the group factors was found to be quite small after the variance from *g* had been removed; because these group factors are so small the usefulness of interpreting them can be questioned. In contrast, it should be noted that most of the subtests have a substantial amount of specificity for each age group. Hence, though discussion of group factors may be a helpful heuristic for organizing and discussing the results from the Stanford-Binet IV, profile analysis based on individual subtests seems to be a more defensible approach given the findings of the confirmatory factor analyses reported by Thorndike et al. (1986).

Thorndike et al.'s (1986) conclusion that "The results provided good support for the theoretical rationale underlying the test battery" (p. 87) can not be endorsed. Rather, there appears to be very meager support for the underlying structure of the Stanford-Binet IV as evidenced by the confirmatory factor analysis. Furthermore, as pointed out by Slate (1986) the three-level hierarchical model of intelligence proposed by Thorndike et al. is not supported very well. However, before abandoning the Stanford-Binet IV as a test with poor construct validity, further factor analyses based on more conventional methods and more appropriate age groupings need to be done. Only with further independent analyses will any consensus about the appropriateness of each subtest to its respective Reasoning Area be answered. Because the confirmatory factor analyses done by Thorndike et al. did not substantiate the theoretical structure of the Stanford-Binet IV, exploratory factoring would seem advisable, so that the data can "speak" for themselves. Therefore, in this thesis exploratory rather than confirmatory factor analyses have been used.

Research on The Stanford-Binet IV

The most substantial comparison of the Stanford-Binet IV with other tests has been done by Thorndike et al. (1986) and is reported in the *Technical Manual* of the Stanford-Binet IV. Although the Fourth Edition has been compared to several other tests including the WPPSI, WAIS-R, and K-ABC, only some of the comparisons with the WISC-R and Stanford-Binet L-M will be considered here as they are directly relevant to the current thesis.

First, the Stanford-Binet IV and the Stanford-Binet L-M were both given to 139 "normal" examinees. Thorndike et al. found that the mean for the Composite score on Fourth Edition was 105.8 (SD = 13.8) compared to the L-M mean Total score of 108.1 (SD = 16.7). Because 86 percent of the examinees in this sample were given the Fourth Edition prior to Form L-M, Thorndike et al. caution that the 2.3 point difference may be the result of a practice effect. The correlation between the Total score on Form L-M and the Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning, and Short-Term Memory Areas of the Fourth Edition were .76, .56, .70, and .67 respectively. Indirectly this substantiates a claim that the new test manages to measure more nonverbal areas than the older Stanford-Binet tests--something that was attempted unsuccessfully in the 1937 revision.

In the *Technical Manual* Thorndike et al. (1986) also report the correlations between the Stanford-Binet IV and the WISC-R. Each of the 205 examinees ("normal" children) took both tests in a counterbalanced order. The mean Composite score on the Fourth Edition was 102.4 (SD = 15.3) while the mean FSIQ on the WISC-R was 105.2 (SD = 16.7). The correlations between the WISC-R VIQ, PIQ, and FSIQ were as follows: with the Fourth Edition Composite .78, .73, and .83 respectively; with the Fourth Edition Verbal Reasoning Area .72, .60, and .73 respectively; with the Fourth Edition Abstract/Visual Reasoning Area .68, .67, and .73 respectively; with the Fourth Edition Quantitative Reasoning Area .64, .63, and .69 respectively; and with the Fourth Edition

Short-Term Memory Area .64, .63, and .70 respectively. From this it can be seen that the WISC-R FSIQ and Stanford-Binet IV Composite score are relatively strongly correlated. It is also apparent that the four Areas on the Stanford-Binet IV have some relationship to the various scales on the WISC-R, but the strength of these relationships differs depending on how similar the content for the specific Area of the Stanford-Binet IV is to the particular Scale on the WISC-R.

In a group of 19 gifted children Thorndike et al. (1986) found that the Stanford-Binet IV Composite correlated .69 with the WISC-R FSIQ. The respective means and standard deviations were 116.3 (16.4) and 117.7 (12.1). As would be expected in a gifted sample, the strongest scores on both tests were in the Verbal areas; conversely, the weakest areas were the WISC-R Performance Scale and the Fourth Edition's Abstract/Visual Reasoning Area. In general, the results from the two tests are very comparable.

In another clinical sample the WISC-R and Stanford-Binet IV were administered to 90 children who had been identified as learning disabled. In nearly all the cases the WISC-R was administered before the Fourth Edition. As expected, the WISC-R PIQ was slightly higher than the WISC-R VIQ. However, on the Fourth Edition, only the Short-Term Memory Area appeared to be weak compared to the other three Reasoning Areas; the means for Verbal Reasoning, Abstract/Visual Reasoning, and Quantitative Reasoning Areas were all within about two points. The mean WISC-R FSIQ was 87.8 (SD = 15.7) while for the Fourth Edition Composite the mean was 84.8 (SD = 14.5); further, these two scales correlated at .87. Comparison of these means suggests that the Composite score on the Stanford-Binet IV may be 3 or so points lower than the WISC-R FSIQ. Thus, for learning disabled children the Fourth Edition probably has slightly lower scores than the WISC-R, especially when a possible practice effect is considered (i.e., the WISC-R was nearly always administered first). However, the results are encouraging in that there seems to be a great deal of consistency between the various scores on the two tests.

The next study of interest reported by Thorndike et al. (1986) compared the WISC-R and Stanford-Binet IV in a sample of 61 children identified as mentally retarded. Once again all children were tested first with the WISC-R and then with the Stanford-Binet IV. The mean Composite score on the Fourth Edition was 66.2 (SD = 9.8) while the mean WISC-R FSIQ was 67.0 (SD = 9.9); the correlation between these two scales was .66 which is not surprising given the restricted range of the sample. As in the learning disabled sample, the WISC-R PIQ was stronger that the WISC-R VIQ. Meanwhile, on the Stanford-Binet IV the Verbal Reasoning and Short-Term Memory Areas were slightly weaker than the Abstract/Visual and Quantitative Reasoning Areas. Once again the two tests gave similar results across all of the major scales.

Finally, Thorndike et al. (1986) reported the means and standard deviations achieved by "exceptional groups" (gifted, learning disabled, and mentally retarded) on the Stanford-Binet IV. The striking finding is that the four Reasoning Areas are all very similar with the only notable difference being that the Short-Term Memory Area was usually a few points lower than the other Areas; otherwise, only minor fluctuations of a point or two were noted between the Areas, despite the fact that special groups were being tested. This finding is heartening, suggesting that there is much consistency across the various areas[24].

The only published research article that has appeared on the Stanford-Binet IV in current journals has been by Carvajal and Weyand (1986). Their study was extremely limited in scope and is highly suspect because of their limited sample size. Carvajal and Weyand administered the "General Purpose Abbreviated Battery" of the Stanford-Binet IV (Vocabulary, Bead Memory, Quantitative, Memory for Sentences, Pattern Analysis, and Comprehension) and the complete WISC-R to 23 children ranging in age from 8-5 to 9-10

[24] Alternately, it may be argued that the Stanford-Binet IV is nothing more than a *g* - test; therefore, one would expect to find relatively little variation in the major Areas, as they are really measuring only *g* and not group factors. This type of argument does have some basis given the confirmatory factor analyses reported by Thorndike et al. (1986).

years. Their sample was a third-grade class described as being "middle class". In all cases, the WISC-R was given first and followed by administration of the Stanford-Binet IV. The mean Stanford-Binet IV Composite score was 113.3 (SD = 11.3) with a range from 87 to 133. Meanwhile for the WISC-R the mean FSIQ was 115.0 (SD = 12.6) with a range of 92 to 142.

Carvajal and Weyand (1986) found that the Stanford-Binet IV Composite score and the WISC-R Full Scale score correlated .78. They also reported some correlations between WISC-R and Stanford-Binet IV subtests: the Stanford-Binet IV Vocabulary test correlated .65 with the WISC-R Vocabulary subtest; the Stanford-Binet IV Comprehension test correlated .42 with the WISC-R Comprehension subtest; the Stanford-Binet IV Pattern Analysis test correlated .18 with the WISC-R Block Design subtest; and the Stanford-Binet IV Quantitative test correlated .14 with the WISC-R Arithmetic subtest. As noted by Carvajal and Weyand the standard deviations in their small sample were lower than those found in the standardization samples for either the WISC-R or the Stanford-Binet IV. As such, they qualify their findings by stating "The homogeneity of the study group may have reduced the correlations" (p. 966).

Further research on the Stanford-Binet IV has been done by Meloff (1987) in an unpublished thesis. In her thesis Meloff describes the Stanford-Binet IV and looks at the discriminative ability of the test in special populations (learning disabled, mentally retarded, gifted, and normal groups). The findings from Meloff's thesis that are most applicable to the current thesis are the correlations between subtests on the Stanford-Binet IV. Unfortunately, these correlations are not broken down into the age groups used in this thesis. Therefore, her reported correlations do not allow for possible developmental trends or differences in correlations at given age levels. However, Meloff suggests that her correlations are generally quite similar to those found by Thorndike et al. (1986).

Meloff (1987) does report some unusually low correlations (e.g., r = .18 between Matrices and Copying; r = .27 between Copying and Memory for Objects). This brings out an important point for any researcher that plans to do work with the Stanford-Binet IV. The low correlations that Meloff reported may be a result of inherent methodological difficulties. It should be noted that the 119 subjects in Meloff's sample ranged in age from 3-6 to 17-5 years, and hence there may have been some compounding of error resulting from correlating subtests across such a wide age range. That is, the Copying subtest is most applicable for children between the ages of 2-0 through 9-11 years while Matrices is most applicable for children age 8 through adult. Hence, correlating these two subtests is methodologically unwise in that the overlap in age ranges (in terms of the average child) is only two years. This restriction of range undoubtably accounts for the low correlations found by Meloff between some of the subtests on the Stanford-Binet IV. Consequently, any further correlational studies should ensure that subtests are appropriate for all ages that are being correlated.

Further Critiques and Reviews on the Stanford-Binet IV

The remaining literature on the Stanford-Binet IV has been restricted to brief review articles, most of which have been of limited scope. One article published in *Communique* was a commentary by Fagan (1986) about the premature release of test kits. His major criticism focussed on the lack of a technical manual and errors in norm tables of the first manual. Since that time, however, the *Technical Manual* (Thorndike et al., 1986) has been published and a second manual with correct norms has been released.

A more substantial criticism of the Stanford-Binet IV has been made by Slate (1986). He listed several complaints that indeed seem to be justified. His first criticism has to do with the measurement of *g* as opposed to group factors. Apparently he did not have the *Technical Manual* available when he was writing his article, despite the fact that he mentions that it was recently published. Nevertheless, Slate's (1986) criticism of the inclusion of group factors on a test measuring a strong general factor is still relevant:

> In addition, it was mentioned that the revised Stanford-Binet essentially measures 'g,' but to what percentage of the variance does this 'essentially' refer? If 'g' accounts for about the same variance that 'g' accounts for on the WISC-R (about 82%; Kaufman, 1975), then the factor areas that the revised Stanford-Binet is supposed to measure (i.e. visual reasoning, abstract/visual reasoning, quantitative reasoning, short-term memory) are not, in fact, measured. This would mean that the Stanford-Binet is a measure of 'g' and 'g' alone (p. 3).

Slate's concern about the role of *g* can be addressed more fully now that the *Technical Manual* has provided the needed information about the confirmatory factor analyses done on the Stanford-Binet IV. Indeed, the relatively small amount of variance accounted for by the group factors does bring up the question of whether or not they deserve mention as separate constructs; in this sense the Fourth Edition does appear to be primarily a measure of *g* rather than of *g* and group factors.

What Slate has failed to clarify is the fact that the factoring of the WISC-R that he referred to (i.e., Kaufman, 1975) used a totally different method than that used on the Stanford-Binet IV. That is, Kaufman's analysis of the standardization data for the WISC-R used principal factoring with varimax rotations; when varimax rotations are used the variance from the general factor is dispersed onto the group factors. Thus, the general factor that Kaufman reported was the first unrotated principal factor. This general factor was not extracted by Kaufman before the factors were rotated. Conversely, Thorndike et al. (1986) did extract the general factor before extracting further factors. As such, the factoring used by Kaufman (1975) is not directly comparable to that done by Thorndike et al. (1986). Anyone comparing the factorial results from these two tests needs to be aware of how differences can be cause by using different factoring procedures.

Similar ignorance concerning the factoring of the Stanford-Binet IV has been propagated by Sim (1987). Sim attempts to rebut Slate's criticism that the general factor on Stanford-Binet IV accounts for 82% of variance (as per Kaufman, 1975). Sim makes the following statement, apparently lacking an understanding or appreciation of factor analysis:

> Critics claim that the Fourth Edition does not adequately measure the four specific abilities it purports to measure. They have pointed out that if the amount of variance covered by the "g" factor is similar to the amount on the WISC-R, 82%, this would leave the four specific abilities to account for a paltry 18% of the variance. However, as reported in the Stanford-Binet IV manual, variance attributed to the "g" factor ranges from 42-50%. Thus, the four specific abilities actually account for approximately 50-58% of the variance (p. 19).

This statement is not totally correct, and it is surprising that such ignorance could be published in a scholarly journal. First, it must be recognized that Kaufman's 82% is referring to percentage of common factor variance, not percentage of total subtest variance; in such a case, the common factor variance does add to 100%. However, Thorndike et al. (1986) are referring to percentage of **total** subtest variance; this does not add up to 100% for the general factor and group factors--the unique variance, including error variance, is included in this 100%. Whereas Sim suggests that the group factors on the Fourth Edition account for 50-58% of the variance he apparently fails to understand that this figure contains not only the variance from the group factors (a small amount) but also the specificity and error variance. Mr. Sim needs to brush up on his factor analysis.

Returning to Slate's (1986) review of the Stanford-Binet IV, it is worth mentioning that if Slate had been cognizant of the factoring method used by Thorndike et al. (1986), he should have referred to the factoring of the WISC-R standardization sample done by Wallbrown et al. (1975) rather than that done by Kaufman's (1975). Wallbrown et al.'s hierarchical factoring method is at least more comparable with the method used by

Thorndike et al. (1986). As a reminder, Wallbrown et al. (1975) found that the general factor on the WISC-R accounted for 36% of the total subtest variance while the Verbal factor accounted for 6% and the Perceptual-Spatial factor accounted for 5% of the variance. This finding is comparable to that for the Stanford-Binet IV where *g* accounts for 42-50% of the total subtest variance while group factors account for anywhere from 2.2 to 9% of the total subtest variance depending on the specific age group that is being referred to. Thus, comparing the factoring of the Stanford-Binet IV with the WISC-R indicates that the amount of variance accounted for by the general and group factors is not all that different if similar factoring procedures are used on both tests.

What does all this mean? First, using the British method of extracting a general factor and then factoring the residual matrix to find group factors, both the Stanford-Binet IV and the WISC-R are basically measures of general intelligence--though the Stanford-Binet IV may be slightly more a *g* -test than the WISC-R, especially for children over 6 years of age. As Osberg (1986) points out, "Additional factor analyses will no doubt be performed by others using alternative techniques and assumptions. Results from these analyses should provide further documentation and provide users with additional information for interpreting test results" (p. 3). Totally different results will be found on the Stanford-Binet IV if a factoring procedure like Kaufman's (1975) is used because the variance from the group factors will be spread onto the group factors.

A further complaint by Slate is the cost of the Fourth Edition (about $327 American in 1986) as compared to the cost of the WAIS-R (about $135 American in 1986). This complaint fails to take into account the fact that the Stanford-Binet IV covers the same age span covered by three Wechsler tests (WPPSI, WISC-R, and WAIS-R). Hence this criticism is one that was not well though through. Osberg (1986), for example, pointed out that the WISC-R and WPPSI together cost approximately [illegible] American funds in 1986). Thus, the cost issue is a straw-man that is easily dismissed.

Yet another criticism by Slate (1986) focuses on the possibility of several different abbreviated batteries being used on the Stanford-Binet IV: "the flexibility that permits one examiner to select a different set of subtests than another to obtain an Intelligence Quotient " (p. 3). This is another poorly chosen criticism. In the *Scoring Guide* for the Stanford-Binet IV[25], Thorndike et al. fail to clarify the issue of which subtest should be included in a battery. However, abbreviated batteries should be considered just that, just as abbreviated batteries from the WISC-R are considered short-forms and not full battery tests. If Thorndike and his colleagues have not made it clear enough, then common sense should clarify that in administering the Stanford-Binet IV, all subtests that are appropriate for a given age level should be given. Any other combinations of subtests should be clearly marked as an "Abbreviated Battery".

Slate's (1986) criticism about the subtests used in an given administration of the Stanford-Binet IV would have been more effective if he had gotten to the heart of the issue. That is, because different subtests are appropriate for different ages and ability levels, the Stanford-Binet IV may indeed measure slightly different things at different age levels (see Vernon, 1987 for a discussion of this); thus, results across different ages may be difficult to compare. The latter is a justifiable hypothesis, but it needs to be verified empirically, especially if the Stanford-Binet IV turns out to be primarily a test of general intelligence.

Other commentaries on the Stanford-Binet IV have been given by Barnard (1987), and Sim (1987). They add no important information to the understanding of the Stanford-Binet IV but rather restated Slate's concerns or attempt to feebly displace them. A more informative commentary on the Fourth Edition has been given by Telzrow (1987). She considers the use of the new Stanford-Binet with preschool children. She notes that with 2-year-olds it is possible that no basal might be obtained; as such, the scores should be

25 All references to the *Scoring Guide* are in reference to: Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Guide for Administering and Scoring the Fourth Edition.* Chicago: Riverside Publishing Company. To avoid confusion with the *Technical Manual*, which is also published in 1986, the above source is not listed in the reference list.

considered "estimates" as outlined by Thorndike et al. (1986). She also criticizes the organization of the Fourth Edition items into subtests rather than retaining the age level organization of the earlier Binet scales; although this criticism may have some bearing for two-year-olds, it does not apply to older children. Further, Telzrow notes that at the youngest age levels the Fourth Edition measures *g* and possibly Verbal Fluency and Visuo-Spatial Performance rather than all four Areas that exist on the Scale. Taken together, her comment do point out that the Stanford-Binet IV may have some limitation with two- and three-year-old children, especially those with below average intellectual capabilities. Time will tell if her concerns are valid.

The most thorough and also most scathing review of the Stanford-Binet IV has been done by Philip Vernon. Given that the authors of the Fourth Edition claim that their test is partially structured after Vernon's model of intelligence, it is fitting that he has reviewed the new Binet Scale. The first two lines from the abstract of Vernon's (1987) article set the tone for the remainder of the article: "The recently published fourth edition of the Stanford-Binet Scale should not be so entitled, since it is more similar in structure and content to the Wechsler scales. It lacks the large variety of items and the flexibility of administration of the Terman-Binet tests" (p. 251).

A reading of Vernon's (1987) article leaves an impression that one of the major problems for older clinicians may be resistance to change: "Testers may find it difficult to learn the administration details and to train new students, and thus may prefer to rely mainly on WISC-R and WAIS-R" (p. 251); he further states that "The complexity of instructions and administration is clearly greater than that of L-M or Wechsler scales; and testers will need to spend a lot of time in learning how to give it" (p. 252). Change is often met with resistance, for the new challenges what was previously held: "Having used and/or taught the the previous editions of the SB for 53 years, I was naturally perturbed to find that Thorndike et al. had given up the traditional Binet format and adopted a structure much more like that of the Wechsler scales, where the items are grouped into subtests, each measuring a different aspect of ability" (p. 253). Clearly Dr. Vernon is sentimental towards the older Stanford-Binet tests and maintains that despite the "haphazard arrangement of item" it gives greater flexibility. Vernon's words will have to be reviewed in ten or so years, once younger clinician's trained on the Stanford-Binet IV have been in the field for a few years.

Another of Vernon's criticisms of the Fourth Edition is that it may measure different things at different ages: "The continuity of the NS scores is questionable also insofar as different sets of subtests are given at different age levels..." (1987, p. 254)[26]. There is some truth in his statement that the four reasoning Areas are comprised of different subtests at different age levels, and that this may result in a "seriously lack of comparability in the content of area scores at different ages" (p. 252). Even though this may be a problem, it was also present on Form L-M where there was considerable change in items at different age levels. In a summary statement, Vernon acknowledges the pragmatic need for changing content: "Thus the content, both of the four area scores and of the composite, does alter, much as it did in SB and L-M. I conclude, then, that this is not a matter that could be solved by psychometric ingenuity, but is a necessary characteristic of human mental growth" (p. 254). Because of the problem of changing content and weakness of certain item types, Vernon (1987) suggests that testers omit the Absurdities, Paper Folding and Cutting, Equation Building, and Memory for Objects subtests. This suggestion needs further research to substantiate it.

In his review, Vernon presents a Table (see "Table 1" on p. 253 of Vernon, 1987) which summarizes the age range for given subtests. Unfortunately information for this table was taken from the *Scoring Guide* for the Stanford-Binet IV and there are some problems in that the age ranges that are specified; these age ranges for some subtests are not accurate in terms of "average" children. For example, Matrices, Memory for Digits,

[26] Vernon uses NS to refer to the Stanford-Binet IV, and SB to refer to the Stanford-Binet.

and Memory for Objects are not appropriate for children age 5 and up as Vernon suggests; rather they are appropriate for children of age 8 and up[27]. This leads one to wonder if Dr. Vernon has actually had much experience in administering the Fourth Edition or if most of his comments were derived from examination of the items (i.e., his statement of "In my own judgement, based mainly on inspection of the subtest items, the weakest ones, which should be dispensed with if possible..." (p. 252)). Readers should carefully note this important error in Dr. Vernon's article and in the *Scoring Guide* for the Stanford-Binet IV.

Regarding the factor analyses conducted on the Stanford-Binet IV, Vernon (1987) suggests "That is, the mean group-factor loadings for any subtest is less than one quarter of the *g* variance, on average. It is doubtful whether this 11% figure is statistically significant. And it is probably too small to justify the authors' belief that the NS measures four diagnostically useful group factors besides *g* " (p. 254). Vernon also comments that he is suspicious of the high *g* loadings found for the Short-Term Memory subtests on the Fourth Edition. Overall, Vernon does not give a favorable review of the Fourth Editions factor structure: "The WISC and WAIS area scores seem to be preferable to the NS for another reason. Many factorial studies have in fact shown that the Wechsler tests measure three distinct group factors, not two" (p. 255). This last comment needs to be taken cautiously, as further factor analyses are needed on the Stanford-Binet IV before one can conclude which factors are being measured, especially given the type of confirmatory analysis done by Thorndike et al. (1986).

In terms of face validity of the subtests and Areas, Vernon endorsed the Verbal Reasoning Area though he found fault with the Absurdities subtest in that it is entirely pictorial. He gave the Abstract/Visual Reasoning Area a pat on the back, suggesting that it provides a heterogeneous sampling of nonverbal skills "which may well be superior to the WISC-WAIS performance tests, though less extensive than Kaufman's (1983)

---

[27] See the Methodology section of this thesis for a detailed description of appropriate ages; this contention can also be verified from Thorndike et al. (1986) by inspecting the intercorrelations for each age level and the number of children taking each subtest.

Simultaneous Processing battery" (Vernon, 1987, p. 255). He criticized the Quantitative Reasoning Area suggesting that it is dependent on "New Math" teaching and that none of the subtests appear to involve a substantial degree of mathematical reasoning. Finally, he noted that memory items on Form L-M accounted for only 12% of the total items, where they account for nearly 25% of all items on the Fourth Edition. Thus, Vernon suggest that the Short-Term Memory Area be down-played.

Vernon's (1987) overall impression of the Stanford-Binet IV can be seen in the following quote: "But it is misleading to call the battery a Stanford-Binet scale. Indeed it marks the end of Terman's 70-year era. It is based on a novel theory of the organization of abilities, and unfortunately this is inadequately confirmed. But the main drawbacks are the greater complexity that testers and tutors will find in learning and administering the scale, and the problems raised by omitting some of the subtests" (p. 257). Despite these criticisms, Vernon suggests that a "considerable" number of Canadian psychologist should use the Fourth Edition and after a few years meet and reach a consensus about its use.

H. A Brief Look at Short-Term Memory

Because the confirmatory factor analyses conducted by Thorndike et al. (1986) suggest that the four subtests on Short-Term Memory Area of the Stanford-Binet IV do not all load on the same factor, it is relevant to briefly consider research in the area of memory. For many clinicians, the inclusion of the Short-Term Memory Area on the Stanford-Binet IV is a welcomed addition; problems in memory are often taken as possible indicators of brain dysfunction, disease, or injury. In fact, in clinical and neurological samples, it is not unusual for clinicians to administer a separate memory test when they are using one of the Wechsler intelligence tests (e.g., the use of the Wechsler Memory Scale as discussed by Cohen &Sandberg, 1977).

Since the inclusion of the digit span type items by Binet and Simon on the 1905 Binet-Simon scale, the inclusion of digit-span subtests had been quite common on intelligence tests. Indeed, historically many intelligence tests have included some types of memory items. Horn (1976) in his review of human abilities in the early 1970s suggested that "At least since the time of the pioneering work of Woodrow in the 1930s it has been known (by some) that there is only a low relationship (r ≈ 0.35) between commonly accepted measures of intelligence and short-term acquisition of the kind measured in span memory, recognition memory, paired-associates learning, and serial learning tasks" (p. 451). Further, Horn suggested that when "memory primaries" are considered in second-order analyses, a broad memory factor can be found that is distinguishable from Gf, Gc, and Gv. In the concluding remarks of his review, Horn (1976) stated that "Memory theory is becoming part of theory of intellect" (p. 478). Thus, though earlier work by Horn and Cattell did not include a Memory factor, Horn (1976) acknowledged that such a factor probably exists. Therefore, the inclusion of the Short-Term Memory Area on the Stanford-Binet IV has support not only from previous Stanford-Binet scales, but also from Horn's acknowledgement of its presence.

Many traditional theories of memory postulate two mechanisms, one short-term, and the other long-term. However, Baddeley (1976) clearly stated that the conception of memory as having only two components is a "gross oversimplification" (p. 169). He suggested that in the 1970s memory was beginning to be considered more of an active system than it had been in the past: "whereas the characteristic 1960s view of memory was that of a store that had the sole function of holding information which might or might not be used in other information-processing tasks, the tendency in the 1970s is to regard memory as an integral part of other information-processing tasks, such as perception, pattern recognition, comprehension, and reasoning" (Baddeley, 1976, p. 187). Thus, in the 1970s and 1980s there has been a tendency to consider models of memory which attribute a more active role to the memory system; that is, short-term memory is usually replaced by "working memory". Whereas short-term memory is traditionally conceived of as a passive storage buffer, working memory refers to a more active component in the processing system: "Working memory is assumed to have processing as well as storage functions; it serves as the site for executing processes and for storing the products of these processes..." (Daneman & Carpenter, 1980, p. 450). It is also assumed that processes and structure compete for a shared limited capacity in working memory.

In a different vein Baddeley (1967) suggested that there is "clear evidence for quite separate visual, auditory, and kinesthetic memory systems. Although we know considerably less about any of these than we know about verbal memory, it is already becoming clear that they do not represent simple unitary systems either" (p. 187). Thus, as suggested by Baddeley, it is possible that there are really several types of short-term memory, depending on the medium by which items are expressed. Couple this with a conception of working memory, and one can explain why the Bead Memory subtest does not load on a Memory factor. Bead Memory involves not only visual memory but also the ability to processing information through conceptual strategies; hence Bead Memory probably exerts a greater strain on working memory than does other memory subtests which require more passive storage. Therefore, Bead Memory may have more in common with other subtests which require visual processing skills and or efficiency of categorical skills than with subtests requiring only passive short-term memory.

In summary, the area of short-term memory is moving away from a passive buffer concept towards a more interactive definition of "working memory". In the latter, both storage and processing compete for a limited amount of space. Further, there is some evidence that different types of short-term memory may exist especially in terms of stimuli that are presented auditorily as opposed to visually. Thus, it is entirely possible that the subtests on the Short-Term Memory Area from the Stanford-Binet IV may not load on the same factor.

CHAPTER III

RESEARCH DESIGN AND PROCEDURES

A. Introduction

This thesis is concerned with the construct and concurrent validity of the Stanford-Binet IV. To achieve this goal, a series of principal components factor analyses were conducted. The examinees were given both the WISC-R and the Stanford-Binet IV in a counterbalanced order; the factoring of the Stanford-Binet IV provides an assessment of the construct validity of this new intelligence test. To ensure that sampling variations are not responsible for the resultant factor structure, the WISC-R was also factored to provide a reference structure; by comparing the current WISC-R factor structure to that found by other researchers, the adequacy of the current sample can be estimated, and possible implications for the factorial results from the Stanford-Binet IV can be determined.

Besides factoring the WISC-R and Stanford-Binet IV, a joint factor analysis of these two tests was done. As pointed out by Kieth (1987) this is an important venture, for it allows a direct comparison of the similarity of constructs measured on the two tests. Furthermore, with the inclusion of more subtests in the factored battery, it is possible that new factors will emerge--ones for which there was an insufficient number of similar tests on either battery alone to allow for a common factor to emerge.

Because of the structure of the Stanford-Binet IV, and because of possible developmental trends, the total sample was divided into three groups. The rationale for these age groupings is given in some detail, as each group was administered slightly different tests on the Stanford-Binet IV. Also, the adaptive testing format of the Stanford-Binet IV creates a problem with missing data, and an explanation of how this was dealt with is given.

B. Instruments

The Stanford-Binet IV and the WISC-R were both used in this study. All 12 WISC-R subtests were administered to each child, though a few children had missing data on the Digit Span or Mazes subtests. From the Stanford-Binet IV, children were administered all subtests that were appropriate to them; because of differing ages and differing cognitive abilities, not all children took the same subtests on the Stanford-Binet IV.

All children in this study were tested by graduate students enrolled in a testing course (Educational Psychology 524). All testing was conducted under the supervision of clinically trained psychologists. Testing procedures followed those outlined in the appropriate manuals for the WISC-R and Stanford-Binet IV. The testing was done between October 1986 and May of 1987. Although the WISC-R and Stanford-Binet IV were only two tests in a larger battery given to these children, only these two test are relevant to this thesis. Despite an attempt to counterbalance the order of administration of the two tests, pragmatic concerns resulted in a slight imbalance: 57% of the children first took the WISC-R followed by the Stanford-Binet IV, while 43% took the Stanford-Binet IV first followed by the WISC-R. The length of time between one administration and the subsequent testing with the other instrument ranged from 1 to 100 days. The mean length of time between the two administrations was 37 days (SD = 18.0 days).

C. Subjects

The subjects included 168 children who had been referred to the Education Clinic at the University of Alberta for intellectual assessment. The Education Clinic is open to all clients whether self-referred, referred by parent, teacher, or other professional. It is a free service offered to the general public with the understanding that graduate students being trained in testing and counseling will perform the services as needed. As such, the current sample is clinical in nature, containing not only gifted children, mentally retarded children,

children with learning problems, but also "normal" children whose parents were simply interested in knowing their child's intellectual potential. The demographics of this sample, including reason for referral, are given below in Table 7. Because some referrals were made through school districts etc., information was not always available, and was recorded as "unknown".

The nature of the current sample from Table 7 can be compared to the information provided in the *Technical Manual* (Thorndike et al., 1986) for the respective percentage in the U.S. population that fits into the appropriate categories. Although the "unknown" categories in Table 7 are fairly large, the overall trends in the current sample are not totally unlike those in the 1980 U.S. Census. Especially important is the fact that the sample is not skewed in terms of an over-representation of children whose parents have more than a high school education, nor in terms of children whose parents work in managerial/professional occupations. Race and ethnic group were fairly diverse and included not only Native Canadians and Caucasian children, but also those from Negro and Asian decent. There was, however, an apparent under-representation of children from rural areas. The reasons for referral were diverse with the largest single category being children with general learning problems in school.

The ages of the children in the sample ranged in age from 6-0 to 16-11 years[28] including 101 males and 67 females; the Stanford-Binet IV Composite IQs ranged from 36 to 141; and the WISC-R Full Scale IQs ranged from 40 to 145.

[28] This age range is for testing on the Stanford-Binet IV.

Table 7

Demographics of the Sample

Characteristic	Percent	Characteristic	Percent
Father's Education		Mother's Education	
Less Than High School Graduate	19.6	Less Than High School Graduate	15.5
High School Graduate	19.6	High School Graduate	28.0
1 to 3 years of College	13.1	1 to 3 years of College	13.7
College Graduate or Greater	17.9	College Graduate or Greater	12.5
Unknown	29.8	Unknown	30.4
Ethnic Background		Number of Siblings	
Caucasian	79.2	None	6.0
Native Canadian	5.4	One	29.8
Negro	1.2	Two	31.0
Asian	4.8	Three	17.9
Other	4.8	Four or More	10.7
Unknown	4.8	Unknown	4.8
Father's Occupation		Mother's Occupation	
Managerial / Professional	25.0	Managerial / Professional	19.0
Technical & Sales	7.7	Technical & Sales	7.1
Service Occupations	31.0	Service Occupations	28.0
Precision Production	6.0	Precision Production	1.8
Operator / Fabricator	7.7	Operator / Fabricator	3.0
Other	13.1	Other	35.1
Unknown	9.5	Unknown	6.0
Parent's Marital Status		Community Size	
Married	72.0	299,000 +	71.4
Separated or Divorced	7.2	25,000 - 99,000	4.8
Widowed	1.2	2,500 - 24,999	16.7
Common-Law	2.4	Less than 2,500	7.1
Unknown	17.3		
Referral Reason		Referral Source	
School Learning Problems	35.7	Parents	55.4
Behavior/Developmental Problems	5.4	Medical	1.2
Reading Problems	10.1	School	41.7
Giftedness	16.1	Other	1.2
Mental Retardation	16.7		
Parental Interest	16.1		

D. Choosing of Age Groups

The decision to use three age groups in the current study was arrived at by (a) determining the appropriate ages where most subjects took the same subtests on the Stanford-Binet IV, and (b) after examination of the pattern of Stanford-Binet IV subtests taken by subjects of various age levels. To facilitate the determination of ages to be included in each age group, the entry Levels for each subtest were inspected. The appropriate age range for each subtest was then estimated using the Vocabulary subtest which serves as the Routing Test by which the child's entry level is determined. A few assumptions were used in arriving at the appropriate ages for each subtest: (a) that the "average" child at any age level would obtain a score of 50 on the Vocabulary subtest; (b) that a subtest is most appropriate at a given age level when a child with the lowest specified entry level (e.g., entry Level I for Matrices) obtains a score of 50, which would be the "average" for children that age; and (c) that the diagonal[29] of the Entry-Level Chart on the back of the Record Booklet for the Stanford-Binet IV could be used to estimate the approximate chronological age for a given entry Level.

Thus, the procedure for determining the most appropriate age range was rather complex. First, the lowest entry Level for a given subtest was noted; for example, Level I for the Matrices subtest. Second, an inspection of the Entry-Level Chart on the back of the Record Booklet was made to find the most likely chronological age for the Level in concern; for Level I, this would be between 8-0 and 8-11 years--which is near the diagonal of the chart. Third, to verify this, the highest pair of items administered on the Vocabulary subtest needed to reach that entry level was read from the Entry-Level Chart; for the age 8-0 to 8-11 this is 23-24 or 25-26. Fourth, the number of items passed (a range) needed to be estimated from the highest pairs that were administered. This was done assuming that the child passed all items except the last three or four, thus giving a ceiling. In our

---

29 By diagonal, it is meant from the top left to the bottom right of the Entry-Level Chart with the bottom right being taken as the point in the last row between P and Q.

example, this would mean that the child had a score between 19 and 22 on the Vocabulary subtest. Fifth and finally, the Scoring Guide for the Stanford-Binet IV was consulted to see if this range contained a score of 50 on the Vocabulary subtest at the given chronological age. Using the above method, the ages for each subtest were calculated. These ages are presented in Table 8.

Table 8
Entry Levels For Each of the 15 Subtests

Subtest	Entry Level	Subtest	Entry Level
Vocabulary	A-Q [1]	Quantitative	A-Q [1]
Comprehension	A-Q [1]	Number Series	I-Q [2]
Absurdities	A-L [4]	Equation Building	M-Q [5]
Verbal Relations	M-Q [5]	Bead Memory	A-Q [1]
Pattern Analysis	A-Q [1]	Memory For Sentences	A-Q [1]
Copying	A-J [3]	Memory For Digits	I-Q [2]
Matrices	I-Q [2]	Memory For Objects	I-Q [2]
Paper Folding	M-Q [5]		

[1] An entry level of A begins at age 2-0; the average adult has an entry level of about P; entry levels beyond P are for older children and adults who are above average in intelligence. Entry levels designated as "A-Q" apply to all people from 2 years through adulthood.
[2] An entry level of I is for the average child of about 8-0 to 8-11 years.
[3] An entry level of J is for the average child of about 9-0 and 9-11 years.
[4] An entry level of L is for the average child of about 11-0 to 12-11 years.
[5] An entry level of M is for the average child of about 13-0 to 14-11 years.

For those who wish to check the accuracy of Table 8, a simple procedure can be followed. For the subtest in question, the reader can consult the tables in the *Technical Manual* (Thorndike, et al., 1986, pp. 110-126) that give the intercorrelations between Tests, Areas, and Composite for each age level (2, 3, 4, 5....18-23 years). In Thorndike et al.'s tables, the numbers below the diagonal give the number of subjects who took the given pair of subtests in the standardization sample. One needs to note the maximum number below the diagonal, as this comes close to representing the total sample size at that age; then for the given subtest, one flips through the age levels until it is found that nearly all of the subject at that age level take the subtest in question. For our example of Matrices, it is seen that roughly 176 of 415 seven-year olds took this test, while 365 of 439 eight-year-olds took the subtest. Clearly, the Matrices subtest, with entry Level I is most appropriate for children over the age of 8-0 years, which coincides with Table 8.

A further reason for using age groups is that by doing so, one is able to observe developmental changes in intelligence if they occur (Kieth, 1987). Also, because the children at different age levels take different subtests on the Stanford-Binet IV, it is possible that slightly different constructs are being measured at each age level (Vernon, 1987). Thus, it is doubly important that any correlational studies, or any factor analytic studies on the Stanford-Binet IV, use appropriate age groupings that are in agreement with Table 8. Only children who take the same subtests should be grouped together--otherwise spurious findings may result [30] .

From examination of Table 8, and the current sample, it was decided that the most appropriate age groupings for the current study would be from 6-0 through 8-11 years, 9-0 through 12-11 years, and 13-0 through 16-11 years. These three groups will at times be referred to as the "younger group", the "middle group", and the "older group" respectively. It should be noted that these groupings do not match with those used by Thorndike et al. (1986) in the confirmatory factoring that was conducted on the standardization sample[31]. As previously pointed out, it is maintained that Thorndike et al.'s groupings are inappropriate and result in both some unnecessary error, and some loss of pertinent information. Further, their use of median correlations for the factor analyses in the age groups tends to obscure the fact that there were large differences in the number of

---

30 An exception to this rule would be if the six test Abbreviated Battery (Vocabulary, Comprehension, Pattern Analysis, Quantitative, Bead Memory, and Memory for Sentences) was used in the study. Because these six test cover the entire age span from 2 through adult, no age groupings would be necessary if one was not concerned with possible developmental trends.

31 Thorndike et al. (1986) used the age groups of: 2 through 6 years; 7 through 11 years; and 12 through 18-23 years.

individuals taking certain subtests; this can be a serious methodological problem (see Gorsuch, 1983).

Appendix A.1 gives the number and percentage of subjects in each age group taking each of the 15 subtests on the Stanford-Binet IV. This appendix was used in conjunction with Table 8 in determining which subtests to include in the factoring of the Stanford-Binet IV at each of the three age levels. Appendix A.1.1, A.1.2, and A.1.3 give the number of subjects that took each pair of tests appropriate at that age level. These appendices are important for two reasons. First, they provide the number of cases on which the pairwise correlations (reported in the results section) are based. Secondly, they indicate which subtests tended to be taken by the same individuals and therefore, which subtests are most appropriate to be included in the same factor analysis. For comparison sake, the percentage of subjects taking each subtest as a function of Thorndike et al.'s (1986) groupings[32] are given in Appendix A.2; corresponding numbers of pairwise cases are provided in Appendix A.2.1 and A.2.2.

Because of the need for three age groupings in this thesis, the size of each sample has been reduced considerably. In the 6-0 through 8-11 year group, the sample size is 46. This includes 27 males and 19 females. The mean age for the group[33] is 7-4 years (SD = 10 months) with a range from 6-0 to 8-11 years. In the 9-0 to 12-11 year group, the sample size is 66. This includes 47 males and 19 females. The mean age for this middle group is 10-11 years (SD = 14 months) with a range from 9-0 to 12-10 years. Finally, in the 13-0 to 16-11 year group, the sample size is 56. This includes 27 males and 29 females. The mean age for the older group is 15-0 years (SD = 15 months) with a range from 13-2 to 16-11 years.

---

32 Actually, the age 6-0 through 6-11 has been added to Thorndike et al.'s age group of 7 through 11 thus giving a grouping of ages 6 through 11, and 12 through 16-11 years; the reason for this was to make Thorndike et al.'s groupings appropriate for the full range of the WISC-R which covers the span of 6-0 through 16-11 years.

33 Unless otherwise specified, the ages referred to in this thesis are the ages at time of testing on the Stanford-Binet IV.

E. Handling Missing Data

The adaptive testing format of the Stanford-Binet IV makes it somewhat difficult to calculate correlations and to perform factor analyses on this scale. Because not all children take the same subtests, the problem of missing data emerges. Looking at Appendix A.1 it can be seen that for each of the three age groups, some tests are not taken by all children. This type of problem is not usually encountered when correlations and factor analyses are run on the WISC-R, because each child takes at least the 10 mandatory subtests and often the two optional subtests as well.

There are several ways to deal with missing data such as found on the Stanford-Binet IV. From the *Technical Manual* (Thorndike et al., 1986) it is obvious that correlations were calculated on a pairwise basis; that is, each correlation coefficient is computed using all cases which have complete data for the pair of subtests being correlated, even if that record has missing values on other subtests. This can be seen from the differing n's for each pair of subtests when intercorrelations were reported at each age level (Thorndike et al., pp. 110-127). This method has some merit, especially for correlations; however, as warned by Gorsuch (1983) this can present problems in factoring the correlation matrix if the n's differ from one another very much. Because calculating pairwise correlations utilizes a maximum amount of data, it will be used for calculating correlations between the the subtests on the Stanford-Binet IV[34], between the subtests on the WISC-R, and between the Stanford-Binet IV and WISC-R subtests taken together. Further, because few cases were missing from the WISC-R data, pairwise correlations were also used as the basis for factoring the WISC-R.

However, for the factoring of the Stanford-Binet IV, the amount of missing data, especially for the two older age groups, necessitated an alternate and more appropriate method of handling the problem. Dropping all cases for which there was not complete

[34] The number of cases for each pairwise comparison between subtests on the Stanford-Binet IV for each of the three age groups is given in Appendix A.1.1 - A.1.3.

data would have reduced the total sample size too much. Besides, a more effective procedure as outlined by Timm (1970) exists for dealing with missing data. This procedure involves replacing missing data with estimates derived from a multiple regression analysis. First, a regression equation is obtained by using existing data for the subtests where no variables are missing[35]. Then the resulting regression equation is used in conjunction with existing scores for the individual, and the missing data point is estimated. The use of this type of multiple regression procedure for dealing with missing data has proven to be one of the best overall procedures for maintaining true correlations and leaving the variance unchanged as well (Timm, 1970). Therefore, for the factoring of the Stanford-Binet IV, and the joint factoring of the Stanford-Binet IV and the WISC-R the multiple regression procedure was used to estimate missing values.

A step-wise multiple regression procedure based on pairwise correlations was used to calculate the regression equations which would be used to estimate missing data. A cut-off criteria of .05 was chosen; that is, in attempting to add a new subtest to the equation, that subtest had to make a significant contribution (at the .05 level) to the squared multiple correlation, otherwise it would not be entered into the equation. Regression equations were computed separately for each variable at each of the three age levels. Table 9 give the regression equations and squared multiple correlation (R^2) for each subtest in the age group 6-0 through 8-11 years; Table 10 gives the regression equation and the corresponding R^2 for each subtest in the age group 9-0 through 12-11 years; and Table 11 gives the regression equation and the corresponding R^2 for each subtest in the age group 13-0 through 16-11.

[35] Or alternately, as used in this study, the pairwise correlations can be used to calculate the multiple correlations from which the regression equation is derived.

Table 9

Regression Equations Used to Estimate Missing Values For the Age Group 6-0 Through 8-11 Years

Regression Equation For Predicting Missing Scores	R^2 Values For The Regression Equation
Vocabulary = 3.906 + .450 Comp + .277 Copying + .214 Mem For Sent [a]	R^2 = .68
Comprehension = 7.224 + .642 Vocabulary + .260 Pattern Analysis	R^2 = .54
Absurdities = 0.863 + .594 Quantitative + .428 Vocabulary	R^2 = .46
Pattern Analysis = 11.434 + .446 Comprehension + .324 Absurdities	R^2 = .39
Copying = 15.374 + .620 Vocabulary	R^2 = .34
Matrices: No Variables Entered by Stepwise Regression at a .05 criterion [b, c]	R^2 = N/A
Quantitative = 29.871 + .419 Absurdities [a]	R^2 = .34
Number Series = 36.853 + .340 Pattern Analysis [b]	R^2 = .72
Bead Memory = 34.600 + .330 Pattern Analysis [a]	R^2 = .10
Memory For Sentences = 13.029 + .728 Vocabulary	R^2 = .36
Memory For Digits = 16.854 + .383 Vocabulary + .324 Quantitative [b]	R^2 = .57
Memory For Objects = 36.788 + .356 Memory For Sentences [b]	R^2 = .24

[a] Even though there was no missing data for Vocabulary, Quantitative, and Bead Memory, regression equations were calculate to assist clinicians who may for some reason not be able to administer these subtest, but still desire a rough estimation of the child's possible performance on these subtests.

[b] For Matrices, Number Series, Memory For Digits, and Memory For Objects, the independent variables used to predict the criterion excluded the other three in this list. For example in attempting to build a regression equation for Number Series, the subtests of Matrices, Memory For Digits and Memory For Objects were not allowed to enter the equation as independent variables. This was done for pragmatic reasons--because all four of these subtests tended to be missing for the same subjects. If they had been included in the regression process and had been entered into the regression equation, then the predicted dependent variable would have been uncalculatable since the variables used to predict the dependent were often not available. For the remaining eight variables the Matrices, Number Series, Memory For Digits, and Memory For Objects were not included in the regression analysis, as they are not the most appropriate for this age level.

[c] For Matrices, even when all twelve variables were allowed to enter into the stepwise regression as independent variables, none of the subtest surpassed the .05 criteria; as such, no regression equation was calculated.

Table 10

Regression Equations Used to Estimate Missing Values For the Age Group 9-0 Through 12-11 Years

Regression Equation For Predicting Missing Scores	R^2 Values For The Regression Equation
Vocabulary = -2.595 + .380 Comp + .363 Mem For Sent + .327 Num Series	$R^2 = .75$
Comprehension = 4.909+ .468 Absurdities + .433 Vocabulary	$R^2 = .69$
Absurdities = -1.524 + .465 Comprehension + .335 Copying + .284 Matrices	$R^2 = .77$
Pattern Analysis = 6.457 + .550 Bead Memory + .374 Quantitative	$R^2 = .61$
Copying = 14.626 + .579 Absurdities	$R^2 = .38$
Matrices = 8.499 + .483 Absurdities + .340 Vocabulary	$R^2 = .61$
Quantitative = 12.553 + .696 Number Series	$R^2 = .45$
Number Series = 8.942 + .284 Absurdities + .276 Vocab + .264 Quantitative	$R^2 = .66$
Bead Memory = 2.467 + .580 Pattern Analysis + .304 Matrices	$R^2 = .55$
Memory For Sentences = - 5.811 + .487 Vocab + .568 Memory For Digits	$R^2 = .69$
Memory For Digits = 15.073 + .515 Memory For Sentences + .223 Copying	$R^2 = .58$
Memory For Objects = 27.390 + .438 Matrices	$R^2 = .30$

Table 11

Regression Equations Used to Estimate Missing Values For The Age Group 13-0 Through 16-11 Years

Regression Equation For Predicting Missing Scores	R^2 Values For The Regression Equation
Vocabulary = 6.297 + .600 Comprehension + .322 Memory For Sentences [a]	R^2 = .77
Comprehension = - .259 + .344 Quant + .336 Vocab + .298 Number Series	R^2 = .87
Absurdities = 14.097 + .416 Comprehension + .295 Copying [b]	R^2 = .75
Pattern Analysis = 18.317 + .335 Quantitative + .307 Bead Memory	R^2 = .76
Copying = 11.606 + .561 Bead Memory [b]	R^2 = .63
Matrices = 9.525 + .569 Quantitative + .269 Bead Memory	R^2 = .79
Quantitative = - 4.331 + .650 Comprehension + .425 Matrices [a]	R^2 = .85
Number Series = 12.573 + .397 Quantitative + .401 Comprehension	R^2 = .79
Bead Memory = - 10.592 + .745 Pattern Analysis + .432 Matrices [a]	R^2 = .71
Memory For Sentences = - 1.499 + .280 Comp+ .343 MFD + .307 Vocab	R^2 = .76
Memory For Digits = 12.806 + .451 MFS + .343 Quantitative	R^2 = .65
Memory For Objects = 16.306 + .[illegible]1 MFS + .334 Memory For Digits	R^2 = .50

[a] Even though there was no missing data for Vocabulary, Quantitative, and Bead Memory, regression equations were calculate to assist clinicians who may for some reason not be able to administer these subtest, but still desire a rough estimation of the child's possible performance on these subtests.

[b] For the Absurdities and Copying Subtest, all 11 other subtest were allowed to enter as independent variables in the regression equation. For the remaining subtests only 9 subtests (excluding Absurdities and Copying) were used as independent variables. The reason for this being that inclusion of these two subtests as independent variables would often make it impossible to estimate scores--because a fairly large number of children this age do not take these two subtests.

Table 12

Regression Equations Used To Estimate Missing Values On the WISC-R[a]

All Ages

Regression Equation For Predicting Missing Scores	R^2 Values For The Regression Equation
Digit Span = 2.273 + .407 Information + .210 Arithmetic + .123 Coding	$R^2 = .60$
Coding = 1.212 + .310 Arithmetic + .233 Vocabulary + .292 Digit Span	$R^2 = .38$
Mazes = 3.079 + .388 Block Design + .232 Picture Arrangement	$R^2 = .32$

[a] These regression equations were used for the WISC-R to estimate missing values only when the WISC-R was jointly factored with the Stanford-Binet IV; otherwise, the factoring done on the WISC-R was based on correlation matrices computed on pairwise correlations. For the entire sample (n=168) there were 3 missing data points on Digit Span, 3 on Coding subtest, and 11 on Mazes.

From Tables 9-11 it should be noted that for the younger group the R^2 values tend to be relatively low. This indicates that although there is some overlap of variance between the subtests, this amount is not nearly as high as in the two older groups. The implication from this is that estimates of missing data in the youngest age group will be less accurate than in the older groups. Further, the test with the lowest R^2 in both the two older groups is Memory For Objects; this suggests that it has the least variance in common with the other subtests, and consequently will have the poorest regression estimates when data is missing.

F. Data Analysis

All computations were done with the Statistical Package for the Social Sciences-Revised (SPSSx). According to Pollane & Schnittjer (1977) SPSS (the forerunner of SPSSx) was found to be one of the best packages for running principal components analyses in terms of relative cost, options available, pedagogical aids, accuracy, and output. Similarly, MacCallum (1983) found that SPSS had no computational problems when doing principal components factoring.

Analysis of the data was conducted in several stages.

1. Means and standard deviations were calculated for all scales and subtests on both the WISC-R and Stanford-Binet IV for each of the three age groups.
2. Pearson product-moment correlations based on pairwise comparison were computed separately for each of the three age groups:
   a. Intercorrelations between subtests and scales on the WISC-R
   b. Intercorrelations between subtests, Areas, and the Composite on the Stanford-Binet IV.
   c. Intercorrelations between WISC-R subtests and scales with the Stanford-Binet IV subtests, Areas, and Composite.
3. Principal components factor analysis of all 12 WISC-R subtests--followed by varimax rotation of factors; this was done separately for each of the three age groups and missing data was handled in a pairwise fashion. Further, the entire sample (n = 168) was then factored in a single age group for the WISC-R to check on the effect that dividing the sample into three age groups might have on the factor structure of each age group taken separately (the results of the latter factoring are presented in Appendices B.1, B.2, and B.3).
4. The subtests from the Stanford-Binet IV were subjected to separate principal components factor analysis at each of the three age levels. Missing data was dealt with by using the multiple regression procedure (for comparison sake, the pairwise method was also used to handle missing data and the results from these analyses are placed in Appendices C.1-C.7). For each age group, components were rotated using the varimax and quartimax criteria. (The direct oblimin criterion was also used to rotate components, but given limitations of space and negative correlations between components, these rotations are reported only in Appendices D.1-D.6). The subtests included in the factor analyses of the different age groups varied:

a. For the 6-0 through 8-11 age group, 8 subtests were included in the analysis: Vocabulary, Comprehension, Absurdities, Pattern Analysis, Copying, Quantitative, Bead Memory, and Memory for Sentences.

b. For the 9-0 through 12-11 age group, 12 subtests were included in the analysis: Vocabulary, Comprehension, Absurdities, Pattern Analysis, Copying, Matrices, Quantitative, Number Series, Bead Memory, Memory for Sentences, Memory for Digits, and Memory for Objects.

c. For the 13-0 through 16-11 age group, 10 subtests were included in the analysis: Vocabulary, Comprehension, Pattern Analysis, Matrices, Quantitative, Number Series, Bead Memory, Memory for Sentences, Memory for Digits, and Memory for Objects.

5. A joint principal components analysis of the WISC-R and Stanford-Binet IV was done at each of the three age levels. Components were rotated by the varimax and quartimax criteria. All 12 WISC-R subtests were included at each of the three age groups; the Stanford-Binet IV subtests included for each age group were as outlined in 4 a-c above.

Principal components analyses were used in this thesis as an approximation of principal factoring; more specifically, a truncated principal components method was used in this thesis. Because principal components analysis is simpler and less expensive than principal factoring, and because the two do not give greatly differing results (Lee & Comrey, 1979; Mulaik, 1972; Velicer et al., 1982) the former method was used. Furthermore, principal component analyses would have been necessary in order to determine the number of factors to extract and rotate, even if principal factoring was used; therefore, by using principal components analyses an additional step was omitted. Also, the problem of communality estimates and possible Heywood cases was avoided by using the principal components method. Because of the high reliabilities of the subtests from both the Stanford-Binet IV and the WISC-R, and because the resulting communalities from the principal components analyses are high, the use of this method receives additional support.

Although Thorndike et al. (1986) used a version of confirmatory factor analysis, the factor structure that emerged did not clearly substantiate the theoretical structure of the Stanford-Binet IV. Given these problems, it seemed that an exploratory factor analysis of this test was called for. That is, the relationships between the subtests were allowed to determine the emergent structure of the test; however, in general terms, it was believed that the factor structure of the Stanford-Binet IV that would emerge would not be totally different from the hypothesized structure as outlined by Thorndike et al. (1986). To allow for comparison with Thorndike et al.'s factorial results, the quartimax criterion for rotating factors was used; the quartimax method is actually close to the procedure used by Thorndike et al. in that a general factor is allowed to emerge and the resulting group factors are found from the residual matrix left over after the general factor has been extracted.

Similarly, to allow for comparison with the known WISC-R factor structure (e.g., Kaufman, 1975) varimax rotation of the principal components was undertaken. Although the general factor is lost in this procedure, it does allow for a clearer identification of group factors, though they do contain variance from the general factor. Nonetheless, the use of varimax rotations with the WISC-R has resulted in clinically useful information, and therefore probably can do the same with the Stanford-Binet IV.

The joint factoring of the WISC-R and Stanford-Binet IV was done in order to help clarify the relationship between the various subtests on each test. The importance of this aspect of this thesis cannot be overstated. First, when factor analyzing variables, the number of factors extracted does bear some relationship to the number of variables included in the analysis; at least two and preferably three variables are needed to adequately define a factor. Thus, by combining the Stanford-Binet IV and WISC-R subtests, the number of variables is increased, and factors can be more adequately defined by the variables.

McDonald and Mulaik (1979) in discussing the common factor method briefly discuss the case of the addition of more variables to a first set. Let us assume that a factor pattern emerges for the WISC-R (with n variables); this will be called the "core". The Stanford-Binet IV tests (the m variables) are added as variables to form a combined matrix that is then factored. "Although the effect of adding more variables may be to add further common factors, so that, say, there are r common factors altogether, the requirement of consistent loadings is that one must be able to fit the model and choose a rotation in such a way that the factor loadings of the n core variables are unaltered in the context of the m variables" (McDonald & Mulaik, 1979, p. 302). Thus, it is possible that when the m + n variables are factored, that the unique variance of some variables may be redefined into common variance and a new factor will emerge.

By jointly factoring the Stanford-Binet IV and WISC-R it is possible that the same factors found on one of the tests will be obtained, or that factors from both tests will emerge, or that a set of totally new factors will be found. If the addition of the Stanford-Binet IV subtests to the WISC-R subtests results in the same factors as typically found on the WISC-R (or vice versa), further validation is given to the generalizability of the WISC-R (or Stanford-Binet IV) factor structure as a basic structure of human intelligence. By jointly factoring the two tests the nature of the Stanford-Binet IV will be further clarified. Further, by determining the relationship between the subtests on the Stanford-Binet IV and the WISC-R factor structure, clinicians using the WISC-R will have the option of using certain tests from the Stanford-Binet IV (or vice versa) to supplement the WISC-R subtests.

The number-of-factors problem in this thesis is handled in a pragmatic way. Actually, a combination of different rules is used as suggested by several researchers (Gorsuch, 1983; Hakstian et al., 1982). Although several different rules are used, the primary emphasis was on a combined usage of the eigenvalue-one and scree test criteria (Hakstian et al., 1982). However, as suggested by many different researchers (Carroll, 1985; Comrey, 1978; Gorsuch, 1983; Hakstian et al., 1982) various number-of-factors solutions around those suggested by the eigenvalue-one and scree test criteria were also examined; also, as suggested by Walkey (1983) the number of factors hypothesized to underlie the test was taken into consideration and rotated for inspection even if not warranted by the eigenvalue-one or scree test criteria. The ultimate criteria for deciding the number of factors to accept as the "correct" number was made on the basis of psychological meaningfulness.

Typically when interpreting factors only variables with loadings greater or equal to .30 are used in naming factors. The reason for this being that a loading of .30 explains less than 10% of the variance of that variable (Nunnally, 1967). However, because of the very small sample sizes in the three age groups, a more stringent level of loadings was used. As pointed out by Loo (1983), if the true population correlation is 0 then the formula for the standard error of r is σ_r and the formula for calculating this is as follows:

$$\sigma_r = 1 / \sqrt{N-1}$$

Thus, with a sample size of 46, $\sigma_r = .15$. Therefore, in order to achieve an alpha level of .01, correlations as high as .45 may happen purely by chance. When the sample size is increased to 56, $\sigma_r = .13$ which means that to be depart from a null hypothesis that the sample correlation (ρ) departs from zero (α=.01), a correlation must be .39. With a sample size of 66, $\sigma_r = .12$ which means that correlations as high as .36 do not lead to rejection of the null hypothesis. Therefore, as a rough guide in this thesis, only factor loadings above .40 were used to help interpret factors, though lower loadings were considered if they clearly help to define a factor.

G. Limitations of This Study

The small sample sizes for the three age groups presents a serious limitation on the results from the principal components factor analyses. Generally it is desirable to have a minimum of 100 subjects (Gorsuch, 1983) for any type of factor analysis. However, given pragmatic concerns, the current sample sizes are not that large. As such, the results must be taken cautiously, and as a preliminary finding that will need further validation from factoring of larger samples. The main problem with small samples is that correlations tend to be less "stable"; that is, larger differences can occur purely by chance. However, as Baggaley (1982) and Loo (1983) have pointed out, when intercorrelations between subtests are high the stability of correlations tends to increase and smaller samples can be used. Baggaley (1982) suggests that when average intercorrelations are in the .5 range, the minimum number of subjects needed is approximately [illegible] to 3.0 times the number of variables depending on whether the number of variables is large (above 20) or small (around 10).

As a further consideration, a formula presented by Carroll (1985), though not taking into account the strength of intercorrelations, suggests that roughly three components will be justifiable from the current sample sizes. Taking into consideration the high intercorrelations between subtests, four or five factors may also tentatively be considered.

Another factor to remember is that heterogeneous samples tend to inflate a general factor and conversely minimize group factors. Hence, given the wide range in abilities of the children included in this thesis, it is not unlikely that the correlations and general factor may be higher than in a "normal" sample that is normally distributed.

Overall, the results from this thesis should be taken cautiously as the sample sizes for the three age groups are small and the sample is quite heterogeneous. However, because no previous factor analytic work has been done on the Stanford-Binet IV, it is important that some preliminary understanding of its factor structure in a clinically referred population be investigated.

CHAPTER IV

RESULTS AND DISCUSSION

Because of the amount of data that is presented in this thesis, the arrangement of the results section has been divided into several areas that correspond to the description of the "data analysis" phase as outlined in the research and procedures chapter. The first section deals with the effect that order of administration has on scores for the WISC-R and Stanford-Binet IV. The primary question is whether a previous administration of one of these two tests will result in a slight increase in IQ scores on the test that is administered second; in other words, is there a practice effect when these two tests are administered in a relatively short period of time. Also reported in the first section is a plot of the WISC-R Full Scale IQ against the Stanford-Binet IV Composite score; this is provided to give an index of how similar the two tests are, to indicate the concurrent validity of the Stanford-Binet IV against the WISC-R, and how much variability there is between the IQs on these two different tests, and the an.

The second section provides the means and standard deviations on the WISC-R and Stanford-Binet IV for each of the three age groups. Also provided is a histogram of the WISC-R FSIQ and Stanford-Binet IV Composite IQ for each of the three age groups; these are given to clarify the range of scores on each test for each age group. Examination of the histograms provides an indication of the degree of homogeneity or heterogeneity in the three age groups.

The third section deals with the correlations on the two tests. Three tables are provided for each age group: one gives the correlations between the WISC-R subtests and scales; another gives the correlations between the Stanford-Binet IV subtests, Areas, and Composite; and the final table gives the correlations between the WISC-R subtests and scales and the Stanford-Binet IV subtests, Areas, and Composite. For the sake of future reference, the tables have been organized by test rather than age level; that is, all the

WISC-R correlations are given first, then the Stanford-Binet IV correlations, and finally the joint correlations between the WISC-R and Stanford-Binet IV. After all nine tables have been presented, each table is discussed individually.

The fourth section provides the results of the principal components analyses of the WISC-R for each of the three age groups. A brief discussion of the component structure for each age group is given after the appropriate tables; the emphasis is on how similar the WISC-R component pattern in the specific age group is when compared to the the factor pattern that is usually found.

The fifth section provides the results of the principal components analyses of the Stanford-Binet IV. Of the three different criteria used for rotating the factors (varimax, quartimax, and direct oblimin) only the results from varimax and quartimax methods are presented in this chapter, while the results from the direct oblimin rotations have been given in appendices. Within each age group, the unrotated components are given first, then the scree plot, then the quartimax rotations, and finally the varimax rotations.

The sixth section provides the results of the joint principal components analyses of WISC-R and Stanford-Binet IV. In the beginning of this section the eigenvalues, and unrotated components for all three age groups, are provided. Then, by age group the scree plot, the varimax rotated components, and the quartimax rotated components are provided; the results from each age group are discussed separately.

A. Effect of Order of Administration

As mentioned in Chapter 3, there was an attempt to give the two tests in a counterbalanced order. Because of pragmatic problems the two tests were not given in a completely counterbalanced order, nor was there a standard length of time between administration of the two tests. The length of time between one administration and the subsequent testing with the other instrument ranged from 1 day to 100 days. The mean length of time between the two administrations was 37 days (SD = 18.0). Similarly, of the

168 subjects, 96 received the WISC-R followed by the Stanford-Binet IV (referred to as "Order 1"), while 72 receive the Stanford-Binet IV followed by the WISC-R (refered to as "Order 2"). The means and standard deviations for Order 1 and Order 2 are reported in Table 13. Because division of the sample into three age groups based on order of test administration would have resulted in rather small sample sizes, the effect of ordering is examined independent of age group. Thus, only the Area and Composite scores on the Stanford-Binet IV are given[36].

Examination of Table 13 indicates that a slight practice effect may exist when both the WISC-R and Stanford-Binet IV are given to the same child within a short period of time; however, this effect is small, and not immediately clear. Before interpreting Table 13, it should be noted that for normal children, Thorndike et al. (1986) found that the Stanford-Binet IV Composite score was 2.8 points lower than the WISC-R FSIQ when the two tests were given in a counterbalanced order; similarly, in a sample of learning disabled children, Thorndike et al. found that the Stanford-Binet IV Composite score was 3.0 points lower than the WISC-R FSIQ when the WISC-R was administered before the Stanford-Binet IV.

36 Comparison of subtests would be inappropriate for checking the effect of ordering as children of different ages took different subtests. Also, as pointed out by Vernon (1987) it should be remembered that the Area scores on the Stanford-Binet IV may measure slightly different things at different ages--because different subtests are include in computing the SAS for the Area. Despite this possible complication, the Area scores have been treated as if they measure the same construct at each age level.

Table 13

Effect of Order of Administration on Test Score Results[a]

ORDER 1[b]

Stanford-Binet: Fourth Edition (Given After the WISC-R)			**WISC-R** (Given First)		
Areas and Composite	**Mean**	**SD**	**Scales**	**Mean**	**SD**
Composite IQ	101.0	21.9	Full Scale IQ	101.2	23.6
Verbal Reasoning SAS	103.1	20.2	Verbal Scale IQ	99.3	23.9
Abstract/Visual Reasoning SAS	100.2	20.0	Performance Scale IQ	102.7	21.2
Quantitative Reasoning SAS	101.4	21.3			
Short Term Memory SAS	98.2	20.6			

ORDER 2[c]

Stanford-Binet: Fourth Edition (Given First)			**WISC-R** (Given after the Binet IV)		
Areas and Composite	**Mean**	**SD**	**Scales**	**Mean**	**SD**
Composite IQ	95.0	20.3	Full Scale IQ	98.1	21.8
Verbal Reasoning SAS	98.7	19.6	Verbal Scale IQ	96.0	21.2
Abstract/Visual Reasoning SAS	94.1	19.3	Performance Scale IQ	100.5	20.8
Quantitative Reasoning SAS	95.6	18.4			
Short Term Memory SAS	93.6	20.8			

[a] Age Range 6-0 to 16-11 (all subjects, n = 168)

[b] For "Order 1", 96 children received the WISC-R followed by the Stanford-Binet IV.

[c] For "Order 2", 72 children received the Stanford-Binet IV followed by the WISC-R.

Returning to Table 13, examination of Order 1 suggests that when the WISC-R is given before the Stanford-Binet IV the WISC-R FSIQ and Binet IV Composite score are nearly identical. This relationship was further investigated by performing, a forced regression analysis; the dependent variable being the WISC-R FSIQ and the independent variable being the Stanford-Binet IV Composite score; the regression equation is:

WISC-R FSIQ = 1.02 Stanford-Binet IV Composite IQ - 1.76 and $R^2 = .89$

In contrast, examination of Order 2 suggests that when the WISC-R is given after the Stanford-Binet IV, the mean WISC-R FSIQ is 3.1 points higher than the mean Stanford-Binet IV Composite score. To substantiate this, a forced regression analysis was conducted with the dependent variable being the WISC-R FSIQ and the independent variable being the Stanford-Binet IV Composite score; the regression equation is:

WISC-R FSIQ = 1.00 Stanford-Binet IV Composite IQ + 2.92 and [illegible]7

If there was no practice effect, then one would expect one test to be lower than the other (by the same amount) regardless of which test was administered first. On the other hand, if there was a practice effect but no difference in the means, then whichever test was given second should have been higher. From inspection of Table 13, neither a differences in means, nor a practice effect can explain the current findings. Rather, it appears that the Stanford-Binet IV Composite score is 1.5-2.0 points lower than the WISC-R FSIQ[37] and that a practice effect of 1.5 to 2.0 points exists, regardless of which test is given first.

Comparing the Stanford-Binet IV Verbal Reasoning IQ and the WISC-R VIQ indicates that the former is higher than the latter regardless of which test was administered first. However, the mean for the Verbal Reasoning Area on the Stanford-Binet IV

---

37 The mean effect with the WISC-R FSIQ being higher than the Stanford-Binet IV Composite score can be verified by taking the respective means for the entire sample (98.4 versus 99.8). Thus the WISC-R FSIQ was 1.4 points higher even though only 43% of the sample took the WISC-R after the Stanford-Binet IV.

appears to be 3 points higher than the mean VIQ on the WISC-R[38] ; this would leave a practice effect of about 1 point regardless of which test is taken first.

As a final comparison, the WISC-R PIQ is found to be higher than the Stanford-Binet IV Abstract / Visual Reasoning Area regardless of which test is given first. Overall, the mean WISC-R PIQ appears to be about 4 points higher than the mean Stanford-Binet IV Abstract / Visual Reasoning Area[39] ; also, there appears to be a practice effect of about 2 points regardless of which test is given first.

To summarize the effects of ordering, it appears that the Stanford-Binet IV Composite score is about 1.5 to 2.0 points lower than the WISC-R FSIQ, that the Stanford-Binet IV Verbal Reasoning are is about 3 points higher than the WISC-R VIQ, and that the Abstract / Visual Reasoning Area is 4 points lower than the WISC-R PIQ. These findings are in concordance with those reported by Thorndike et al. (1986) for their sample of non-exceptional children (p. 62) and for learning disabled children (p. 75). Furthermore, the current findings suggest that there may be a small practice effect of 1 to 2 points (regardless of which test is administered first) for scales that are similar on the two tests. It is worth noting that the similarity between scores on the major scales of these two tests, combined with the small practice effect, may make these two tests extremely useful if repeated administrations are necessary in a short period of time. By using the Stanford-Binet IV as an alternate for the WISC-R and vice versa, problems resulting from large practice effects (up to 8 points on test-retest on the Stanford-Binet IV Composite score, Thorndike et al., 1986; up to 7 points on test-retest on the WISC-R FSIQ, Wechsler,

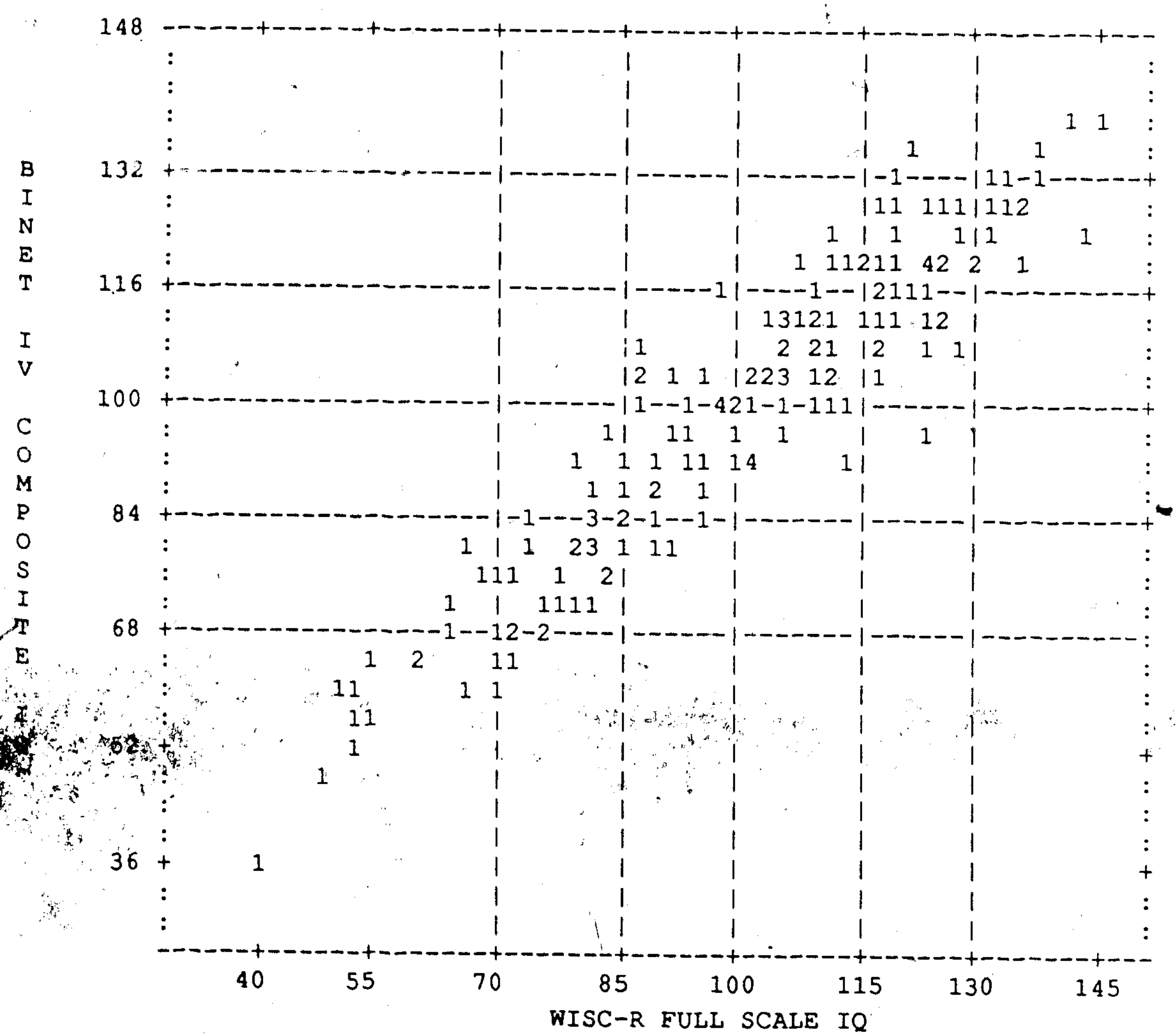
[38] The mean effect with the Stanford-Binet IV Verbal Reasoning Area being higher than the WISC-R VIQ can be verified by taking the respective means for the entire sample (101.2 versus 97.9). Thus the Verbal Reasoning Areas was 3.3 points higher even though only 43% of the sample took the WISC-R after the Stanford-Binet IV.

[39] The mean effect with the WISC-R PIQ being higher than the Stanford-Binet IV Abstract / Visual Reasoning Area can be verified by taking the respective means for the entire sample (101.8 versus 97.6). Thus the WISC-R PIQ was 4.2 points higher even though only 43% of the sample took the WISC-R after the Stanford-Binet IV.

1974) can be reduced, as long as appropriate allowances are made for differences between the scores on the WISC-R and Stanford-Binet IV as indicated above.

In order to provide a clearer picture of how the Stanford-Binet IV Composite score and the WISC-R Full Scale IQ compare on an individual basis rather than just mean differences, Figure 5 was constructed. This figure confirms that the two major scales from the respective tests are indeed quite similar, even though for some individuals the two scores are quite different. Overall, the general relationship between the Stanford-Binet IV Composite and WISC-R FSIQ is very promising.

Figure 5
Stanford-Binet IV Composite IQ Versus WISC-R Full Scale IQ



B. Means and Standard Deviations for Each of the Three Age Groups

Tables 14-16 provide the means and standard deviations for the Stanford-Binet IV and the WISC-R for each of the three age group. Several important trends in the standard deviations of the Stanford-Binet IV Composite and Areas scores, and the WISC-R FSIQ, VIQ, and PIQ should be noted. First, because a clinical sample was used, the standard deviations tend to be higher than those found for the respective standardization samples. This is not surprising given that the range of children's abilities in the current sample is broad, ranging from children being considered for placement in programs for the gifted, to those being considered for placement in programs for the mentally retarded. For the 13-0 to 16-11 age group, the standard deviations are extremely high, indicating extreme heterogeneity; as such, correlations and factor loadings will tend to be inflated in this age group and the results must be interpreted cautiously.

A surprising trend for the standard deviations is that the Area and Composite scores on the Stanford-Binet IV tend to be slightly lower than those on the WISC-R FSIQ, VIQ, and PIQ. This fact is interesting in that the standard deviations for the respective standardization samples are 16 for the Stanford-Binet IV and 15 for the WISC-R. Although this pattern may be inherent to the current sample, Thorndike et al. (1986) found the same trend in non-exceptional children (p. 62), and in learning disabled children (p. 75). Further studies will be needed to determine whether the standard deviations on the Stanford-Binet IV are truly lower than those on the WISC-R or whether the current findings are a result of sampling variability.

The standard deviations on the Quantitative Reasoning Area of the Stanford-Binet IV appear to deviate from the standard deviations on the other Areas. For the 6-0 to 8-11 year-old group, the standard deviation on Quantitative Reasoning Area is lower than the other Areas by 3 or more units; the probable reason is the low standard deviation on the Quantitative subtest which is often the only subtest from the Quantitative Reasoning Area that is given to children at this age. In contrast, the standard deviation on the Quantitative

Reasoning Area for the 13-0 to 16-11 age group is higher than the other Area scores by 3 or more units. Overall, the variability in the standard deviations of the Quantitative Reasoning Area for different age groups raises some questions as to its standardization, as well as to whether the same deviation score means the same thing at different age levels. Based on these preliminary results, it is suggested that in comparing the Quantitative Area SAS scores to other Area SAS scores, the clinician should make allowances for possible differences in the meaning of deviation score, because of differing normal scores. Vernon's (1987) criticism of the Quantitative Reasoning Area as being too dependent on "new math" teaching may explain differing variability at different ages because of different school curriculum emphases at different ages and ability levels.

The standard deviations for the subtests on the Stanford-Binet IV vary somewhat; however, in interpreting these differences caution must be used, especially for subtests that are not taken by all children in that age group. For example, the Matrices, Number Series, Memory for Digits, and Memory for Objects subtests in the youngest age group were taken only be older children and those with higher ability. Therefore, the homogeneity of those taking these subtests is high, and standard deviations would be expected to be lower, because of the homogeneous nature of these children. For this same reason, the standard deviations for Absurdities and Copying subtests in the 13-0 to 16-11 age group are also lower than other subtests.

Table 14

Means and Standard Deviations For the Stanford-Binet IV and WISC-R For the Age Group 6-0 to 8-11 Years[a]

Stanford-Binet IV			WISC-R		
Areas and Composite	**Mean**	**SD**	**Scales**	**Mean**	**SD**
Composite IQ	109.0	14.2	Full Scale IQ	111.3	17.4
Verbal Reasoning SAS	111.1	15.4	Verbal Scale IQ	108.8	17.1
Abstract/Visual Reasoning SAS	104.6	15.3	Performance Scale IQ	111.6	16.5
Quantitative Reasoning SAS	106.9	11.4			
Short-Term Memory SAS	107.4	16.2			
Binet IV Subtests	**Mean**	**SD**	**WISC-R Subtests**	**Mean**	**SD**
Vocabulary	53.6	7.4	Information	11.3	3.1
Comprehension	55.7	7.9	Similarities	11.8	3.8
Absurdities	55.3	8.4	Arithmetic	10.1	2.8
Pattern Analysis	54.2	8.7	Vocabulary	11.8	3.9
Copying	48.6	7.9	Comprehension	12.0	3.0
Matrices	54.9[b]	5.2	Digit Span	10.7	2.4
Quantitative	53.0	6.0	Picture Completion	12.1	2.9
Number Series	55.3[b]	4.8	Picture Arrangement	12.6	3.6
Bead Memory	52.5	9.2	Block Design	12.2	3.0
Memory For Sentences	52.0	9.1	Object Assembly	11.0	2.4
Memory For Digits	54.5[b]	5.4	Coding	10.1	4.4
Memory For Objects	55.3[b]	6.7	Mazes	12.0	3.2

[a] The sample size for this age was n = 46; 27 of these children were male and 19 were female.

[b] Interpret these means and standard deviations with caution as the number of children taking these subtests was much lower than the total number of subjects in this age group (n's of 25, 19, 30, and 28 respectively for Matrices, Number Series, Memory For Digits, and Memory For Objects). It is important to note that only older children, and brighter children in this age group took these four subtests--See Table 8 for entry levels and Appendix A.1 for the percentage of children taking each subtest.

Table 15

Means and Standard Deviations For the Stanford-Binet IV and WISC-R For the Age Group 9-0 to 12-11 Years[a]

Stanford-Binet IV			WISC-R		
Areas and Composite	**Mean**	**SD**	**Scales**	**Mean**	**SD**
Composite IQ	94.2	17.8	Full Scale IQ	96.3	20.1
Verbal Reasoning SAS	98.3	18.5	Verbal Scale IQ	94.6	21.2
Abstract/Visual Reasoning SAS	95.1	17.9	Performance Scale IQ	98.5	18.3
Quantitative Reasoning SAS	94.4	15.7			
Short-Term Memory SAS	91.7	16.8			
Binet IV Subtests	**Mean**	**SD**	**WISC-R Subtests**	**Mean**	**SD**
Vocabulary	48.3	9.1	Information	8.3	3.6
Comprehension	49.2	8.9	Similarities	10.4	4.7
Absurdities	49.9	8.6	Arithmetic	7.8	3.4
Pattern Analysis	49.0	8.4	Vocabulary	9.0	4.1
Copying	43.5[b]	8.1	Comprehension	10.2	3.7
Matrices	49.0[b]	8.6	Digit Span	7.8	3.1
Quantitative	46.5	7.9	Picture Completion	10.9	3.2
Number Series	48.8[b]	7.6	Picture Arrangement	10.9	3.2
Bead Memory	45.8	9.1	Block Design	10.0	4.1
Memory For Sentences	45.0	9.2	Object Assembly	9.5	3.8
Memory For Digits	48.0	7.5	Coding	7.4	3.7
Memory For Objects	48.9	6.9	Mazes	11.4	4.1

[a] The sample size for this age was n = 66; 47 of these children were male and 19 were female.

[b] Less than 90% of the subjects in this age group took these three tests; see Appendix A.1 for details.

Table 16

Means and Standard Deviations For the Stanford-Binet IV and WISC-R For the Age Group 13-0 to 16-11 Years[a]

Stanford-Binet IV

Areas and Composite	Mean	SD
Composite IQ	94.7	26.8
Verbal Reasoning SAS	96.4	22.3
Abstract/Visual Reasoning SAS	94.7	23.9
Quantitative Reasoning SAS	97.7	27.7
Short-Term Memory SAS	92.3	24.7

Binet IV Subtests	Mean	SD
Vocabulary	49.1	11.1
Comprehension	47.9	11.2
Absurdities	45.2	7.7 [b]
Pattern Analysis	48.8	8.7
Copying	38.0	9.0 [b]
Matrices	49.4	11.2
Quantitative	47.8	12.5
Number Series	51.0	10.4 [b]
Bead Memory	47.1	12.7
Memory For Sentences	43.8	10.5
Memory For Digits	48.9	10.6
Memory For Objects	47.6	9.5
Verbal Relations	55.2	6.7 [c]
Paper Folding & Cutting	55.5	7.2 [c]
Equation Building	64.0	8.1 [c]

WISC-R

Scales	Mean	SD
Full Scale IQ	94.8	26.7
Verbal Scale IQ	92.8	25.8
Performance Scale IQ	97.6	24.6

WISC-R Subtests	Mean	SD
Information	8.0	4.3
Similarities	9.9	5.0
Arithmetic	8.6	4.4
Vocabulary	8.4	4.4
Comprehension	9.4	4.3
Digit Span	8.3	3.4
Picture Completion	9.9	3.8
Picture Arrangement	9.9	4.1
Block Design	10.1	4.6
Object Assembly	9.7	4.4
Coding	8.5	4.5
Mazes	10.1	3.8

[a] The sample size for this age was n = 56; 27 of these children were male and 29 were female.

[b] Less than 80% of the subjects in this age group took these three tests; see Appendix A.1 for details.

[c] Less than 55% of the subjects in this age group took these three tests; see Appendix A.1 for details. The means and standard deviations for these three subtests should be interpreted very cautiously.

Examining the means for the different age groups, it is apparent that the Stanford-Binet IV Composite IQ is about 2 points lower than the WISC-R FSIQ for the two youngest age groups (Tables 14 & 15) though for the oldest age group the two scores are nearly identical. Figures 6 through 11 graphically represent the frequencies of the Stanford-Binet IV Composite and WISC-R FSIQ for each of the three age groups. Overall, the shape of the distributions for the two scales on the two tests is similar. However, though the curve is roughly bell-shaped, there is more heterogeneity to the distribution than would be expected on a "truly" normal curve. Hence, factor loadings and the size of the general factor may be slightly inflated.

Next, a comparison of the Stanford-Binet IV Verbal Reasoning SAS with the WISC-R VIQ indicates that for all three age groups the Verbal Reasoning SAS is about 2.3 to 3.7 higher. Conversely, in all three groups the WISC-R PIQ is 2.9 to 7.0 points higher than the Stanford-Binet IV Abstract / Visual Reasoning Areas SAS. These findings correspond to mean differences that were found when ordering effects were also considered.

Figure 6
Histogram of the Binet IV Composite Score for the 6-0 to 8-11 Age Group

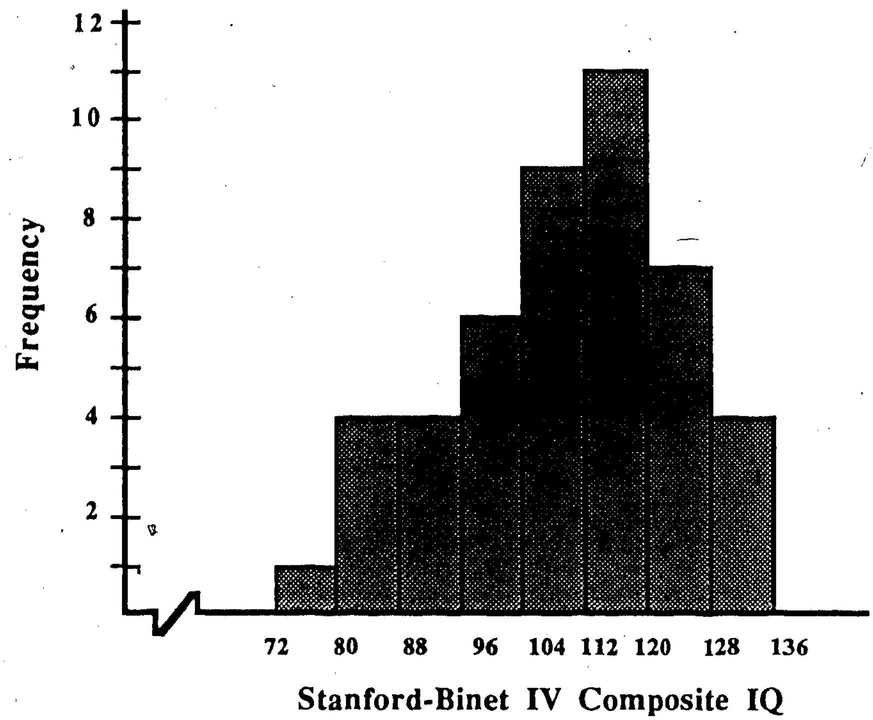


Figure 7
Histogram of the WISC-R Full Scale Score for the 6-0 to 8-11 Age Group

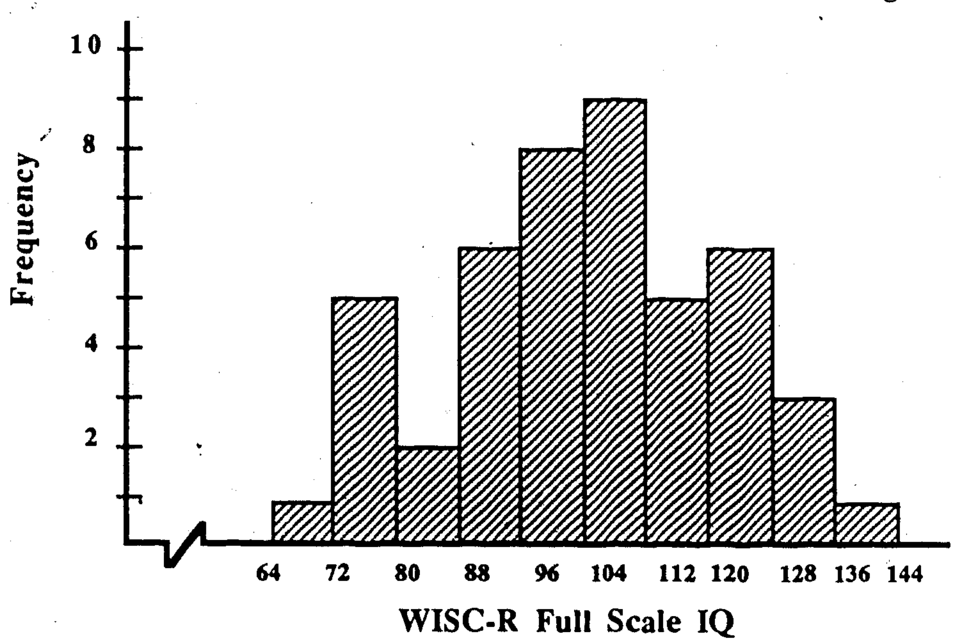


Figure 8
Histogram of the Binet IV Composite Score for the 9-0 to 12-11 Age Group

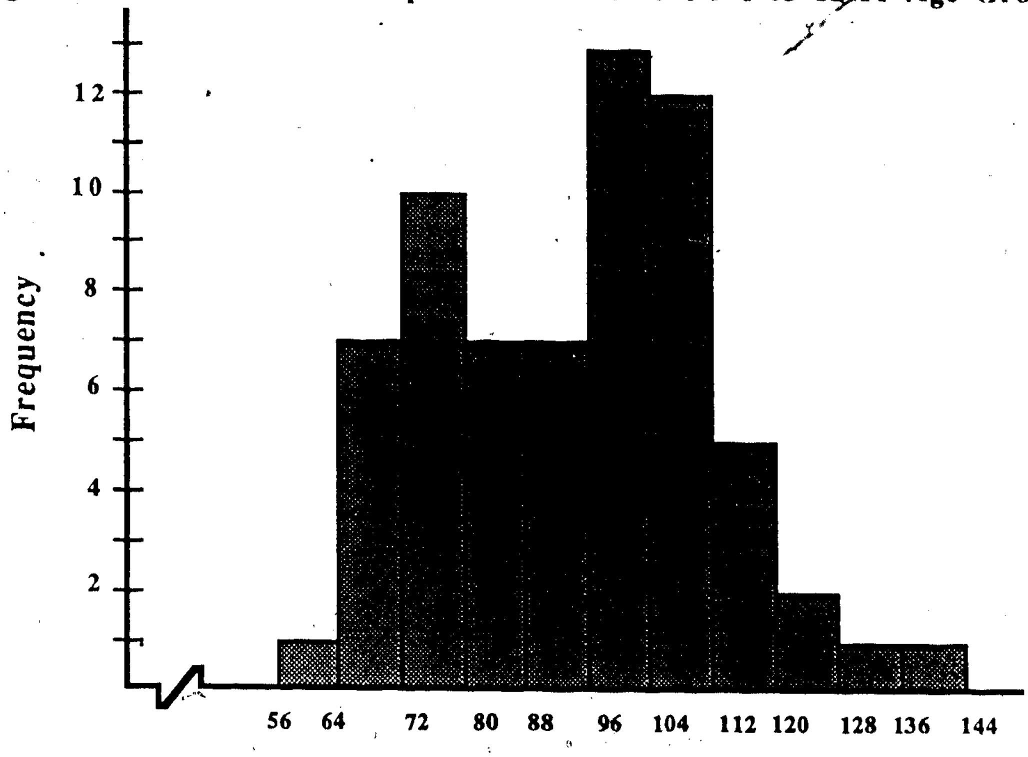


Figure 9
Histogram of the WISC-R Full Scale Score for the 9-0 to 12-11 Age Group

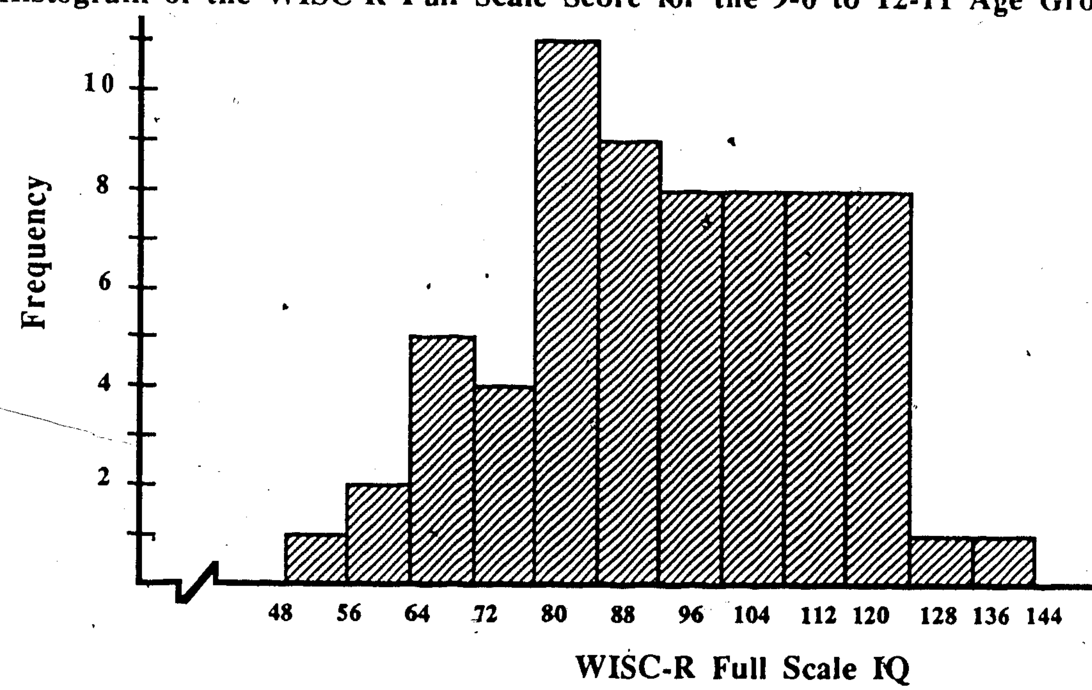


Figure 10

Histogram of the Binet IV Composite Score for the 13-0 to 16-11 Age Group

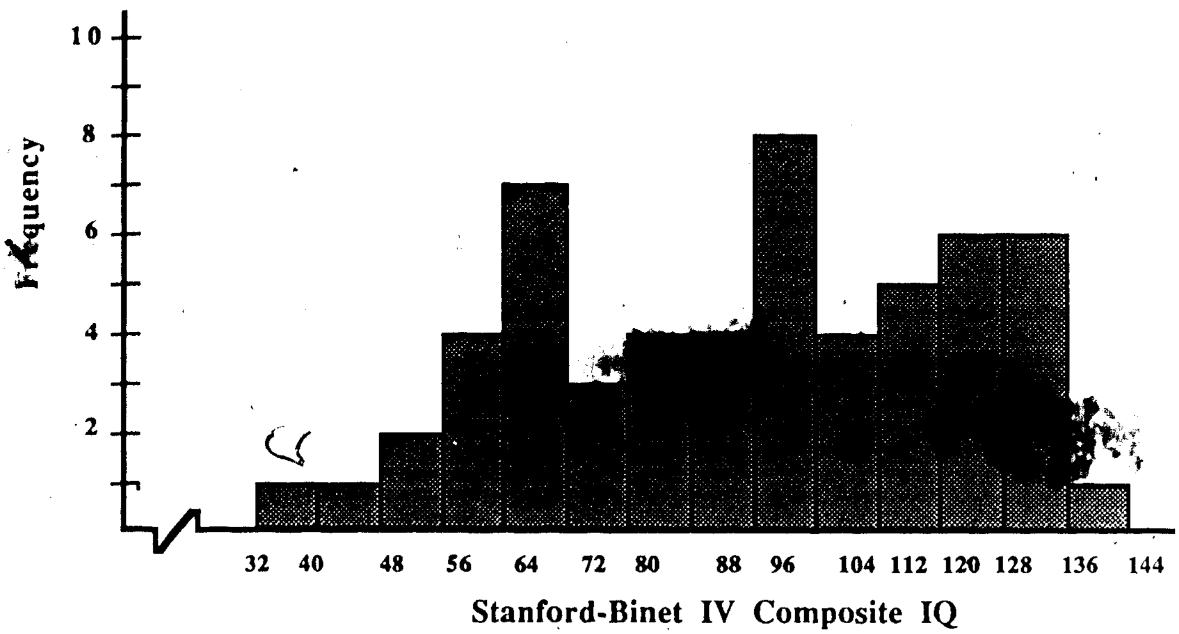
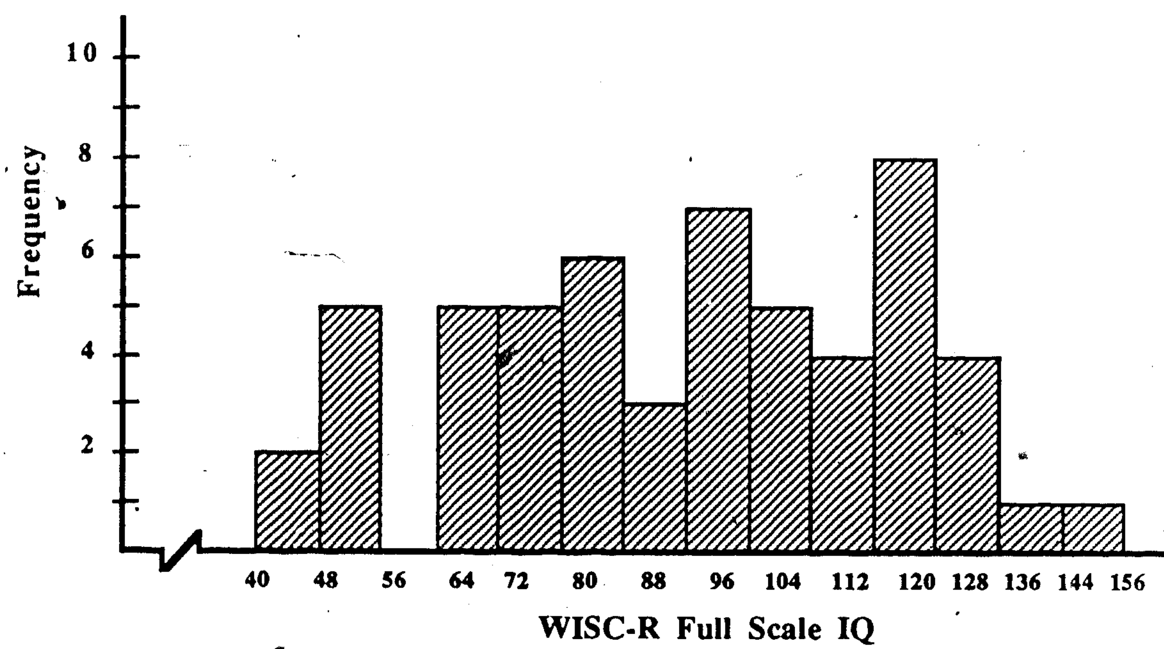


Figure 11

Histogram of the WISC-R Full Scale Score for the 13-0 to 16-11 Age Group



In terms of the means for individual subtests on the Stanford-Binet IV, high and low means must be interpreted with caution; as mentioned previously, some subtests are not taken by all children in a given age group, and hence will have either lower or higher means than other subtests; these subtests can be determined from Table 8 as well as from Appendix A.1. Briefly, the subtests with means that deviate unexpectedly[40] from other means are as follows: in the 6-0 to 8-11 age group, the Copying subtests has a low mean; for the 9-0 to 12-11 age group, the Quantitative, Bead Memory, and Memory for Sentences are slightly lower than other subtests; in the age group 13-0 to 16-11 the Memory for Sentences subtest is lower than other subtests.

For the WISC-R subtests, the means for all subtests tended to be similar, though the Information, Arithmetic, Digit Span, and Coding subtests were generally lower than the means for other subtests in all three groups. This suggests that the third factor of Freedom From Distractibility may be present in the sample (Kaufman, 1975, 1979); the presence of this factor is most pronounced in the 9-0 to 12-11 age group. Because a clinical sample was used in this thesis, it is not surprising that the means on these four subtests are lower than other means.

---

40 "Unexpectedly" is meant as a subtest that was appropriate for the entire range of ages in the group and also that nearly all of the children in the group took the subtest (i.e., over 90% of the children). For appropriateness of each subtest for children of a given age, see Table 8. For example, the mean of 38.0 for the Copying subtest in the 13-0 to 16-0 group is not unexpected; because this subtest is most applicable for "average" children under 10 years of age, in the 13-0 to 16-11 age group, only children with low IQs would take the Copying subtest and hence a low mean is expected.

C. Intercorrelations Between Subtests, Areas, Scales, and Composite on the Stanford-Binet IV and the WISC-R

The tables reported in this section provide the correlations for the Stanford-Binet IV and WISC-R. The first three tables (Tables 17-19) report the correlations between the WISC-R subtests and scales; Tables 20-22 report the correlations between the Stanford-Binet IV subtests, Areas, and Composite; finally, Tables 23-25 report the correlations between the Stanford-Binet IV subtest, Areas, and Composite with the WISC-R subtests and scales. All correlations reported in Tables 17-25 are based on pairwise correlations; the number of pairwise cases for each comparison can be obtained from Appendix A.1.1 - A.1.3. After all nine tables of correlations are reported, a brief discussion of each is given with the emphasis being on unexpected correlations that differ from those reported by Wechsler (1974) or Thorndike et al.(1986).

Table 17
Intercorrelations Between the WISC-R Subtests and Scales For the Age Group 6-0 to 8-11 Years

WISC-R Tests and Scales	Verbal Tests						Performance Tests						Scale Scores (IQs)		
Age 6-0 to 8-11 Years	Information	Similarities	Arithmetic	Vocabulary	Comprehension	Digit Span	Picture Completion	Picture Arrangement	Block Design	Object Assembly	Coding	Mazes	Verbal IQ	Performance IQ	Full Scale IQ
Information		.62	.68	.67	.41	.45	.21	.55	.49	.29	.44	.36	.81	.54	.74
Similarities			.61	.64	.64	.53	.18	.56	.45	.40	.39	.19	.87	.56	.79
Arithmetic				.54	.59	.65	.31	.49	.38	.43	.59	.37	.81	.62	.78
Vocabulary					.58	.43	.27	.64	.41	.40	.53	.28	.84	.60	.80
Comprehension						.51	.31	.57	.26	.36	.42	.14	.76	.52	.70
Digit Span							.29	.53	.51	.64	.53	.09	.62	.66	.68
Picture Completion								.37	.19	.28	.25	.17	.30	.55	.44
Picture Arrangement									.51	.61	.43	.30	.67	.78	.78
Block Design										.59	.30	.47	.51	.69	.63
Object Assembly											.42	.21	.49	.79	.67
Coding												.35	.59	.73	.70
Mazes													.31	.42	.39
Verbal IQ														.72	.94
Performance IQ															.91
Full Scale IQ															

Table 18
Intercorrelations Between the WISC-R Subtests and Scales For the Age Group 9-0 to 12-11 Years

WISC-R Tests and Scales	Verbal Tests						Performance Tests						Scale Scores (IQs)		
Age 9-0 to 12-11 Years	Information	Similarities	Arithmetic	Vocabulary	Comprehension	Digit Span	Picture Completion	Picture Arrangement	Block Design	Object Assembly	Coding	Mazes	Verbal IQ	Performance IQ	Full Scale IQ
Information		.70	.69	.77	.74	.71	.46	.29	.65	.47	.25	.36	.87	.58	.80
Similarities			.66	.72	.78	.49	.59	.39	.62	.46	.30	.31	.89	.64	.84
Arithmetic				.67	.66	.59	.47	.29	.62	.47	.36	.28	.81	.61	.78
Vocabulary					.80	.58	.63	.26	.64	.52	.39	.18	.90	.68	.86
Comprehension						.50	.58	.27	.66	.52	.31	.34	.90	.65	.85
Digit Span							.25	.25	.45	.32	.34	.25	.64	.44	.60
Picture Completion								.40	.56	.60	.31	.25	.63	.76	.74
Picture Arrangement									.27	.47	.29	.28	.36	.64	.52
Block Design										.71	.39	.51	.72	.80	.81
Object Assembly											.31	.49	.57	.84	.74
Coding												.12	.36	.63	.52
Mazes													.34	.44	.41
Verbal IQ														.73	.94
Performance IQ															.91
Full Scale IQ															

Table 19
Intercorrelations Between the WISC-R Subtests and Scales For the Age Group 13-0 to 16-11 Years

WISC-R Tests and Scales	Verbal Tests						Performance Tests						Scale Scores (IQs)		
Age 13-0 to 16-11 Years	Information	Similarities	Arithmetic	Vocabulary	Comprehension	Digit Span	Picture Completion	Picture Arrangement	Block Design	Object Assembly	Coding	Mazes	Verbal IQ	Performance IQ	Full Scale IQ
Information		.87	.78	.89	.83	.78	.61	.64	.78	.72	.61	.54	.94	.81	.91
Similarities			.82	.87	.86	.76	.67	.73	.81	.70	.68	.66	.95	.86	.94
Arithmetic				.74	.80	.71	.63	.73	.75	.66	.50	.61	.89	.78	.87
Vocabulary					.82	.76	.56	.61	.75	.64	.60	.51	.93	.77	.89
Comprehension						.74	.58	.64	.70	.63	.64	.59	.92	.76	.88
Digit Span							.51	.60	.69	[illegible]	.61	.63	.81	.74	.80
Picture Completion								.61	.69	.62	.57	.74	.65	.83	.76
Picture Arrangement									.69	.72	.42	.69	.72	.82	.79
Block Design										.73	.57	.65	.82	.89	.88
Object Assembly											.58	.69	.72	.88	.82
Coding												.59	.65	.78	.75
Mazes													.62	.78	.72
Verbal IQ														.86	.97
Performance IQ															.96
Full Scale IQ															

Table 20
Intercorrelations Between the Stanford-Binet IV Subtests, Areas, and Composite For the Age Group 6-0 to 8-11 Years

Binet IV Tests, Areas, and Composite	Verbal Reasoning Area			Abstract/ Visual Reasoning Area			Quantitative Reasoning Area		Short-Term Memory Area				Areas and Composite				
Age 6-0 to 8-11 Years	Vocabulary	Comprehension	Absurdities	Pattern Analysis	Copying	(Matrices)*	Quantitative	(Number Series)*	Bead Memory	Memory for Sentences	(Memory for Digits)*	(Memory for Objects)*	Verbal Reasoning	Abstract/Visual Reasoning	Quantitative Reasoning	Short-Term Memory	Composite
Vocabulary		.74	.56	.47	.58	.26	.42	.54	.20	.60	.68	.44	.88	.63	.48	.56	.80
Comprehension			.52	.57	.41	.02	.37	.42	.30	.50	.56	.08	.87	.58	.42	.51	.74
Absurdities				.52	.41	.39	.59	.51	.23	.44	.53	.44	.82	.58	.61	.50	.77
Pattern Analysis					.43	.34	.36	.61	.31	.34	.54	.30	.61	.85	.41	.45	.73
Copying						.01	.41	.50	.17	.32	.50	.43	.54	.79	.46	.40	.67
(Matrices)*							.14	.18	.14	.34	.31	.33	.28	.53	.10	.40	.40
Quantitative								.56	-.01	.33	.58	.26	.54	.41	.96	.22	.62
(Number Series)*									.26	.29	.49	.36	.57	.68	.85	.48	.78
Bead Memory										.27	.19	.37	.28	.35	.05	.73	.47
Memory for Sentences											.63	.49	.60	.42	.34	.79	.68
(Memory for Digits)*												.38	.66	.56	.61	.69	.75
(Memory for Objects)*													.38	.46	.40	.76	.59
Verbal Reasoning														.69	.59	.61	.90
Abstract/Visual Reasoning															.46	.54	.84
Quantitative Reasoning																.28	.68
Short-Term Memory																	.78
Composite																	

* Items in brackets and with an asterick indicate variables for which there is a consistent pattern of missing data for this age; therefore, the correlations for these variables, though reported, must be interpreted very cautiously.

Table 21
Intercorrelations Between the Stanford-Binet IV Subtests, Areas, and Composite For the Age Group 9-0 to 12-11 Years

Binet IV Tests, Areas, and Composite Age 9-0 to 12-11 Years	Verbal Reasoning Area			Abstract/ Visual Reasoning Area			Quantitative Reasoning Area		Short-Term Memory Area				Areas and Composite				
	Vocabulary	Comprehension	Absurdities	Pattern Analysis	Copying	Matrices	Quantitative	Number Series	Bead Memory	Memory for Sentences	Memory for Digits	Memory for Objects	Verbal Reasoning	Abstract/Visual Reasoning	Quantitative Reasoning	Short-Term Memory	Composite
Vocabulary		.77	.72	.59	.39	.70	.58	.72	.59	.73	.55	.42	.91	.71	.70	.75	.85
Comprehension			.77	.49	.35	.64	.56	.66	.47	.59	.47	.48	.93	.62	.66	.65	.79
Absurdities				.66	.62	.74	.63	.73	.60	.60	.49	.52	.90	.81	.76	.72	.89
Pattern Analysis					.54	.58	.54	.63	.71	.52	.46	.50	.63	.87	.67	.71	.80
Copying						.47	.48	.54	.52	.37	.48	.46	.48	.81	.55	.55	.67
Matrices							.51	.69	.60	.59	.44	.55	.76	.83	.66	.70	.82
Quantitative								.67	.32	.50	.38	.40	.64	.63	.93	.52	.75
Number Series									.48	.55	.52	.41	.77	.76	.90	.65	.86
Bead Memory										.43	.40	.48	.60	.73	.46	.76	.71
Memory for Sentences											.73	.45	.70	.60	.57	.84	.75
Memory for Digits												.41	.55	.56	.49	.81	.67
Memory for Objects													.52	.59	.45	.73	.63
Verbal Reasoning														.77	.76	.77	.92
Abstract/Visual Reasoning															.77	.79	.93
Quantitative Reasoning																.64	.88
Short-Term Memory																	.89
Composite																	

Table 22
Intercorrelations Between the Stanford-Binet IV Subtests, Areas, and Composite For the Age Group 13-0 to 16-11 Years

Binet IV Tests, Areas, and Composite	Verbal Reasoning Area			Abstract/ Visual Reasoning Area			Quantitative Reasoning Area		Short-Term Memory Area				Areas and Composite				
Age 13-0 to 16-11 Years	Vocabulary	Comprehension	Absurdities	Pattern Analysis	Copying	Matrices	Quantitative	Number Series	Bead Memory	Memory for Sentences	Memory for Digits	Memory for Objects	Verbal Reasoning	Abstract/Visual Reasoning	Quantitative Reasoning	Short-Term Memory	Composite
Vocabulary		.86	.82	.77	.73	.77	.81	.76	.72	.80	.65	.62	.94	.81	.83	.81	.89
Comprehension			.82	.80	.64	.83	.90	.87	.76	.81	.69	.59	.96	.85	.91	.83	.93
Absurdities				.76	.73	.77	.76	.79	.77	.79	.66	.61	.91	.82	.81	.82	.88
Pattern Analysis					.78	.78	.82	.82	.81	.73	.73	.63	.82	.91	.84	.83	.89
Copying						.60	.63	.62	.79	.65	.60	.57	.74	.87	.69	.75	.79
Matrices							.86	.83	.78	.72	.71	.61	.85	.94	.89	.83	.92
Quantitative								.87	.75	.77	.75	.61	.89	.88	.98	.84	.94
Number Series									.77	.70	.65	.51	.89	.88	.95	.83	.95
Bead Memory										.67	.66	.59	.80	.84	.79	.87	.86
Memory for Sentences											.76	.66	.85	.77	.78	.89	.86
Memory for Digits												.66	.72	.76	.76	.89	.82
Memory for Objects													.64	.68	.63	.84	.73
Verbal Reasoning														.88	.92	.88	.96
Abstract/Visual Reasoning															.91	.88	.96
Quantitative Reasoning																.86	.97
Short-Term Memory																	.94
Composite																	

Table 23

Intercorrelations Between the Stanford-Binet IV Subtests, Areas, and Composite with the WISC-R Subtests and Scales For the Age Group 6-0 to 8-11 Years

Age 6-0 to 8-11 Years	WISC-R Verbal Tests						WISC-R Performance Tests						WISC-R Scale Scores (IQs)		
Binet IV Tests, Areas, & Composite	Information	Similarities	Arithmetic	Vocabulary	Comprehension	Digit Span	Picture Completion	Picture Arrangement	Block Design	Object Assembly	Coding	Mazes	Verbal IQ	Performance IQ	Full Scale IQ
Vocabulary	.65	.65	.60	.77	.58	.40	.20	.62	.51	.42	.47	.32	.79	.61	.76
Comprehension	.54	.64	.51	.62	.57	.25	.22	.63	.40	.17	.40	.51	.69	.51	.65
Absurdities	.48	.58	.53	.49	.47	.33	.39	.55	.40	.50	.41	.50	.60	.59	.64
Pattern Analysis	.39	.37	.40	.41	.22	.25	.35	.40	.49	.40	.45	.64	.44	.58	.55
Copying	.44	.42	.50	.53	.38	.37	.23	.56	.32	.41	.45	.32	.56	.55	.60
(Matrices)*	.35	.17	.28	.31	.15	.42	.37	.14	.53	.46	.34	.31	.33	.55	.47
Quantitative	.55	.47	.55	.39	.29	.36	.05	.44	.44	.39	.40	.33	.55	.49	.56
(Number Series)*	.46	.52	.59	.62	.24	.37	.01	.55	.34	.54	.24	.42	.67	.46	.61
Bead Memory	.18	.29	.28	.31	.48	.28	.38	.47	.29	.28	.25	.25	.36	.45	.43
Memory for Sentences	.56	.63	.44	.60	.46	.47	.14	.41	.58	.27	.24	.27	.64	.43	.59
(Memory for Digits)*	.58	.64	.63	.62	.50	.73	.38	.61	.68	.61	.45	.49	.73	.73	.76
(Memory for Objects)*	.43	.51	.39	.61	.29	.45	.16	.51	.40	.66	.42	.07	.55	.56	.60
Verbal Reasoning	.65	.73	.63	.72	.62	.38	.32	.69	.51	.43	.48	.51	.81	.66	.80
Abstract/Visual Reasoning	.51	.48	.52	.55	.37	.35	.38	.54	.54	.49	.53	.59	.60	.69	.69
Quantitative Reasoning	.58	.50	.61	.44	.32	.35	.05	.47	.43	.42	.40	.37	.59	.50	.60
Short-Term Memory	.52	.61	.52	.65	.58	.48	.36	.62	.53	.39	.40	.35	.69	.62	.71
Composite	.69	.73	.70	.74	.60	.49	.36	.73	.62	.53	.56	.56	.84	.77	.87

* Items in brackets and with an asterick indicate variables for which there is a consistent pattern of missing data for this age; therefore, the correlations for these variables, though reported, must be interpreted very cautiously

Table 24

Intercorrelations Between the Stanford-Binet IV Subtests, Areas, and Composite with the WISC-R Subtests and Scales For the Age Group 9-0 to 12-11 Years

Age 9-0 to 12-11 Years	WISC-R Verbal Tests						WISC-R Performance Tests						WISC-R Scale Scores (IQs)		
Binet IV Tests, Areas, & Composite	Information	Similarities	Arithmetic	Vocabulary	Comprehension	Digit Span	Picture Completion	Picture Arrangement	Block Design	Object Assembly	Coding	Mazes	Verbal IQ	Performance IQ	Full Scale IQ
Vocabulary	.80	.72	.63	.82	.82	.60	.51	.34	.66	.57	.36	.25	.86	.67	.84
Comprehension	.68	.73	.68	.71	.77	.52	.51	.34	.53	.48	.44	.34	.81	.63	.79
Absurdities	.71	.67	.57	.66	.77	.54	.58	.33	.64	.62	.31	.41	.77	.68	.79
Pattern Analysis	.58	.57	.58	.51	.58	.42	.55	.20	.78	.59	.32	.38	.63	.67	.69
Copying	.45	.43	.37	.45	.47	.40	.59	.38	.60	.61	.40	.45	.51	.69	.64
Matrices	.74	.71	.65	.73	.79	.44	.62	.26	.71	.60	.32	.40	.82	.69	.82
Quantitative	.62	.62	.51	.57	.63	.43	.58	.30	.51	.37	.07	.30	.67	.50	.64
Number Series	.76	.72	.73	.76	.69	.53	.51	.17	.67	.59	.20	.43	.83	.60	.79
Bead Memory	.51	.46	.44	.52	.60	.27	.47	.03	.74	.58	.35	.30	.57	.60	.63
Memory for Sentences	.75	.59	.62	.68	.63	.69	.37	.21	.46	.28	.34	.20	.72	.46	.65
Memory for Digits	.59	.49	.58	.57	.48	.83	.28	.13	.50	.30	.34	.13	.61	.44	.58
Memory for Objects	.49	.49	.42	.47	.48	.39	.50	.18	.45	.42	.39	.27	.53	.51	.56
Verbal Reasoning	.80	.76	.68	.81	.86	.60	.59	.38	.66	.61	.40	.36	.89	.72	.88
Abstract/Visual Reasoning	.71	.68	.68	.69	.75	.52	.68	.32	.84	.70	.41	.45	.79	.81	.86
Quantitative Reasoning	.75	.72	.67	.69	.72	.53	.59	.29	.65	.53	.17	.42	.81	.61	.78
Short-Term Memory	.76	.65	.66	.72	.72	.69	.51	.19	.70	.51	.46	.29	.79	.65	.78
Composite	.84	.78	.74	.80	.84	.65	.65	.33	.79	.65	.40	.42	.91	.78	.91

Table 25

Intercorrelations Between the Stanford-Binet IV Subtests, Areas, and Composite with the WISC-R Subtests and Scales For the Age Group 13-0 to 16-11 Years

Age 13-0 to 16-11 Years	WISC-R Verbal Tests						WISC-R Performance Tests						WISC-R Scale Scores (IQs)		
Binet IV Tests, Areas, & Composite	Information	Similarities	Arithmetic	Vocabulary	Comprehension	Digit Span	Picture Completion	Picture Arrangement	Block Design	Object Assembly	Coding	Mazes	Verbal IQ	Performance IQ	Full Scale IQ
Vocabulary	.83	.83	.78	.90	.82	.72	.61	.63	.78	.62	.57	.49	.89	.77	.87
Comprehension	.87	.88	.86	.85	.87	.80	.70	.68	.77	.69	.69	.65	.93	.84	.92
Absurdities	.79	.85	.81	.77	.76	.69	.68	.77	.78	.70	.55	.55	.86	.81	.87
Pattern Analysis	.79	.81	.78	.70	.79	.70	.70	.69	.84	.73	.69	.67	.83	.86	.88
Copying	.71	.72	.57	.72	.55	.62	.56	.67	.74	.60	.62	.47	.71	.74	.75
Matrices	.75	.84	.76	.77	.80	.75	.68	.74	.76	.74	.71	.70	.84	.88	.90
Quantitative	.83	.85	.79	.83	.84	.83	.62	.68	.80	.72	.61	.65	.89	.82	.89
Number Series	.84	.87	.82	.76	.80	.76	.56	.68	.76	.60	.62	.56	.90	.81	.91
Bead Memory	.75	.75	.80	.67	.64	.71	.70	.69	.80	.68	.59	.62	.78	.83	.83
Memory for Sentences	.81	.79	.71	.85	.75	.82	.56	.52	.70	.71	.62	.58	.85	.76	.83
Memory for Digits	.70	.72	.67	.68	.72	.82	.54	.64	.68	.69	.55	.62	.76	.74	.77
Memory for Objects	.60	.55	.61	.63	.54	.62	.39	.59	.54	.67	.51	.51	.63	.65	.66
Verbal Reasoning	.89	.92	.86	.91	.87	.80	.68	.72	.83	.71	.66	.61	.96	.87	.95
Abstract/Visual Reasoning	.82	.88	.80	.80	.82	.77	.69	.79	.85	.78	.74	.73	.89	.92	.93
Quantitative Reasoning	.86	.87	.82	.84	.85	.83	.67	.72	.82	.72	.65	.66	.91	.86	.92
Short-Term Memory	.84	.83	.81	.82	.76	.86	.65	.69	.80	.79	.68	.67	.88	.86	.90
Composite	.89	.91	.86	.88	.86	.85	.70	.76	.86	.79	.72	.70	.95	.92	.97

Intercorrelations Between the WISC-R Subtests and Scales[41]

The intercorrelations between the WISC-R subtests and scales for the age group 6-0 to 8-11 years are reported in Table 17; all intercorrelations were significant ($p < .05$) with the following exceptions: Picture Completion was not significantly related to Information ($p=.078$), Similarities ($p=.111$), Block Design ($p=.102$) or to Mazes ($p=.130$); furthermore, Mazes was not significantly related to Similarities ($p=.112$), Comprehension ($p=.182$), Digit Span ($p=.292$) or to Object Assembly ($p=.082$).

The intercorrelations between the WISC-R subtests and scales for the age group 9-0 to 12-11 years are reported in Table 18; all intercorrelations were significant ($p < .05$) with the following exceptions: Mazes was not significantly related to Vocabulary ($p=.078$) or to Coding ($p=.182$).

The intercorrelations between the WISC-R subtests and scales for the age group 13-0 to 16-11 years are reported in Table 19; all intercorrelations were significant ($p < .05$) with no exceptions; in fact all intercorrelations were significant even at the .001 level.

Intercorrelations Between the Stanford-Binet IV Subtests, Areas, and Composite

The intercorrelations between the Stanford-Binet IV subtests, Areas, and Composite for the age group 6-0 to 8-11 are reported in Table 20; the number of pairwise cases used to calculate each correlation can be obtained from Appendix A.1.1. Although the intercorrelations between Matrices, Number Series, Memory for Digits, and Memory for Objects with all other subtests and Areas has been reported, these correlations must be interpreted very cautiously because they are not appropriate for many children in this age

[41] Following Thorndike et al. (1986, see note 2, p. 110), intercorrelations between subtests and scales on the WISC-R or between subtests and Areas, or subtests and Composite on the Stanford-Binet IV have not been adjusted for overlap. The reason for this being that "the adaptive testing format of the Fourth Edition does not require each examinee to take the same test" (Thorndike et al., 1986, p. 110).

group and the number of pairwise cases for these correlations is often low[42]. For the eight subtests appropriate for this age level (see Chapter 3) all intercorrelations between subtests, Areas, and the Composite were significant (p < .05) with the following exceptions: Bead Memory was not significantly related to Vocabulary (p=.088), Absurdities (p=.061), Copying (p=.131), Quantitative (p=.464) or to the Quantitative Reasoning Area (p=.363); further, the Quantitative subtest was not significantly related to the Short-Term Memory Area (p=.067).

Comparing the intercorrelations for the 6-0 to 8-11 age group (Table 20) with those reported by Thorndike et al. (1986, pp. 114-116) a striking similarity can be noted. Only a few correlations differ to some degree. First, the intercorrelations between Bead Memory and nearly all other subtests, Areas, and the Composite are much lower in the current sample; the correlations between the Quantitative subtest and the Short-Term Memory Area is also slightly lower in the current sample; similarly, the correlations between the Quantitative Reasoning Area and both the Short-Term Memory Area and the Composite are lower in this sample; and the Short-Term Memory Area has a lower correlation with the Composite than reported by Thorndike et al.. Conversely, some of the correlations in the current sample are higher than those reported by Thorndike et al.: the Vocabulary subtest correlates slightly higher with both the Copying subtest and the Abstract / Visual Reasoning Area; similarly, the Comprehension subtest correlates more highly with both the Pattern Analysis subtest and the Abstract / Visual Reasoning Area than reported by Thorndike et al.; also, the Absurdities subtest is more highly correlated with both the Quantitative

42 For the four subtest taken by only some children in this age group (Matrices, Number Series, Memory for Digits, and Memory for Objects) many correlations were not significant (p>.05): Matrices was not significantly related to Vocabulary (p=.107), Comprehension (p=.463), Copying (p=.476), Quantitative (p=.252), Number Series (p=.227), Bead Memory (.254), Memory for Sentences (p=.052), Memory for Digits (p=.066), Memory for Objects (p=.055), the Verbal Reasoning Area (p=.090) or to the Quantitative Reasoning Area (p=.310); Memory for Digits was not significantly related to Bead Memory (p=.160); Memory for Objects was not significantly related to Comprehension (p=.344), Pattern Analysis (p=.058), Quantitative (p=.092) or to Number Series (p=.066); and Number Series was not significantly related to Bead Memory (p=.137) or to Memory for Sentences (p=.116).

subtest and the Quantitative Reasoning Area than reported by Thorndike et al. (1986). Overall, however, the correlations from the 6-0 to 8-11 age group are nearly interchangeable with those found by Thorndike et al.

The intercorrelations between the Stanford-Binet IV subtests, Areas, and Composite for the age group 9-0 to 12-11 are reported in Table 21; the number of pairwise cases used to calculate each correlation can be obtained from Appendix A.1.2. For the twelve subtests appropriate for this age level (see Chapter 3) all intercorrelations between subtests, Areas, and the Composite were significant ($p < .05$). Comparing the findings from the current sample (Table 21) with reported intercorrelations by Thorndike et al. (1986, pp. 117-120) reveals that the two sources indeed report similar results. A few of the correlations in the current sample are slightly higher than those reported by Thorndike et al. (1986): First, the Absurdities subtest correlates more strongly with nearly all subtests, Areas, and the Composite; second, the Matrices subtest has slightly higher correlations with Vocabulary, Absurdities, Memory for Sentences, Memory for Objects, and the Verbal Reasoning Area; third, the Pattern Analysis subtest has higher correlations with the Bead Memory, Memory for Objects, and Short-Term Memory Area; fourth, the Number Series subtest has a slightly higher correlation with the Vocabulary, Absurdities, and Verbal Reasoning Area; fifth, the Memory for Digits subtest has a higher correlation with the Vocabulary and Memory for Sentences subtests; sixth and finally, the Abstract / Visual Reasoning Area has a higher correlation with the Vocabulary and Absurdities subtests. Taken as a whole, however, the correlations reported in Table 21 are indeed very similar to those reported by Thorndike et al. for the corresponding age range.

The intercorrelations between the Stanford-Binet IV subtests, Areas, and Composite for the age group 13-0 to 16-11 are reported in Table 22; the number of pairwise cases used to calculate each correlation can be obtained from Appendix A.1.3. For the ten subtests appropriate for this age level (see Chapter 3) all intercorrelations between tests, Areas, and the Composite were significant ($p < .05$); in fact, all correlations were significant even at the .001 level. Comparison of Table 22 with the intercorrelations for the appropriate age range in Thorndike et al. (1986, pp. 121-124) indicates that the correlations in the current sample (13-0 to 16-11 age group) are inflated. This is to be expected given the extremely large standard deviations for this group on subtests, Areas and the Composite on the Stanford-Binet IV (see Table 16); and also, the extreme heterogeneity of this age group (see Figure 10). Therefore, the factorial findings for this age group must be regarded cautiously, as the artificially high correlations may operate to produce an artifactual general factor.

Intercorrelations Between the Stanford-Binet IV Subtests, Areas, and Composite and the WISC-R Subtests and Scales

The intercorrelations between the Stanford-Binet IV subtests, Areas, and Composite with the WISC-R subtests and scales for the age group 6-0 to 8-11 are reported in Table 23. Although twelve tests from the Stanford-Binet IV have been included, the Matrices, Number Series, Memory for Digits, and Memory for Objects subtests are not applicable to all children in this age group, and therefore have far lower numbers of cases on which the pairwise correlations are based; also, because of the type of children who take these subtests at this age level, a truncation of range is present which will tend to lower actual correlations. Therefore, the correlations of these four tests with other tests and scales must be interpreted with extreme caution--to see how these subtests relate to others, one would be best advised to inspect Table 24 which reports the intercorrelations for the 9-0 to 12-11 age group--where these four subtests are appropriate for the "average" child.

The intercorrelations between the eight Stanford-Binet IV subtests appropriate for the youngest age (see Chapter 3), the four Areas, and the Composite with the twelve WISC-R subtests and three scales were all significant ($p < .05$) with the following exceptions: the Picture Completion subtest was not significantly related to the Stanford-Binet IV Vocabulary subtest (p=.096), the Stanford-Binet IV Comprehension subtest (p=.074), the Copying subtest (p=.061), the Quantitative subtest (p=.379), the Memory for Sentences subtest (p=.171) or to the Quantitative Reasoning Area (p=.[illegible] the Object Assembly subtest was not significantly related to the Binet IV Comprehension subtest (p=.126); the Pattern Analysis subtest was not significantly related to the WISC-R Comprehension subtest (p=.071); the Bead Memory subtest was not significantly related to the Information subtest (p=.111) or to the Mazes subtest (p=.051); and the Memory for Sentences subtest was not significantly related to the Coding subtest (p=.056).

The intercorrelations between the Stanford-Binet IV subtests, Areas, and Composite with the WISC-R subtests and scales for the age group 9-0 to 12-11 are reported in Table 24. The intercorrelations between the twelve Stanford-Binet IV subtests appropriate for this age (see Chapter 3), the four Areas, and the Composite with the twelve WISC-R subtests and three scales were all significant ($p < .05$) with the following exceptions: Picture Arrangement was not significantly related to the Pattern Analysis subtest (p=.052), the Number Series subtest (p=.108), the Bead Memory subtest (p=.420), the Memory for Digits subtest (p=.146), the Memory for Objects subtest (p=.087) or to the Short-Term Memory Area (p=.064); the Coding subtest was not significantly related to the Quantitative subtest (p=.280), the Number Series subtest (p=.071) or to the Quantitative Reasoning Area (p=.086); and the Mazes subtest was not significantly related to the Memory for Sentences subtest (p=.059) or to the Memory for Digits subtest (p=.165).

The intercorrelations between the Stanford-Binet IV subtests, Areas, and Composite with the WISC-R tests and scales for the age group 13-0 to 16-11 are reported in Table 25. The intercorrelations between the ten Stanford-Binet IV subtests used for this age (see Chapter 3), the four Areas, and the Composite with the twelve WISC-R subtests and three scales were all significant ($p < .05$). Also, the intercorrelations between the Copying and Absurdities subtests with the other subtests and scales were all significant[43].

---

[43] Though the number of children in this age group is 56, the number of pairwise cases for the intercorrelations ranged from 29-31 for Copying and from 42-44 for Absurdities.

D. Principal Components Analyses of the WISC-R

In this section, the results from the principal components analyses of the WISC-R are reported. Each age group will be dealt with separately. All factoring done on the WISC-R alone was based on handling missing data in a pairwise fashion--very few data points were missing. For each group, a table providing the unrotated principal components (full component model) is given. From this table, the size of consecutive eigenvalues was obtained and used in both the eigenvalue-one and scree test criteria; the scree plot is also provided for inspection. Next, one or more tables are provided where factors have been rotated by the varimax criterion. These tables report truncated principal components with several number-of-factors solutions being reported as suggested by several theoreticians (Caroll, 1985; Comrey, 1978; Gorsuch, 1983; Hakstian et al., 1982; Walkey, 1983). Several number-of-factors solutions are provided to encompass the eigenvalue-one and scree test criterion (Gorsuch, 1983; Hakstian et al., 1982), as well as the number of factors hypothesized to underlie the test (Walkey, 1983). Generally, for the WISC-R, this meant rotating two-, three-, four-, and sometimes five-factor solutions--the two-factor solution being suggested by Wechsler's (1974) construction of the WISC-R with Verbal and Performance scales.

WISC-R Principal Components for the Age Group 6-0 to 8-11 Years

The unrotated principal components for the age group of 6-0 to 8-11 years is reported in Table 26. From this table, it can be seen that three of the principal components have eigenvalues greater than one. Therefore, by the eigenvalue-one criterion, three factors should be extracted and rotated. This also corresponds with Kaufman's (1975) work which suggests that three factors underlie the WISC-R. To determine the number of factors suggested by the scree test, the successive eigenvalues from Table 23 have been plotted in Figure 12.

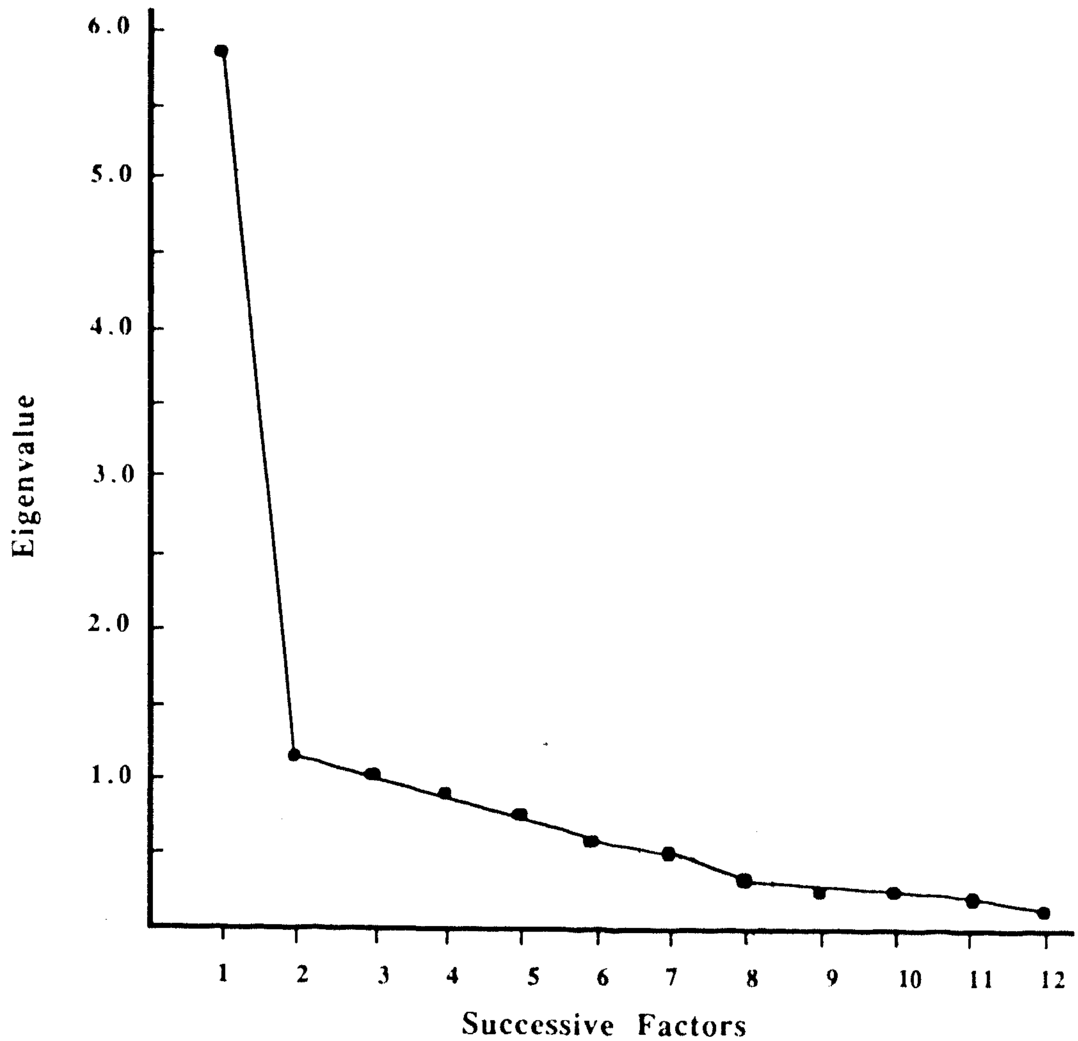
Table 26
Unrotated Principal Components For the WISC-R: Ages 6-0 to 8-11 Years[a]

WISC-R Age 6-0 to 8-11	Unrotated Principal Components											
Tests	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**
Information	.76	.04	-.41	-.08	.09	.27	-.31	-.14	-.04	.00	-.14	.18
Similarities	.77	-.26	-.20	-.23	.18	.11	.15	.23	.27	-.21	.07	.01
Arithmetic	.81	-.10	-.17	.08	-.31	.28	.11	-.21	.09	.10	-.04	-.21
Vocabulary	.78	-.13	-.27	.00	.20	-.26	-.26	.12	-.01	.28	.16	-.07
Comprehension	.71	-.44	-.06	.09	.09	-.16	.43	.02	.17	.10	-.14	.08
Digit Span	.76	-.15	.38	-.17	-.29	.21	.01	-.02	-.17	.00	.25	.10
Picture Completion	.42	-.02	.33	.77	.23	.22	-.09	.11	.03	-.01	.00	.01
Picture Arrangement	.80	.01	.14	.00	.31	-.27	-.07	-.31	-.09	-.25	.04	-.08
Block Design	.66	.51	.18	-.30	.18	.16	.04	.24	-.19	.01	-.12	-.12
Object Assembly	.68	.18	.57	-.19	-.03	-.15	-.02	-.08	.28	.15	-.11	.08
Coding	.68	.02	-.07	.19	-.55	-.29	-.18	.19	-.03	-.16	-.10	.00
Mazes	.43	.73	-.33	.23	-.05	-.09	.28	-.06	.05	.01	.13	.09
Eigenvalue	5.88	1.14	1.05	0.90	0.75	0.57	0.51	0.34	0.27	0.25	0.19	0.13
% of Total Variance	49.0	9.5	8.8	7.5	6.3	4.7	4.2	2.9	2.3	2.1	1.6	1.1
Cummulative % Var.	49.0	58.5	67.3	74.8	81.1	85.8	90.1	92.9	95.2	97.3	98.9	100

[a] Missing data was dealt with on a pairwise basis for this component analysis; n=46.

Figure 12

Scree Plot for the WISC-R for the Age Group 6-0 to 8-11 Years



The scree for this age group has several breaking points, making it difficult to determine the correct number of components suggested by this criterion; however, the upper most break in the scree suggests that five factors should be extracted. Therefore, two-, three-, four-, and five-component solutions were extracted and rotated by the varimax criterion. The five-factor solution is not reported because the fifth rotated principal component turned out to be a singlet with only Picture Completion loading on it; the remaining solutions are presented in Table 27.

Table 27

Two-, Three-, and Four-Component Varimax Rotated Principal Components For the WISC-R: Ages 6-0 to 8-11 Years[a]

WISC-R Age 6-0 to 8-11	Varimax Rotated Principal Components											
	Two Components Extracted			Three Components Extracted				Four Components Extracted				
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Information	.65	.40	.57	.76	.09	.41	.75	.76	.15	.39	-.05	.75
Similarities	.80	.14	.66	.80	.25	.08	.70	.81	.30	.02	-.07	.75
Arithmetic	.76	.29	.67	.74	.31	.22	.70	.74	.25	.20	.21	.70
Vocabulary	.75	.26	.63	.78	.21	.22	.70	.79	.19	.20	.10	.70
Comprehension	.84	-.05	.70	.76	.32	-.15	.70	.76	.19	-.16	.27	.71
Digit Span	.74	.23	.60	.46	.73	-.01	.74	.49	.70	-.11	.17	.77
Picture Completion	.38	.18	.18	.17	.51	.01	.29	.14	.14	.09	.91	.88
Picture Arrangement	.70	.39	.63	.53	.57	.22	.65	.54	.52	.16	.24	.65
Block Design	.34	.77	.70	.19	.56	.62	.73	.22	.72	.49	-.09	.82
Object Assembly	.51	.48	.49	.16	.86	.20	.81	.20	.88	.07	.18	.85
Coding	.59	.34	.47	.54	.33	.26	.47	.54	.25	.25	.30	.51
Mazes	.03	.85	.72	.17	.03	.89	.82	.15	.09	.91	.13	.88
Sum of Squared Loadings	4.81	2.23		3.85	2.59	1.63		3.94	2.39	1.44	1.20	
% of Total Variance	40.1	18.6		32.1	21.6	13.6		32.8	19.9	12.0	10.0	
Cummulative % Var.	40.1	58.7		32.1	53.7	67.3		32.8	52.7	64.7	74.7	

[a] Missing data was dealt with on a pairwise basis for this component analysis; n=46.

From Table 27, it can be seen that the four-component solution also has a singlet--with only Picture Completion having a substantial loading on the fourth component; therefore, the four-component solution is not appropriate. The three-component solution seems the most appropriate in terms of interpretability, and it also is the solution supported by the eigenvalue-one criterion. The two-component solution, on the other hand, appears to compress too much variance onto the first component and the second component consists of only the Block Design, Object Assembly, and Mazes subtests, with an additional moderate loading from Information.

Thus, the three-component solution has been chosen as the correct solution, and it will now be discussed in more detail. The first component is Verbal in nature and will be referred to as Verbal Comprehension as is traditionally the case (see Kaufman, 1975). On this component the Information, Similarities, Arithmetic, Vocabulary and Comprehension subtests have high loadings, just as found by Kaufman (1975) for the WISC-R standardization sample. Furthermore, moderate loadings on this component include Digit Span, Picture Arrangement, and Coding. These moderate loadings are not without precedence: Digit Span has been found to load on the Verbal Comprehension factor (Kaufman, 1975) as h[illegible] Picture Arrangement (Hodges 1982; Karnes & Brown, 1980) and Coding (Karnes & Brown, 1980).

The second component, Perceptual Organiz[illegible] strong or moderate loadings from Digit Span, Picture Completion, Picture Arrangement, Block Design, and Object Assembly; all of these subtests, with the exception of Digit Span are commonly found to have loadings on the Perceptual Organization factor (e.g., Kaufman, 1985). Also, Groff and Hubble (1982) found that Digit Span had its highest loading on the Perceptual Organization factor for children with low intellectual ability.

The third component was fairly unique with Block Design and Mazes having strong loadings and with Information having a moderate loading. Whether or not this component represents Freedom From Distractibility is questionable as the traditional markers of Digit Span, Arithmetic, and Coding did not load on this component. Although Information is known to load on the third factor (Kaufman, 1975; Sapp et al., 1985) as is Block Design (Hodges, 1982; Petersen & Hart, 1979; Sapp et al., 1985), there does not appear to be a clear precedence for Mazes loading on the third factor[44]. Taken as a whole, however, the

[44] It must be remembered, however, that most often the Mazes subtest is omitted from the factor analyses done on the WISC-R, so its loading on the Perceptual Organization factor is best substantiated for "normal" children (i.e., Kaufman, 1975).

third factor has a visual-motor nature that involves planning and nonverbal reasoning; for this reason the third component will be referred to as Visual Planning and Reasoning.

Taken as a whole, the first two components in the 6-0 to 8-11 age group do resemble Verbal Comprehension and Perceptual Organization as defined by Kaufman's (1975) prototypical factor pattern. Slight variations in these two factors from Kaufman's findings are not surprising given that the current sample is clinical in nature--indeed, most of the unique loadings have some precedence in other clinical samples. On the other hand, the third component for this age group does appear to be unique. The inclusion of Mazes in the analysis appears to have had a strong influence in determining the third component; this is not surprising given that several of the correlations between Mazes and other subtests were not significant for this age group. Although the nature of the third component is not totally clear, it does seem to have a Visual Planning and Reasoning element. Overall, the WISC-R factor structure for this age group is similar enough to that found by previous researchers to consider the sample adequate for examining the factor structure of the Stanford-Binet IV.

WISC-R Principal Components for the Age Group 9-0 to 12-11 Years

The unrotated principal components (full component model) for the WISC-R for the age group 9-0 to 12-11 years are reported in Table 28. Only two of these components have eigenvalues greater than one though a third component is close to one (eigenvalue of .95); hence, the eigenvalue-one criteria suggests that a two-components solution is appropriate.

Table 28

Unrotated Principal Components For the WISC-R: Ages 9-0 to 12-11 Years[a]

WISC-R Age 9-0 to 12-11 **Tests**	**Unrotated Principal Components**											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Information	.85	-.28	-.19	.10	.14	-.06	.09	-.07	.00	-.25	-.18	-.16
Similarities	.83	-.12	-.01	-.18	.17	.33	-.07	-.07	-.24	.18	-.14	.04
Arithmetic	.80	-.22	.00	.09	.02	-.03	-.49	.21	.13	-.01	.00	.02
Vocabulary	.86	-.28	.08	-.19	-.05	-.06	.16	-.07	.14	-.09	-.03	.28
Comprehension	.86	-.17	-.10	-.20	-.02	.21	.08	-.15	.22	.10	.19	-.14
Digit Span	.67	-.37	-.01	.48	.17	-.23	.19	.13	-.10	.18	.09	.00
Picture Completion	.71	.23	.18	-.48	-.03	-.07	.16	.37	-.06	-.03	.02	-.06
Picture Arrangement	.47	.51	.36	.13	.58	.03	-.04	-.09	.02	-.10	.08	.03
Block Design	.83	.15	-.17	.01	-.31	-.09	-.11	-.14	-.26	-.14	.18	.02
Object Assembly	.72	.48	-.04	-.07	-.12	-.36	-.05	-.17	.08	.18	-.13	-.03
Coding	.47	.06	.65	.36	-.42	.18	.05	-.01	.02	-.01	-.06	-.04
Mazes	.47	.52	-.53	.33	-.09	.25	.11	.16	.09	-.01	-.03	.07
Eigenvalue	6.36	1.23	0.95	0.85	0.72	0.45	0.38	0.31	0.24	0.21	0.16	0.14
% of Total Variance	53.0	10.3	7.9	7.1	6.0	3.8	3.1	2.6	2.0	1.7	1.3	1.1
Cummulative % Var.	53.0	63.3	71.2	78.3	84.3	88.0	91.2	93.8	95.8	97.5	98.9	100

[a] Missing data was dealt with on a pairwise basis for this component analysis; n=66

To determine the number of components to extract and rotated according to the scree test criterion, the successive eigenvalues from Table 27 have been plotted in Figure 13. Once again, there are several breaking points in the scree. Either a two- or four-component solution seems to be indicated by the scree plot. Therefore, two-, three-, and four-component solutions rotated by the varimax criterion are reported in Table 29.

Figure 13

Scree Plot for the WISC-R for the Age Group 9-0 to 12-11 Years

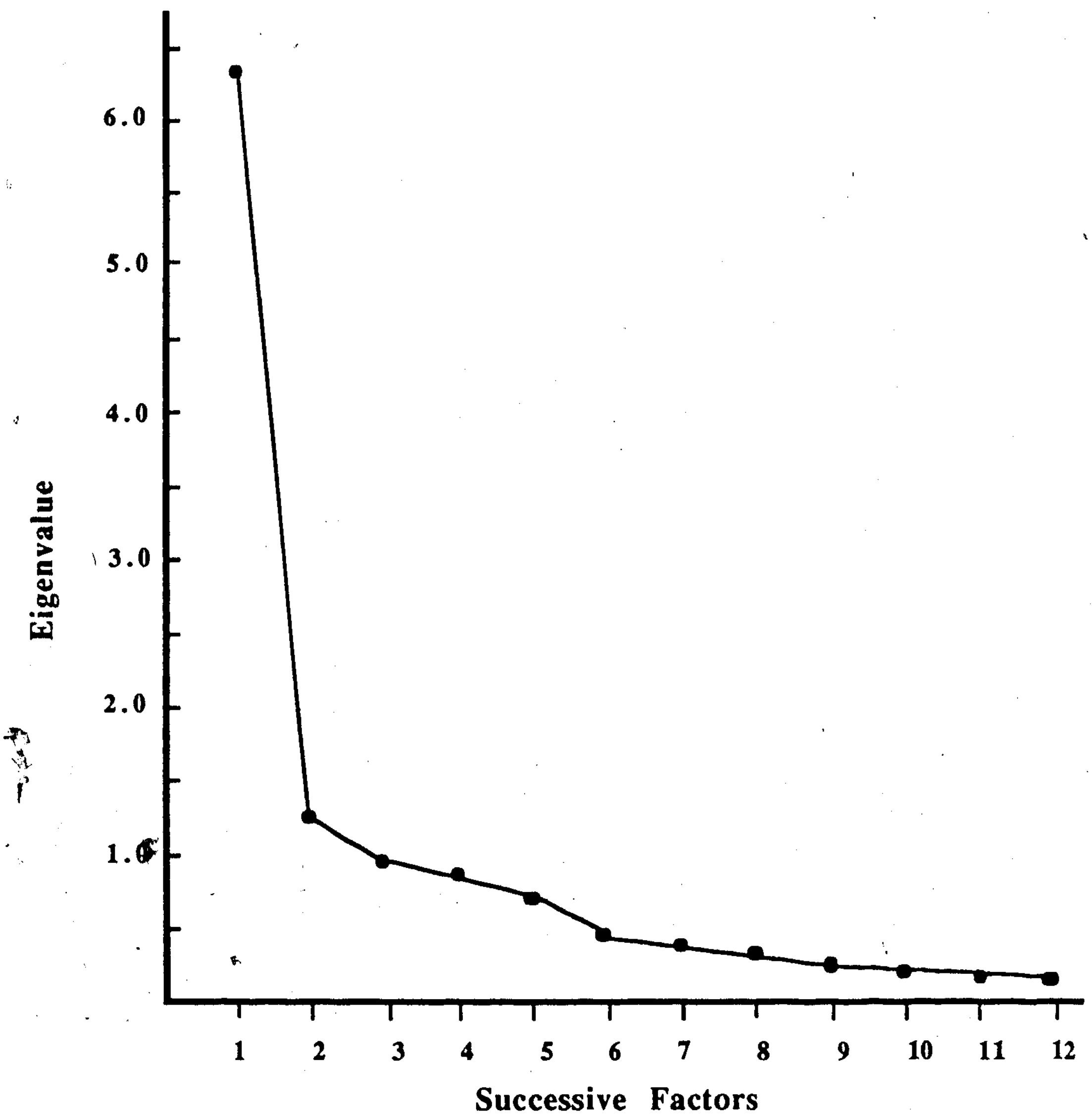


Table 29

Two-, Three-, and Four-Component Varimax Rotated Principal Components For the WISC-R: Ages 9-0 to 12-11 Years[a]

WISC-R Age 9-0 to 12-11	Varimax Rotated Principal Components												
	Two Components Extracted			Three Components Extracted				Four Components Extracted					
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Information	.86	.24	.79	.87	.26	.08	.83	.85	.25	.24	.05	.84
Similarities	.76	.36	.71	.75	.26	.28	.71	.65	.54	.14	.11	.74
Arithmetic	.79	.26	.69	.78	.17	.23	.69	.75	.28	.15	.20	.70
Vocabulary	.87	.24	.82	.85	.10	.30	.83	.75	.53	-.02	.14	.86
Comprehension	.81	.33	.76	.81	.28	.20	.77	.71	.54	.15	.02	.81
Digit Span	.76	.06	.58	.76	.02	.10	.58	.83	-.14	.15	.28	.82
Picture Completion	.46	.59	.56	.43	.34	.54	.59	.24	.84	.11	.21	.82
Picture Arrangement	.11	.69	.49	.06	.34	.71	.62	.01	.34	.33	.64	.64
Block Design	.61	.58	.71	.61	.54	.27	.74	.54	.44	.47	.15	.74
Object Assembly	.34	.80	.76	.32	.65	.48	.76	.22	.57	.55	.30	.76
Coding	.36	.31	.23	.28	-.16	.74	.65	.30	.07	-.06	.83	.78
Mazes	.11	.69	.49	.15	.87	.00	.77	.18	.06	.92	.03	.88
Sum of Squared Loadings	4.79	2.80		4.65	1.98	1.92		3.98	2.37	1.64	1.41	
% of Total Variance	39.9	23.3		38.8	16.5	16.0		33.2	19.8	13.7	11.8	
Cummulative % Var.	39.9	63.1		38.8	55.3	71.3		33.2	53.0	66.7	78.5	

[a] Missing data was dealt with on a pairwise basis for this component analysis; n=66

From Table 29, the two-component solution is the easiest to interpret and follows the prototypic pattern identified by Kaufman (1975) for two-factors. The first component is clearly similar to Verbal Comprehension as outlined by Kaufman. High loadings on this factor include Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Digit Span; moderate loadings include Picture Completion and Block Design--both of which have small loadings on the first component in normal samples (Kaufman, 1975). The second component includes strong or moderate loadings from all the Performance scales subtests with the exception of Coding.

In the three-component solution, the first component is clearly Verbal Comprehension, with high loadings from all the Verbal Scale subtests; once again, moderate loadings on this component from Picture Completion and Block Design are noted. The second and third components have loadings from the Performance Scale subtest; in fact, three of the six Performance Scale subtests load on each of these two components: the second component having loadings from Block Design, Object Assembly, and Mazes; and the third component having loadings from Picture Completion, Picture Arrangement, and Coding just as found by Petersen and Hart (1979) in their group of slow learners and learning disabled children; furthermore, Object Assembly has a moderate loading on this third component. Worth noting, scores on Picture Arrangement, Coding, and Object Assembly all are influenced by speed; to a lesser extent the 20 second time limit for Picture Completion makes time an important aspect. Thus, the third factor may be similar to Freedom from Distractibility, or it may alternately be conceived of as Broad Speediness symbolized by Gs (Horn &Cattell, 1966; Undheim, 1981a); the latter is perhaps more defensible in that neither the Digit Span nor Arithmetic subtests load on the third component.

Comparing the three-component solution to the two-component solution, it can be noted that in the three-component solution the Performance Scale subtests that load on the second component are different than those that load on the second component in the two-component solution. By extracting three components rather than two, the Perceptual Organization component is fractionated into two smaller components. Similarly, when a four-component solution is examined, this fractionating of the Performance Scale subtest onto the second, third, and fourth components is even more evident.

Because the three- and four-component solutions tend to fractionate the Perceptual Organization component, and because the eigenvalue-one criteria suggests only two components, the two-component solution is taken as the best. Furthermore, these two components are similar to Kaufman's factors of Verbal Comprehension and Perceptual Organization. Hence, the two-component solution in the current age group validates the current sample (age group 9-0 to 12-11) as being similar to other clinical populations in terms of the WISC-R factor structure.

WISC-R Principal Components for the Age Group 13-0 to 16-11 Years

The unrotated Principal Components for the age group 13-0 to 16-11 years are reported in Table 30. Only one component has an eigenvalue greater than one, and hence the eigenvalue-one criteria suggests that only one factor be extracted--this would be the first unrotated principal component from Table 30. To determine the number of components suggested by the scree test, the successive eigenvalues have been plotted in Figure 14.

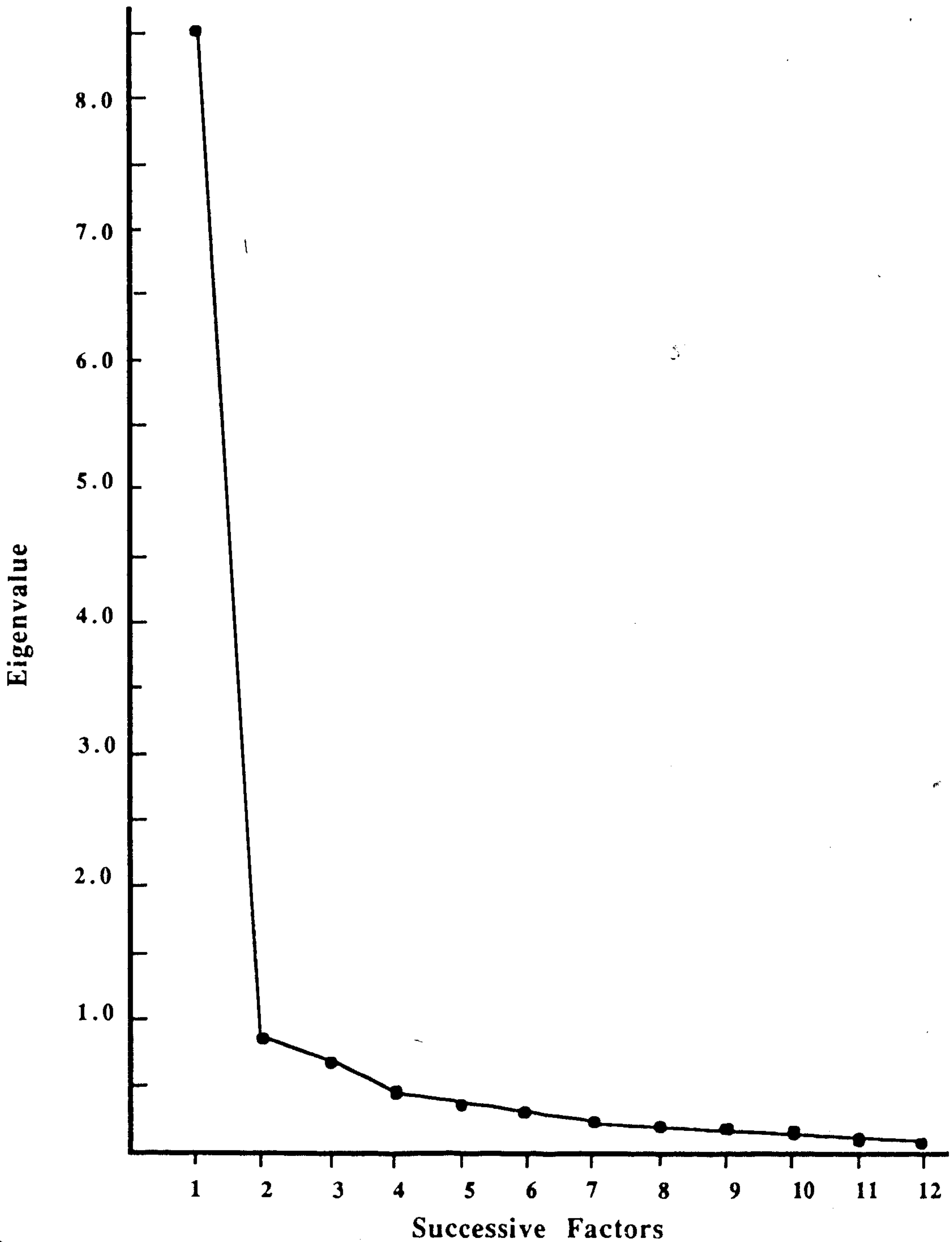
Table 30

Unrotated Principal Components For the WISC-R: Ages 13-0 to 16-11 Years[a]

WISC-R Age 13-0 to 16-11	Unrotated Principal Components											
Tests	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**
Information	.90	-.29	-.03	-.03	-.15	-.03	.11	-.06	.03	-.01	-.25	.05
Similarities	.94	-.14	.00	-.09	.04	.07	-.11	-.06	-.03	-.10	-.01	-.23
Arithmetic	.87	-.09	-.24	-.14	.18	-.02	.10	.30	.02	-.14	.02	.05
Vocabulary	.87	-.35	.01	-.06	-.09	-.03	-.03	-.22	-.03	-.13	.15	.11
Comprehension	.88	-.26	.04	-.09	.21	.09	.11	-.02	-.16	.24	.04	.00
Digit Span	.84	-.17	.06	.31	.13	-.32	.00	.03	.18	.06	.03	-.03
Picture Completion	.76	.43	.11	-.39	-.10	-.11	.15	-.06	.15	.05	.04	-.02
Picture Arrangement	.80	.24	-.39	.10	.11	.25	-.13	-.11	.17	.06	-.01	.05
Block Design	.88	.04	-.09	-.08	-.27	-.11	-.29	.16	-.09	.10	.01	.03
Object Assembly	.83	.21	-.08	.33	-.29	.12	.22	.04	-.08	-.01	.06	-.05
Coding	.72	.03	.63	.07	.02	.22	-.08	.10	.08	-.03	-.01	.05
Mazes	.77	.50	.08	.10	.22	-.15	-.04	-.09	-.20	-.08	-.06	.04
Eigenvalue	8.50	0.86	0.64	0.43	0.35	0.30	0.23	0.21	0.17	0.13	0.09	0.08
% of Total Variance	70.9	7.2	5.4	3.6	2.9	2.5	1.9	1.8	1.4	1.1	0.8	0.7
Cummulative % Var.	70.9	78.1	83.4	87.0	89.9	92.4	94.3	96.0	97.5	98.5	99.3	100

a Missing data was dealt with on a pairwise basis for this component analysis; n=56

Figure 14
Scree Plot for the WISC-R for the Age Group 13-0 to 16-11 Years



From Figure 14, the number of components to extract and rotate according to the scree test is three; a clear break exists in the scree line between the third and fourth eigenvalues. To be consistent with the two younger age groups, two-, three-, and four-component solutions are rotated by the varimax criterion and reported in Table 31.

Table 31
Two-, Three-, and Four-Component Varimax Rotated Principal Components For the WISC-R: Ages 13-0 to 16-11 Years[a]

WISC-R Age 13-0 to 16-11	Varimax Rotated Principal Components											
	Two Components Extracted			Three Components Extracted				Four Components Extracted				
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Information	.88	.35	.90	.86	.32	.25	.90	.84	.30	.19	.26	.90
Similarities	.81	.49	.90	.78	.45	.30	.90	.78	.34	.34	.28	.91
Arithmetic	.73	.49	.76	.75	.51	.05	.82	.74	.39	.37	.02	.84
Vocabulary	.90	.29	.89	.87	.25	.26	.89	.86	.23	.17	.27	.89
Comprehension	.84	.36	.84	.81	.31	.31	.84	.80	.24	.25	.29	.85
Digit Span	.76	.40	.74	.72	.35	.32	.74	.64	.51	.00	.41	.84
Picture Completion	.32	.81	.76	.27	.74	.39	.78	.32	.28	.82	.25	.93
Picture Arrangement	.46	.70	.70	.52	.76	-.08	.85	.46	.72	.36	-.05	.86
Block Design	.65	.59	.77	.64	.57	.21	.78	.63	.43	.42	.18	.79
Object Assembly	.51	.69	.73	.50	.66	.22	.73	.40	.74	.20	.30	.84
Coding	.54	.49	.52	.39	.27	.83	.92	.37	.17	.29	.82	.92
Mazes	.28	.88	.85	.23	.81	.38	.86	.19	.64	.54	.37	.87
Sum of Squared Loadings	5.42	3.96		5.05	3.46	1.51		4.69	2.50	1.77	1.47	
% of Total Variance	45.2	33.0		42.1	28.8	12.6		39.1	20.8	14.8	12.3	
Cummulative % Var.	45.2	78.2		42.1	70.9	83.5		39.1	59.9	74.7	87.0	

[a] Missing data was dealt with on a pairwise basis for this component analysis; n=56

In examining the rotated components for this age group, it must be remembered that the correlations between the WISC-R subtests were extremely high because of the heterogeneity of the sample. As such, a strong general component was expected[45]. From Table 30 the size of the first unrotated component (eigenvalue of 8.50) dwarfs all others, accounting for 70.9% of the total subtest variance. As a result, the remaining components account for relatively little of the total subtest variance. Because of the extremely high correlations between subtests, the factor structure for this age group must be considered cautiously as the factor loadings will be artifactually inflated. This effect can be seen in the fact that only one unrotated component has an eigenvalue greater than one; all other published research on the WISC-R reports at least two components with eigenvalues equal to or greater than one. Hence, the sample for the age group 13-11 to 16-11 is atypical and must be viewed with skepticism.

A one-component solution for this age group can be obtained from Table 30 by taking the first unrotated component. All of the twelve WISC-R subtests have high loadings on this component: only Picture Completion, Coding, and Mazes have loadings in the .72-.77 range while all other subtest have loadings above .80. By conventional interpretation (Kaufman, 1975), this unrotated component can be regarded as general intelligence (*g*). Loadings this high on the first unrotated component have not been reported in the literature, reaffirming the atypical nature of this age grouping in the current sample.

The general factor is so strong that when a two-component solution is examined, the first component resembles Verbal Comprehension, but moderately strong loadings from most of the Performance Scale subtests are also found. Similarly, the second component is similar to Perceputal Organization with most of the Performance Scale subtest having

45 This is the reverse of Thurstone's (1938) problem where the extreme homogeneity of his sample of college students resulted in the masking of a general factor. As such, the extreme heterogeneity of this age group will almost guarantee a strong general factor, and exaggerated loading on the first unrotated component.

strong or moderate loadings, but Similarities and Arithmetic also have moderate loadings on this component; the factor pattern is far more complex than that found by other researchers (e.g, Kaufman, 1975).

In the three-component solution, the first component has its highest loadings from the Verbal Scale subtest, but several of the Performance Scale subtests also have moderate loadings on this component, causing it to have some semblance of a general component; the factor pattern still remains complex. The second component is nearly identical to the second component from the two-component solution; all the Performance Scale subtest with the exception of Coding load on the second component. The third component is a singlet, with only Coding loading on it. The presence of the singlet suggests that overfactoring has occurred and that the three-component solution is not the best.

When four-components are extracted and rotated the pattern of loadings on the second, third, and fourth components indicates that fractionating of the second component from the two-component solution has occurred. Therefore, the four-component solution is complex, difficult to interpret and inferior to one-, and two-component solutions.

Overall, the various solutions for the 13-0 to 16-11 age group indicate that the typical WISC-R factor structure (Kaufman, 1975) does not exist. This is not surprising given the extreme heterogeneity of this age group and the inflated correlations that result from this. Consequently, the atypical nature of this group suggests that any results from it should be regarded as tentative; that is, the pattern of loadings may be in the correct direction, but the magnitude and complexity of the factor loadings is bound to be higher than in other more homogeneous samples. In terms of implications for interpreting the results of the Stanford-Binet IV for this age group, it is possible that the component structure that emerges may give some indications of which subtests load on which components, but the number of components is bound to be obscured by the influence of the artifactually high general factor. Thus, all factorial results from the age group of 13-0 to 16-11 years must be treated as highly tentative.

E. Principal Components Analyses of the Stanford-Binet IV

In this section, the results from the principal components analyses of the Stanford-Binet IV are reported. For organizational purposes, the three age groups will be dealt with consecutively. That is, the age group of 6-0 to 8-11 years will be dealt with first, followed by the age group 9-0 to 12-11 years, and finally the age group 13-0 to 16-11 years. Within each age group, the following order will be maintained: first, the unrotated principal components are reported, then the scree plot is given, then the quartimax rotated components, and finally the varimax rotated components. For the quartimax and varimax rotations, several number-of-components solutions have been reported, as suggested by a combination of the eigenvalue-one criterion, the scree test criterion, and the hypothesized number of factors (which is four).

It should be noted that the quartimax rotated components precede the varimax rotated components. This order was chosen to allow comparison with Thorndike et al.'s findings (1986) prior to using varimax rotations. Although the quartimax criterion for rotation is not as well known as the varimax criterion, the former is highly desirable because it allows for the presence of a general factor simultaneous to group factors (Gorsuch, 1983). Likewise, though the quartimax method is not identical to the procedure used by Thorndike et al. (1986), the general effect is similar, though the order in which the group factors is extracted is not controlled as it was by Thorndike et al.. The varimax rotated components are also reported so that continuity with the WISC-R tradition initiated by Kaufman (1975) can be maintained. In all fairness, to compare the WISC-R with the Stanford-Binet IV it is imperative that the two tests be factored using the same methods. Furthermore, by reporting both the quartimax and varimax rotations, it is easier to demonstrate how choice of a factoring procedure can bias conclusions.

Stanford-Binet IV Principal Components for the Age Group 6-0 to 8-11 Years

Table 32 reports the unrotated principal components for the Stanford-Binet IV for the youngest age group. From this table, it can be seen that only two components in the full component model have eigenvalues of one or greater. Thus, by the eigenvalue-one criterion only two components should be rotated. This contrasts with the WISC-R results in the same group where three components had eigenvalues greater than one. In order to determine the number of components suggested by the scree test, the successive eigenvalues for the components for this age group have been plotted in Figure 15.

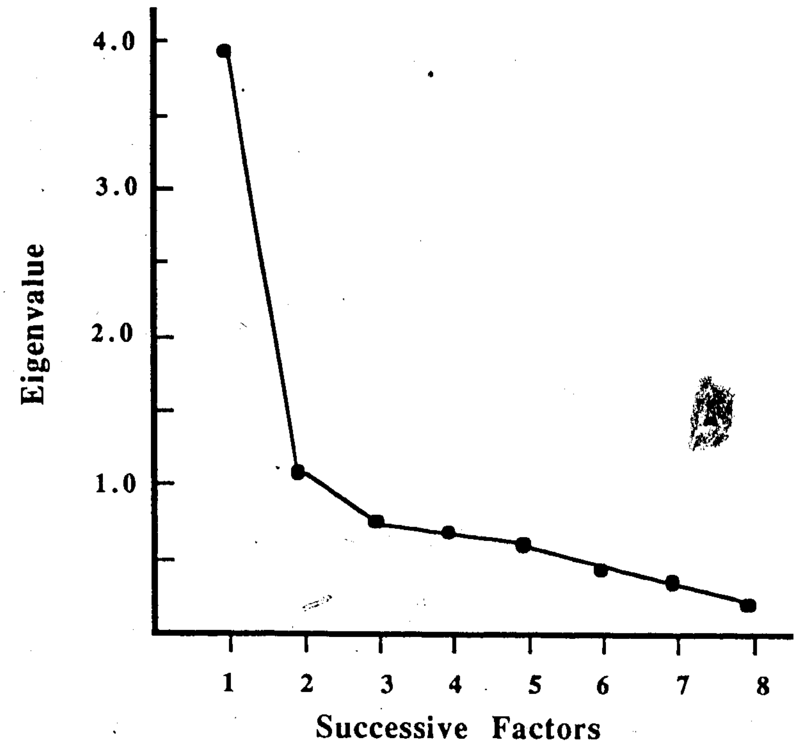
Table 32
Unrotated Principal Components for the Stanford-Binet IV: Ages 6-0 to 8-11 Years[a]

Stanford-Binet IV Age 6-0 to 8-11	Unrotated Principal Components							
Tests	I	II	III	IV	V	VI	VII	VIII
Vocabulary	.85	-.03	-.31	-.20	-.05	-.18	-.07	-.32
Comprehension	.81	.14	-.17	-.08	-.36	-.29	.15	.23
Absurdities	.78	-.17	.20	.35	.03	-.09	-.44	.07
Pattern Analysis	.72	.13	.40	-.05	-.40	.37	.05	-.08
Copying	.67	-.16	.14	-.59	.35	.08	-.05	.12
Quantitative	.63	-.55	.20	.30	.20	-.08	.35	-.05
Bead Memory	.37	.81	.27	.12	.30	-.14	.09	-.05
Memory for Sentences	.68	.16	-.54	.24	.18	.35	.04	.07
Eigenvalue	3.95	1.07	0.75	0.68	0.58	0.41	0.35	0.19
% of Total Variance	49.4	13.4	9.4	8.5	7.3	5.2	4.4	2.4
Cummulative % Var.	49.4	62.8	72.2	80.7	88.0	93.2	97.6	100.0

[a] The unrotated principal components listed in this table are based on the entire sample for this age group (n=46). Missing data were estimated through the regression procedure outlined previously.

Figure 15

Scree Plot for the Stanford-Binet IV at Ages 6-0 to 8-11 Years



Examination of Figure 15 reveals that only one prominent breaks exists in the scree line. This break, like the eigenvalue-one criteria, also suggests that two components should be rotated. However, to be consistent with the findings of Thorndike et al. (1986) for their 2 through 6 year-old group (*g* , Verbal, and Abstract/Visual factors), three components were also rotated; similarly, to be consistent with Thorndike et al.'s findings in their 7 through 11 year-old age group (*g* , Verbal, Memory, and Abstract/Visual factors) and also to be co[illegible]nt with the four Area scores on the Stanford-Binet IV, a four component solutio[illegible] was also rotated. The quartimax rotated two-, three-, and four-component solutions are reported in Table 33.

Table 33

Two-, Three-, and Four-Component Quartimax Rotated Principal Components For The Stanford-Binet IV: Ages 6-0 to 8-11 Years[a]

Stanford-Binet IV Age 6-0 to 8-11	Quartimax Rotated Principal Components											
	Two Components Extracted			Three Components Extracted				Four Components Extracted				
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Vocabulary	.85	.03	.72	.71	.55	.01	.82	.83	.21	.00	.36	.86
Comprehension	.80	.19	.67	.67	.46	.21	.70	.73	.23	.22	.27	.71
Absurdities	.79	-.12	.63	.82	.04	.04	.67	.43	.76	.18	.05	.79
Pattern Analysis	.71	.18	.54	.74	-.09	.37	.70	.29	.48	.46	.41	.70
Copying	.68	-.12	.48	.70	.05	.01	.50	.37	.21	.01	.81	.85
Quantitative	.67	-.51	.70	.78	-.12	-.34	.74	.26	.85	-.18	.11	.83
Bead Memory	.32	.83	.79	.21	.10	.90	.86	.23	-.02	.91	-.01	.88
Memory for Sentences	.67	.20	.49	.45	.76	.08	.78	.88	.13	.07	-.18	.84
Sum of Squared Loadings	3.95	1.09		3.51	1.13	1.11		2.52	1.69	1.16	1.07	
% of Total Variance	49.4	13.6		43.9	14.1	13.9		31.5	21.1	14.5	13.4	
Cummulative % Var.	49.4	63.0		43.9	58.0	71.9		31.5	52.6	67.1	80.5	

[a] The rotated principal components listed in this table are based on the entire sample for this age group (n=46). Missing data were estimated through the regression procedure outlined previously.

Because the two-component solution is supported by both the eigenvalue-one and scree test criteria, it will be examined first. From Table 33 it can be seen that the first component in the two-component solution is a general component with moderate or strong loadings from all subtests except Bead Memory. The second component is bipolar; one pole is Bead Memory, while the other pole is identified by Quantitative. The exact nature of this second component is not clear. Tentatively, it may be considered a Memory or Metamemory component representing Simultaneous Processing and Planning Strategies (Bead Memory) as opposed to Simple Recall of Rehearsed Information taught in school

(Quantitative). If nothing else, the composition of the second component indicates that Bead Memory does not fit that well with the other subtests at this age level; rather, it seems to represent a specific skill that may not be related to general intelligence.

Examination of the three-component quartimax solution for the 6-0 to 8-11 age group indicates that the general component remains intact. With the exception of Bead Memory, all the other subtests have moderate or high loadings on the first component. The second component is clearly Verbal, with loadings from Vocabulary, Comprehension, and Memory for Sentences. The third component is identified primarily by Bead Memory though Pattern Analysis has a weak loading. This suggests that the third component involves a Visual component coupled with Nonverbal Reasoning and perhaps could be considered Abstract / Visual in that the skill required is ability to synthesize visual parts into a whole. In general, the three-component quartimax solution has some similarity to the factor structure found by Thorndike et al. (1986) for their age grouping of 2 through 6 years of age.

Interpretation of the four-component quartimax rotated components for the 6-0 to 8-11 age group is complex. The general factor has been dispersed, and the first factor represents Verbal skills with the Vocabulary, Comprehension, and Memory for Sentences subtests having strong loadings. The second component is identified by high loadings from the Absurdities and Quantitative subtests and a moderate loading from the Pattern Analysis subtest; this component seems to be Acquired Knowledge that is non-verbal in nature. The third component has loadings from Bead Memory and Pattern Analysis and deals with Non-Verbal Reasoning including Ability to Organize Visual Material--something like Broad Visualization. Similarly, the fourth component has a Visual nature, but may relate more to the Visual-motor Ability than Visual Reasoning ability.

Taken as a whole, both the two-, and three-component quartimax rotations for the 6-0 to 8-11 age group hold some promise. The clearest finding, taken from the two-component solution, is that all subtests load on a general factor, with the exception of Bead

Memory, which largely forms the second component. This general factor as represented by the first component is similar to the first unrotated component (see Table 32) and accounts for 49.4 % of the total subtest variance, which is similar to that found by Thorndike et al who found the general factor for the 2 through 6, as well as the 7 through 11 year-old groups to account for 42% of the total subtest variance. Likewise, the three component solution does support the notion that for the younger age group the Stanford-Binet IV may measure *g* , Verbal ability, and Abstract/Visual ability; in the three-component solution the general factor accounts for 43.9% of the total subtest variance while the other two components both account for about 14%.

Table 34

Two-, Three-, and Four-Component Varimax Rotated Principal Components For The Stanford-Binet IV: Ages 6-0 to 8-11 Years[a]

Stanford-Binet IV Age 6-0 to 8-11	Varimax Rotated Principal Components											
	Two Components Extracted			Three Components Extracted				Four Components Extracted				
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Vocabulary	.80	.29	.72	.46	.77	.09	.82	.70	.25	.54	.06	.86
Comprehension	.70	.43	.67	.42	.67	.28	.70	.61	.26	.43	.27	.71
Absurdities	.78	.13	.63	.73	.33	.16	.67	.33	.78	.17	.23	.79
Pattern Analysis	.62	.39	.54	.65	.18	.49	.70	.13	.47	.48	.49	.70
Copying	.69	.10	.48	.63	.30	.12	.50	.17	.20	.88	.04	.85
Quantitative	.79	-.28	.70	.82	.17	-.21	.74	.17	.86	.20	-.14	.83
Bead Memory	.05	.89	.79	.01	.16	.91	.86	.17	-.03	.03	.92	.88
Memory for Sentences	.58	.40	.49	.13	.87	.10	.78	.88	.20	.02	.13	.84
Sum of Squared Loadings	3.56	1.48		2.43	2.08	1.25		1.85	1.78	1.55	1.25	
% of Total Variance	44.5	18.5		30.4	26.0	15.6		23.1	22.3	19.4	15.6	
Cummulative % Var.	44.5	63.0		30.4	56.4	72.0		23.1	45.4	64.8	80.4	

[a] The rotated principal components listed in this table are based on the entire sample for this age group (n=46). Missing data were estimated through the regression procedure outlined previously.

The results from the varimax rotated two-, three- and four-component solutions are reported in Table 34; however, these solutions do not clarify the situation much. In the two-component varimax rotated solution, the first component resembles the general component found in the two-component quartimax solution--all subtests having moderate or strong loadings except Bead Memory, and this component accounts for 44.5 % of the total subtest variance. Finding such a general component when the varimax criterion is used is unusual. The second component is also difficult to interpret, having its strongest loadings from Bead Memory, Comprehension, Memory for Sentences, and Pattern Analysis; as such, it could be conceived of as Ability to Perceive and Organize Material or perhaps Ability to Form Concepts.

The three-component varimax rotated solution is slightly easier to interpret. The first component has loadings from Absurdities, Pattern Analysis, Copying, and Quantitative. This component appears to involve Non-Verbal Reasoning and may be similar to Vernon's Spatial-Mechanical factor or to Fluid Intelligence, though all four subtests also appear to deal with attention to detail and carefulness. The second component is clearly Verbal, with strong loadings from Vocabulary, Comprehension, and Memory for Sentences. The third component once again seems to represent Ability to Perceive and Organize Materials.

The four-component varimax rotated solution is quite similar to the three-component solution, though the order in which the components emerges changes slightly, and there is a splitting of one of the components into two. In the four-component solution, the first component is Verbal, with loadings from Vocabulary, Comprehension, and Memory for Sentences. The second component has its highest loadings from Absurdities, Pattern Analysis and Quantitative, indicating Visual Sequencing of Information. The third component relates to Visual Discrimination and Concept Formation and has high loadings from Copying, Pattern Analysis, Vocabulary, and Comprehension. The fourth component is Abstract / Visual in nature with loadings from Bead Memory and Pattern Analysis.

Taken as a whole, both the two-, and three-component solutions hold merit, just as in the quartimax rotations. The two-component solution for the varimax rotations is probably the best fit with the data given that it is supported by the eigenvalue-one and scree tests criteria. Thus, the first component is a general one, with the second component seemingly relating to Conceptual Thinking and ability to perceive and organize material. In many ways, the two-component varimax solution is similar to the two-component quartimax solution. Both have a strong general component as well as a second component which is influenced most strongly by a very high loading from Bead Memory. However, the three-component varimax solution also deserves mention in that it appears to measure Fluid Intelligence, Verbal Ability, and Ability to Perceive and Organize Material.

Stanford-Binet IV Principal Components for the Age Group 9-0 to 12-11 Years

Table 35 reports the unrotated principal components for the Stanford-Binet IV for the age group of 9-0 through 12-11 years. From this table it can be noted that only one component in the full component model has an eigenvalue greater than or equal to one though a second component comes close (.97) to reaching unity. Thus, by the eigenvalue-one criteria only the first unrotated component should be considered. This contrasts with the WISC-R results where two components had eigenvalues of one or greater and a third component had an eigenvalue of .95.

In order to determine the number of components to rotate using the scree test criterion, the successive eigenvalues from the full component model for this age group were plotted in Figure 16. Examination of this figure indicates that there is more than one break in the scree line. From the scree plot, the most likely number of factors to rotate would be three--which is indicated by the break in the scree that occurs between the third and fourth components.

To be consistent with the findings of Thorndike et al. (1986) (*g* , Verbal, Memory, and Abstract / Visual factors), and with the organization of the Stanford-Binet into four Areas, a four component solution is also reported. For the sake of completeness, and because the second unrotated component almost reaches unity, a two-component solution is also reported. Hence, Table 36 reports two-, three-, and four-component solutions which have been rotated according to the quartimax criterion.

Table 35
Unrotated Principal Components For The Stanford-Binet IV: Ages 9-0 to 12-11 Years[a]

Stanford-Binet IV Age 9-0 to 12-11	Unrotated Principal Components											
Tests	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Vocabulary	.86	-.26	-.04	-.27	-.05	-.02	.01	.17	.13	-.20	.07	-.19
Comprehension	.81	-.22	-.18	-.21	.23	-.17	.31	-.04	-.04	.03	.13	.14
Absurdities	.88	.09	-.15	-.04	.09	-.24	.07	-.04	-.20	.13	-.21	-.14
Pattern Analysis	.78	.27	.03	-.05	-.33	.31	.06	-.10	-.28	-.06	.11	-.02
Copying	.68	.44	-.02	.39	-.09	-.38	-.02	.17	.01	-.08	.09	.03
Matrices	.83	.04	-.11	-.17	.17	-.04	-.43	-.07	.00	.14	.13	.01
Quantitative	.70	-.18	-.43	.34	-.07	.32	.06	.14	.11	.17	.00	-.01
Number Series	.86	-.11	-.24	.09	-.13	-.02	-.09	-.27	.12	-.22	-.14	.09
Bead Memory	.72	.48	.18	-.34	-.20	.06	.08	.07	.22	.12	-.09	.07
Memory for Sentences	.77	-.39	.34	.03	-.03	.05	-.12	.24	-.16	-.03	-.11	.13
Memory for Digits	.69	-.30	.51	.24	-.13	-.09	.08	-.20	.11	.13	.07	-.07
Memory for Objects	.65	.30	.23	.15	.57	.27	.05	-.02	.04	-.10	-.03	-.03
Eigenvalue	7.17	0.97	0.77	0.62	0.62	0.52	0.33	0.27	0.25	0.21	0.15	0.11
% of Total Variance	59.7	8.1	6.5	5.2	5.1	4.4	2.8	2.3	2.0	1.8	1.2	0.9
Cummulative % Var.	59.7	67.8	74.2	79.4	84.6	88.9	91.7	94.0	96.1	97.8	99.1	100

[a] The unrotated principal components listed in this table are based on the entire sample for this age group (n=66). Missing data were estimated through the regression procedure outlined previously.

Figure 16

Scree Plot for the Stanford-Binet IV at Ages 9-0 to 12-11 Years

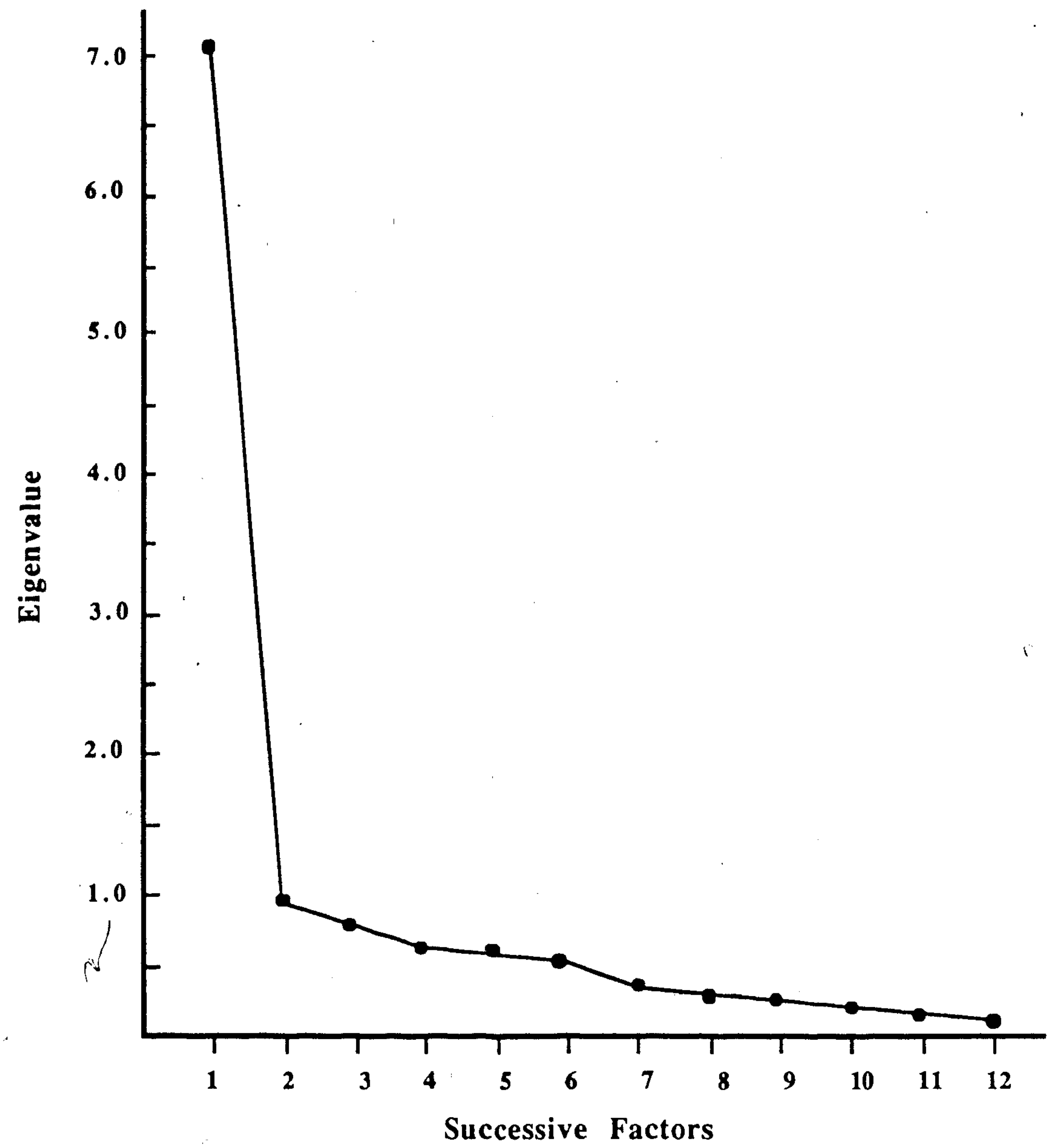


Table 36

Two-, Three-, and Four-Component Quartimax Rotated Principal Components For The Stanford-Binet IV: Ages 9-0 to 12-11 Years[a]

Stanford-Binet IV Age 9-0 to 12-11 Tests	Two Components Extracted I	II	h²	Three Components Extracted I	II	III	h²	Four Components Extracted I	II	III	IV	h²
Vocabulary	.88	-.18	.80	.85	.22	-.19	.80	.87	.13	-.31	-.02	.87
Comprehension	.82	-.15	.70	.82	.08	-.24	.73	.83	.01	-.28	.09	.78
Absurdities	.87	.17	.78	.89	-.06	.04	.81	.89	-.06	.07	.03	.81
Pattern Analysis	.75	.35	.68	.78	-.03	.28	.68	.78	-.03	.21	-.18	.69
Copying	.63	.50	.65	.69	-.18	.39	.65	.64	-.03	.62	.06	.80
Matrices	.83	.11	.70	.84	.00	.02	.71	.86	-.05	-.05	-.06	.74
Quantitative	.72	-.11	.53	.75	-.17	-.35	.71	.71	-.05	.08	.57	.83
Number Series	.86	-.03	.75	.88	-.02	-.18	.80	.86	.01	.00	.27	.81
Bead Memory	.67	.51	.71	.70	-.01	.51	.75	.73	-.10	.16	-.54	.87
Memory for Sentences	.81	-.32	.75	.71	.59	-.09	.87	.72	.58	-.14	.02	.87
Memory for Digits	.72	-.23	.57	.61	.67	.08	.83	.60	.72	.09	.01	.89
Memory for Objects	.62	.36	.51	.62	.10	.41	.56	.61	.16	.40	-.17	.58
Sum of Squared Loadings	7.12	1.03		7.06	0.93	0.92		7.02	0.92	0.83	0.77	
% of Total Variance	59.3	8.6		58.8	7.8	7.7		58.5	7.7	6.9	6.4	
Cummulative % Var.	59.3	67.9		58.8	66.6	74.3		58.5	66.2	73.1	79.5	

a The rotated principal components listed in this table are based on the entire sample for this age group (n=66). Missing data were estimated through the regression procedure outlined previously.

Before examining Table 36, the first unrotated component from Table 35 should be examined, as the eigenvalue-one criteria suggests that a single component solution is warranted. Examination of this unrotated component reveals that all twelve subtests have fairly strong loadings ranging from .65 to .86 on this component; thus, this can be considered a general component synonymous with *g*. It is interesting to note that the subtests with the lowest loadings on the general component are Memory for Objects, Memory for Digits and Copying. These are the same three subtests that Thorndike et al. (1986) found to have the lowest loadings on the general factor in their age grouping of 7 through 11 years.

The two-component quartimax solution can be viewed in Table 36. Once again, the first component is a general one, with loadings ranging from .62 to .88. The subtests with the lowest loadings on this first component are Memory for Objects, Copying, and Bead Memory; the subtests with the highest loadings on the first component are Vocabulary, Absurdities, Matrices, Number Series, and Memory for Sentences. The high loadings from Vocabulary, Number Series, and Memory for Sentences are consistent with Thorndike et al.'s (1986) findings; however the strong loading from Absurdities is unexpected and without precedence; the strong loading from Matrices is an encouraging finding, even though it contrasts with that of Thorndike et al.. The second component in the two-component solution is an Abstract / Visual component with strongest loadings from Copying, Bead Memory, Pattern Analysis, and Memory for Objects.

The three-component solution also retains a strong general component; once again, the subtests with the highest loadings on this component are Vocabulary, Comprehension, Absurdities, Matrices, and Number Series. In contrast, the four subtests from the Short-Term Memory Area, along with Copying, have the lowest loadings. The second component in the three-component model is Auditory Short-Term Memory, with moderate loadings from Memory for Sentences and Memory for Digits. The third component is

Visual / Conceptual, with strongest loadings from Bead Memory, Memory for Objects, Copying, and a moderate negative loading from Quantitative.

In the four-component solution the first component is clearly a general component, just as in the two-, and three-component solutions. Also, the second component in the four-component solution is Auditory Short-Term Memory, with loadings from Memory for Sentences and from Memory for Digits. The difference between the three-, and four-component solutions is that the third component from the three-component solution has been divided into two components in the four-component solution; the third component being Visual Perception / Gestalt Perception with highest loadings from Copying and Memory for Objects; the fourth component is bipolar with one pole being Rote or Passive Memory (as indicated by Quantitative) and the other being Conceptual or Working Memory (as indicated by Bead Memory).

Taken as a whole the quartimax rotated two-, three-, and four-component solutions indicate that the general component is strong and remains intact accounting for 58.5 to 59.3 % of the total subtest variance, depending on how many components are rotated. Of noteworthiness is the fact that the subtests from the Verbal Reasoning Area, along with the Matrices and Number Series subtests tend to have the highest loadings on the general component. Of the various number-of-component solutions, the three-component quartimax rotated solution seems to be the best. Along with the general factor, it has an Auditory Short-Term Memory, as well as a Visual / Conceptual component. Unlike Thorndike et al.'s (1986) findings in their 7 through 11 age group, however, no Verbal factor was found—perhaps because of the strong loadings of the Verbal Reasoning Area subtests on the general component.

Table 37

Two-, Three-, and Four-Component Varimax Rotated Principal Components For The Stanford-Binet IV: Ages 9-0 to 12-11 Years[a]

Stanford-Binet IV Age 9-0 to 12-11	Varimax Rotated Principal Components											
	Two Components Extracted			Three Components Extracted				Four Components Extracted				
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Vocabulary	.81	.37	.80	.67	.30	.51	.80	.78	.21	.40	.24	.87
Comprehension	.76	.36	.70	.73	.27	.37	.73	.74	.18	.27	.34	.78
Absurdities	.61	.64	.78	.65	.56	.26	.81	.59	.50	.21	.40	.81
Pattern Analysis	.41	.72	.68	.40	.68	.24	.68	.45	.64	.19	.21	.69
Copying	.22	.78	.65	.32	.74	.07	.65	.05	.76	.15	.46	.80
Matrices	.61	.57	.70	.61	.50	.30	.71	.66	.43	.21	.28	.74
Quantitative	.65	.33	.53	.81	.19	.12	.71	.35	.17	.18	.80	.83
Number Series	.72	.47	.75	.76	.37	.29	.80	.55	.32	.28	.57	.81
Bead Memory	.25	.81	.71	.20	.81	.23	.75	.53	.74	.10	-.13	.87
Memory for Sentences	.84	.21	.75	.40	.22	.81	.87	.46	.18	.77	.18	.87
Memory for Digits	.72	.23	.57	.21	.28	.84	.83	.19	.28	.87	.16	.89
Memory for Objects	.30	.65	.51	.17	.66	.31	.56	.16	.65	.32	.16	.58
Sum of Squared Loadings	4.51	3.63		3.55	3.13	2.23		3.12	2.71	1.95	1.73	
% of Total Variance	37.6	30.3		29.6	26.1	18.6		26.0	22.6	16.3	14.4	
Cummulative % Var.	37.6	67.9		32.2	58.3	76.9		26.0	48.6	64.9	79.3	

[a] The rotated principal components listed in this table are based on the entire sample for this age group (n=66). Missing data were estimated through the regression procedure outlined previously.

The results of the varimax rotated two- three-, and four-component solutions are presented in Table 37. Examination of the two-component varimax solution is interesting in that these two factors approximate Crystallized Intelligence and Fluid Intelligence.

The first component, accounting for 37.6% of the total subtest variance is Crystallized Intelligence; alternately, it could be conceived of as a Verbal-Educational component (Vernon, 1961). The Vocabulary, Comprehension, Absurdities, Memory for Sentences, Memory for Digits, Quantitative, Number Series, Matrices, and Pattern Analysis subtests load on this component. Vocabulary, Comprehension, Absurdities, and Memory for Sentences are clearly verbal in nature and thus would be expected to load on Crystallized Intelligence. Memory for Digits, Quantitative, and Number Series all involve the use of numbers, and consequently their loading on Crystallized Intelligence is logical.

The loadings from Matrices, Pattern Analysis, and Absurdities on the Crystallized Intelligence component requires some explanation as these subtests are complex and load on both components. First, it should be noted that the loading by Pattern Analysis is fairly small and secondary to its loading on the second component; nonetheless, familiarity with blocks and shapes probably has a learned basis, and hence the small loading on Crystallized intelligence. Second, it should be noted that Matrices has nearly identical loadings on both the first and second components. To those familiar with Raven's Progressive Matrices, this finding may seem confusing; however, closer examination of the Matrices subtest on the Stanford-Binet IV reveals that the first few questions on the Matrices subtest involve the use of pictures of common objects like birds and mice rather than abstract symbols that are used on the Raven's Matrices; therefore, the Matrices subtest is not purely a measure of Fluid Intelligence, but is also influence by formation of categories and objects that are learned; hence, the moderate loading on Crystallized Intelligence. Third, Absurdities also has nearly identical loadings on both components suggesting that it may have not only a verbal element and hence the loading on Crystallized intelligence, but also a non-verbal / spatial element and hence the loading on Fluid Intelligence.

The second component, accounting for 30.3 % of the total subtest variance, is Fluid Intelligence; alternately, it could be conceived of as a Spatial-Mechanical component (Vernon, 1961). The Pattern Analysis, Copying, Bead Memory, Memory for Objects, Absurdities, Matrices, and Number Series subtests load on this component. Pattern Analysis, Copying, Bead Memory, and Memory for Objects are all Spatial in nature, require ability to conceptualize, to perceive the gestalt, and to reason nonverbally--and hence their loadings on Fluid Intelligence are expected. The three remaining subtests (Absurdities, Matrices, and Number Series) are complex, loading on both the first and second components. It is not that surprising that Absurdities loads on both components--its pictorial nature requires nonverbal search strategies, while the verbal element occurs because answers are given verbally. Although Matrices loads on Fluid Intelligence as expected, it also loads on Crystallized Intelligence--indicating that for this age group, an educational / learning element is also involved. Finally, the moderate loading of Number Series on Fluid Intelligence is predictable; Number Series requires not only facility with numbers, but also the ability to identify patterns--which involves nonverbal, Fluid Intelligence.

In the three-component varimax solution the first two components are still Crystallized and Fluid Intelligence and the third component emerges as Auditory Short-Term Memory. The subtests with loadings on Crystallized Intelligence are Vocabulary, Comprehension, Absurdities, Matrices, Quantitative, Number Series, Pattern Analysis, and Memory for Sentences. The Fluid Intelligence component has loadings from Pattern Analysis, Copying, Bead Memory, Memory for Objects, Absurdities, and Matrices. The Auditory Short-Term Memory Component has loadings from Memory for Sentences, Memory for Digits, and Vocabulary.

In the four-component varimax solution, the first component is again Crystallized Intelligence with loadings from Vocabulary, Comprehension, Absurdities, Memory for Sentences, Pattern Analysis, Matrices, Number Series, and Bead Memory. The second component is more restricted than in previous solutions, and is close to Thorndike et al.'s (1986) Abstract / Visual Factor. Subtests with loadings on the second component include Absurdities, Pattern Analysis, Copying, Bead Memory, and Memory for Objects. The third component remains unchanged as Auditory Short-Term Memory; Memory for Sentences, Memory for Digits, and Vocabulary load on this component. The fourth component that emerges is definitely Quantitative in nature with loadings from the Quantitative, Number Series, and Copying subtests.

As a whole, the varimax rotated two-, three-, and four-component solutions for the age group of 9-0 to 12-11 years are encouraging. In the two-component solution the two components that emerge are Fluid and Crystallized Intelligence. In the three-component solution the first two factors are Fluid and Crystallized Intelligence, and the third component is Auditory Short-Term Memory. Finally, in a four-component solution Crystallized Intelligence, Abstract / Visual Reasoning, Auditory Short-Term Memory, and Quantitative Reasoning emerge. To some extent, the findings from this age group do provide support for the construct validity of the Stanford-Binet IV, especially for the theoretical underpinnings of Fluid and Crystallized Intelligence. Similarly, when a four-component solution is examined the division of the Stanford-Binet IV into four Reasoning Areas receives some support; however the subtests that load on each of the four Areas is somewhat different than outlined on the Stanford-Binet IV.

Stanford-Binet IV Principal Components for the Age Group 13-0 to 16-11 Years

Because of the atypical nature of the sample for the age group 13-0 through 16-11 years, the current findings of the component structure of the Stanford-Binet IV for this age group must be examined with extreme caution. As was seen in the component structure of the WISC-R for this age group, an artifactually high general factor exists because of the inflated correlations resulting from an extremely heterogeneous sample.

Table 38
Unrotated Principal Components For The Stanford-Binet IV: Ages 13-0 to 16-11 Years[a]

Stanford-Binet IV Age 13-0 to 16-11	Unrotated Principal Components									
Tests	I	II	III	IV	V	VI	VII	VIII	IX	X
Vocabulary	.89	-.07	-.24	.25	.08	.02	.27	.04	-.02	.01
Comprehension	.93	-.17	-.18	.06	-.06	.01	-.13	.08	.17	-.13
Pattern Analysis	.90	-.10	.18	.02	.18	-.33	-.01	-.10	.01	-.02
Matrices	.92	-.13	.10	-.06	-.22	.12	.07	-.25	-.01	-.06
Quantitative	.93	-.14	-.08	-.10	-.19	-.08	-.02	.04	.05	.21
Number Series	.94	-.17	.01	-.04	-.08	-.02	-.08	.13	-.22	-.06
Bead Memory	.86	-.14	.37	.09	.19	.22	-.03	.06	.04	.06
Memory for Sentences	.87	.21	-.30	-.03	.21	.10	-.15	-.13	-.05	.05
Memory for Digits	.85	.28	.03	-.41	.06	.01	.14	.08	.04	-.05
Memory for Objects	.87	.21	-.30	-.03	.21	.10	-.15	-.13	.01	-.01
Eigenvalue	7.88	0.54	0.39	0.32	0.25	0.19	0.14	0.13	0.09	0.08
% of Total Variance	78.8	5.4	3.9	3.2	2.5	1.9	1.4	1.3	0.9	0.8
Cummulative % Var.	78.8	84.2	88.1	91.2	93.7	95.6	97.1	98.4	99.2	100.0

[a] The unrotated principal components listed in this table are based on the entire sample for this age group (n=56). Missing data were estimated through the regression procedure outlined previously.

Table 38 reports the unrotated principal components for the Stanford-Binet IV for the oldest age group. This table confirms the presence of an artifactually high general component. The first unrotated component accounts for 78.8% of the total subtest variance, just as the first unrotated component on the WISC-R had accounted for 70.9% of the total subtest variance. As a consequence, only the first component in the full component model has an eigenvalue of one or greater. Thus, by the eigenvalue-one criterion no components should be rotated and the first unrotated component should be reported as a general component. Indeed the loadings on this first component range from .85 to .94.

In order to determine the number of components suggested by the scree test, the successive eigenvalues for the components for this age group have been plotted in Figure 17. Examination of Figure 17 reveals that a one-component solution is probably best, though a two-component solution might be entertained. Despite this, two-, three-, and four-component solutions were obtained in order to be consistent with solutions obtained on the two younger age groups. The quartimax rotated two-, three-, and four-component solutions are reported in Table 39.

Examination of Table 39 reveals that even in the two-, three-, and four-component quartimax rotated solutions the general component dwarfs all other components. So much so, that all of the other components are singlets and therefore not legitimately interpretable.

Table 40 reports the varimax rotated two-, three-, and four-component solutions. These solutions provide little help in understanding the structure of the Stanford-Binet IV for this age group. In all of these solutions, the factor structure is complex with nearly all subtests having at least moderate loadings on several components. Thus, no attempt has been made at interpreting the varimax rotations for this age group--they have been provided for the sake of completeness. The decision not to interpret these various solutions is supported by not only the eigenvalue-one but also the scree test criteria; both these criteria identify the single component solution as most appropriate. Therefore, anyone examining and attempting to identify or name the components in Table 40 should do so with extreme caution.

Figure 17
Scree Plot for the Stanford-Binet IV at Ages 13-0 to 16-11 Years

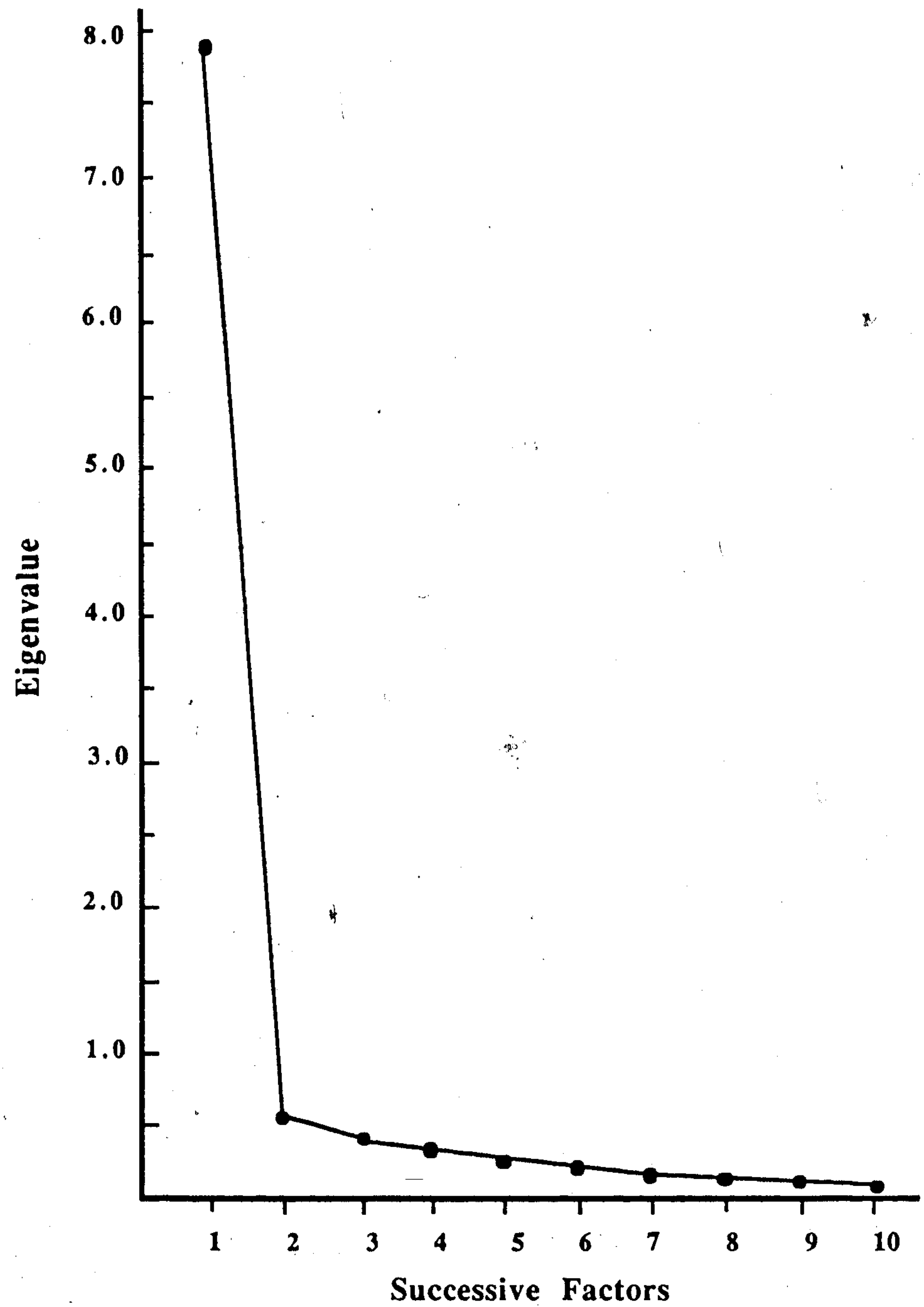


Table 39

Two-, Three-, and Four-Component Quartimax Rotated Principal Components For The Stanford-Binet IV: Ages 13-0 to 16-11 Years[a]

Stanford-Binet IV Age 13-0 to 16-11	Quartimax Rotated Principal Components												
	Two Components Extracted			Three Components Extracted				Four Components Extracted					
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Vocabulary	.89	-.03	.80	.90	-.09	-.21	.86	.90	.05	.20	-.27	.92
Comprehension	.94	-.11	.89	.94	-.16	-.13	.92	.94	-.10	.13	-.12	.93
Pattern Analysis	.90	-.05	.81	.90	.01	.20	.85	.90	.03	-.20	.00	.85
Matrices	.92	-.08	.85	.92	-.05	.13	.86	.92	-.05	-.12	.04	.87
Quantitative	.94	-.08	.89	.94	-.10	-.04	.90	.94	-.12	.05	.05	.91
Number Series	.95	-.12	.91	.95	-.11	.06	.91	.95	-.10	-.05	.00	.92
Bead Memory	.87	-.09	.76	.86	.02	.40	.90	.86	.07	-.40	-.04	.91
Memory for Sentences	.86	.26	.81	.87	.17	-.35	.90	.86	.16	.35	.09	.90
Memory for Digits	.84	.32	.80	.84	.32	-.04	.80	.83	.12	.07	.51	.97
Memory for Objects	.74	.58	.89	.74	.60	-.01	.91	.73	.66	.01	.05	.97
Sum of Squared Loadings	7.87	0.56		7.89	0.55	0.41		7.84	0.52	0.40	0.36	
% of Total Variance	78.7	5.6		78.9	5.5	4.1		78.4	5.2	4.0	3.6	
Cummulative % Var.	78.7	84.3		78.9	84.4	88.5		78.4	83.6	87.6	91.2	

[a] The rotated principal components listed in this table are based on the entire sample for this age group (n=56). Missing data were estimated through the regression procedure outlined previously.

Table 40

Two, Three, and Four-Component Varimax Rotated Principal Components For The Stanford-Binet IV: Ages 13-0 to 16-11 Years[a]

Stanford-Binet IV Age 13-0 to 16-11	Varimax Rotated Principal Components												
	Two Components Extracted			Three Components Extracted				Four Components Extracted					
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Vocabulary	.77	.45	.80	.76	.42	.33	.86	.42	.78	.15	.33	.92
Comprehension	.86	.40	.89	.76	.52	.28	.92	.52	.73	.29	.20	.93
Pattern Analysis	.79	.44	.81	.45	.71	.37	.85	.70	.42	.30	.30	.85
Matrices	.82	.42	.85	.53	.68	.34	.86	.67	.48	.37	.23	.87
Quantitative	.84	.43	.89	.68	.58	.32	.90	.57	.60	.43	.17	.91
Number Series	.87	.40	.91	.62	.66	.30	.91	.65	.57	.36	.19	.92
Bead Memory	.78	.39	.76	.29	.83	.35	.90	.83	.28	.20	.32	.91
Memory for Sentences	.59	.68	.81	.73	.24	.56	.90	.23	.67	.48	.41	.90
Memory for Digits	.54	.72	.80	.45	.42	.65	.80	.39	.30	.78	.35	.97
Memory for Objects	.32	.89	.89	.26	.32	.86	.91	.31	.28	.27	.85	.97
Sum of Squared Loadings	5.44	3.00		3.37	3.21	2.23		3.12	2.93	1.60	1.47	
% of Total Variance	54.4	30.0		33.7	32.1	22.3		31.2	29.3	16.0	14.7	
Cummulative % Var.	54.4	84.4		33.7	65.8	88.1		31.2	60.5	76.5	91.2	

[a] The rotated principal components listed in this table are based on the entire sample for this age group (n=56). Missing data were estimated through the regression procedure outlined previously.

F. Principal Components From the Joint Analysis of the Stanford-Binet IV and the WISC-R

The joint analysis of the Stanford-Binet IV and the WISC-R is an important step in helping to provide further understanding of the component structure of the Stanford-Binet IV. Besides providing information as to whether the component structure of the Stanford-Binet IV is invariant when additional subtests are added, the joint analysis allows a direct comparison of which subtests from the Stanford-Binet IV and WISC-R load on the same components. Furthermore, the joint analysis has the potential to allow more components to be defined than when either the WISC-R or Stanford-Binet IV are factored separately; that is, by increasing the domain of subtests factored, it is possible for new components to emerge as the additional subtests load on what may have previously been a small and insignificant minor factor. The point to remember is that the number and composition of the subtests that are factored does indeed have an impact on the component structure that emerges.

In this section, the results from the principal components analyses of joint factoring of the Stanford-Binet IV and the WISC-R are reported. For organizational purposes, the three age groups are dealt with consecutively. That is, the results from the age group 6-0 to 8-11 years are reported first; these are followed by the results from the age group 9-0 to 12-11 years; finally, the results for the age group 13-0 to 16-11 are reported.

To allow for comparison between the groups, Table 41 has been constructed. In this table the eigenvalues for each component in the full component model[46] are reported for each of the three age groups; the percentage of the total subtest variance that each component accounts for has also been provided. Similarly, the first unrotated component for each of the three age groups has been provided in Table 42. In examining these two tables, the reader should remember that previous analyses have indicated that the 13-0 to

[46] These eigenvalues were obtained from the unrotated components in the full component model; i.e., when as many components as subtests are extracted.

16-11 year-old age group is extremely heterogeneous and hence an artificially strong first component exists with inflated factor loadings on that component.

Table 41
The Eigenvalues and Percent of Total Variance Accounted for by All Principal Components From the Joint Analysis of the Stanford-Binet IV and WISC-R for Each of the Three Age Groups

Component	Age 6-0 to 8-11		Age 9-0 to 12-11		Age 13-0 to 16-11	
	Eigenvalue	% Total Variance	Eigenvalue	% Total Variance	Eigenvalue	% Total Variance
I	9.23	46.2	13.10	54.6	16.15	73.4
II	1.48	7.4	1.76	7.3	1.06	4.8
III	1.45	7.3	1.22	5.1	0.76	3.5
IV	1.29	6.5	1.16	4.8	0.61	2.8
V	1.04	5.2	0.99	4.1	0.48	2.2
VI	0.80	4.0	0.76	3.2	0.40	1.8
VII	0.76	3.8	0.69	2.9	0.39	1.8
VIII	0.66	3.3	0.55	2.3	0.32	1.4
IX	0.53	2.7	0.50	2.1	0.26	1.2
X	0.46	2.3	0.43	1.8	0.25	1.1
XI	0.42	2.1	0.41	1.7	0.23	1.0
XII	0.39	1.9	0.36	1.5	0.22	1.0
XIII	0.32	1.6	0.33	1.4	0.16	0.7
XIV	0.28	1.4	0.30	1.2	0.14	0.6
XV	0.24	1.2	0.26	1.1	0.12	0.6
XVI	0.23	1.1	0.24	1.0	0.11	0.5
XVII	0.17	0.8	0.20	0.8	0.08	0.4
XVIII	0.12	0.6	0.17	0.7	0.08	0.3
XIX	0.08	0.4	0.13	0.6	0.07	0.3
XX	0.06	0.3	0.13	0.5	0.05	0.2
XXI	N/A	N/A	0.11	0.4	0.04	0.2
XXII	N/A	N/A	0.09	0.4	0.03	0.1
XXIII	N/A	N/A	0.08	0.3	N/A	N/A
XXIV	N/A	N/A	0.03	0.1	N/A	N/A

* N/A means that the test was not included in the analysis for the specified age level
* Missing data was estimated through the regression procedure outlined previously.

Table 42
The Unrotated First Principal Component From the Joint Analysis of the Stanford-Binet IV and WISC-R for Each of the Three Age Groups

	Tests	Unrotated First Principal Component		
		Age 6-0 to 8-11	Age 9-0 to 12-11	Age 13-0 to 16-11
WISC-R Tests	Information	.76	.86	.90
	Similarities	.78	.82	.93
	Arithmetic	.78	.78	.88
	Vocabulary	.80	.85	.89
	Comprehension	.69	.87	.88
	Digit Span	.65	.70	.86
	Picture Completion	.40	.69	.74
	Picture Arrangement	.79	.40	.78
	Block Design	.66	.82	.87
	Object Assembly	.62	.69	.81
	Coding	.65	.45	.76
	Mazes	.52	.44	.73
Stanford-Binet IV Tests	Vocabulary	.82	.87	.88
	Comprehension	.74	.81	.93
	Absurdities	.73	.85	N/A
	Pattern Analysis	.64	.76	.89
	Copying	.64	.66	N/A
	Matrices	N/A	.85	.91
	Quantitative	.61	.70	.92
	Number Series	N/A	.85	.93
	Bead Memory	.43	.69	.85
	Memory for Sentences	.67	.75	.86
	Memory for Digits	N/A	.68	.84
	Memory for Objects	N/A	.62	.74

* N/A means that the test was not included in the analysis for the specified age range.
* The regression procedure outlined earlier was used to estimate missing values

Principal Components From the Joint Analysis of the Stanford-Binet IV and the WISC-R for the Age Group of 6-0 to 8-11 Years

From Table 41, the number of components with eigenvalues greater or equal to one can be obtained. Thus, using the eigenvalue-one criterion, five components should be rotated for the youngest age group. This is interesting in that when the Stanford-Binet IV was factored by itself only two components had eigenvalues greater than one, and when the WISC-R was factored by itself only three components had eigenvalues greater than one. Thus, when the two tests are jointly factored more components emerge than when the tests are factored separately. In order to determine the number of components suggested by the scree test, the successive components for this age group have been plotted in Figure 18. Examination of this figure reveals that a five-component solution is also indicated by the scree test criterion.

To allow for comparison with the two-, and three-component solutions of the Stanford-Binet IV and the WISC-R when factored separately, these solutions are also provided for the joint analysis. Thus, two-, three-, four-, and five-component solutions have been presented for the joint analysis of the Stanford-Binet IV and WISC-R for this age group. These solutions are reported in Table 43, Table 44, and Table 45. Because the five-component solution is supported by both the eigenvalue-one and scree test criteria it is examined first.

Figure 18
Scree Plot for the Joint Analysis of
the Stanford-Binet IV and the WISC-R for Ages 6-0 to 8-11

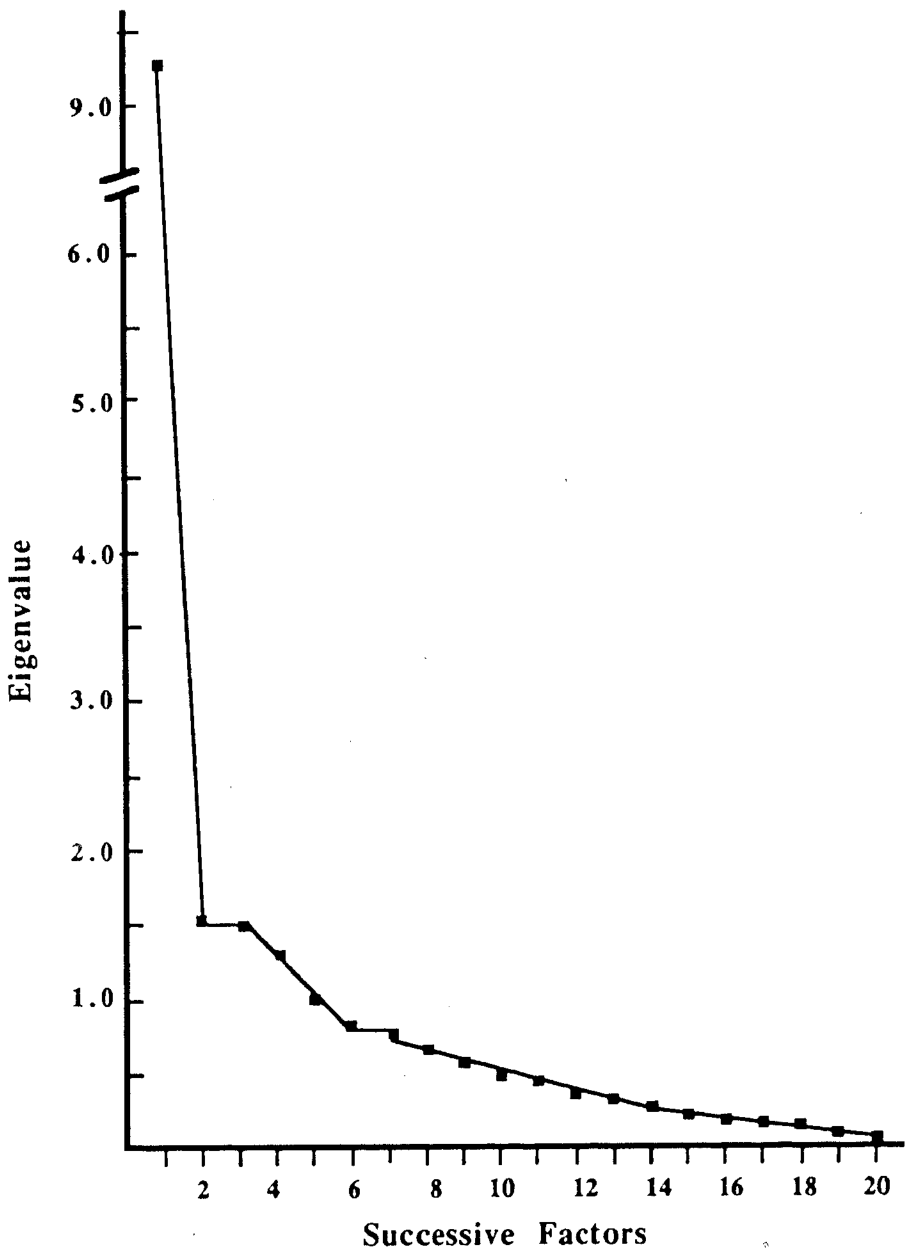


Table 43

Varimax Rotated Two- and Three-Component Solutions for the Joint Analysis of the Stanford-Binet IV and WISC-R: Ages 6-0 to 8-11 Years[a]

Age 6-0 to 8-11		Varimax Rotated Principal Components						
		Two Components Extracted			Three Components Extracted			
	Tests	I	II	h^2	I	II	III	h^2
WISC-R Tests	Information	.70	.33	.59	.77	.28	.06	.68
	Similarities	.82	.20	.71	.82	.10	.21	.73
	Arithmetic	.73	.33	.64	.70	.21	.33	.64
	Vocabulary	.76	.30	.68	.76	.20	.25	.68
	Comprehension	.77	.11	.61	.64	-.07	.50	.67
	Digit Span	.72	.12	.53	.56	-.07	.55	.62
	Picture Completion	.19	.42	.21	.00	-.23	.70	.54
	Picture Arrangement	.67	.42	.63	.58	.25	.52	.67
	Block Design	.42	.56	.48	.43	.48	.27	.48
	Object Assembly	.47	.42	.39	.33	.25	.59	.51
	Coding	.50	.41	.42	.44	.29	.39	.44
	Mazes	.02	.87	.76	.12	.87	.12	.78
Stanford-Binet IV Tests	Vocabulary	.74	.37	.69	.80	.30	.14	.75
	Comprehension	.55	.50	.56	.63	.46	.10	.62
	Absurdities	.46	.62	.59	.49	.54	.27	.60
	Pattern Analysis	.17	.85	.76	.22	.80	.27	.76
	Copying	.52	.38	.41	.52	.29	.24	.41
	Quantitative	.47	.40	.38	.59	.40	-.06	.51
	Bead Memory	.31	.31	.19	.08	.11	.74	.56
	Memory for Sentences	.66	.23	.48	.69	.16	.11	.52
Sum of Squared Loadings		6.63	4.09		6.32	2.95	2.92	
% of Total Variance		33.2	20.5		31.6	14.8	14.6	
Cummulative % Var.		33.2	53.7		31.6	46.4	61.0	

a The rotated principal components listed in this table are based on the entire sample for this age group (n=46). Missing data were estimated through the regression procedure outlined previously.

Table 44

Varimax Rotated Four-Component Solution for the Joint Analysis of the Stanford-Binet IV and WISC-R: Ages 6-0 to 8-11 Years[a]

Age 6-0 to 8-11		Varimax Rotated Principal Components				
	Tests	I	II	III	IV	h^2
WISC-R Tests	Information	.70	.34	.26	-.09	.68
	Similarities	.79	.31	.08	.10	.74
	Arithmetic	.57	.54	.17	.10	.66
	Vocabulary	.77	.26	.18	.17	.72
	Comprehension	.68	.25	-.08	.45	.74
	Digit Span	.34	.78	-.11	.20	.78
	Picture Completion	.04	.23	.23	.68	.56
	Picture Arrangement	.52	.45	.23	.37	.67
	Block Design	.29	.50	.45	.06	.54
	Object Assembly	.09	.81	.20	.23	.76
	Coding	.30	.55	.26	.16	.49
	Mazes	.15	.08	.86	.11	.79
Stanford-Binet IV Tests	Vocabulary	.79	.25	.28	.06	.77
	Comprehension	.76	-.06	.46	.19	.83
	Absurdities	.43	.36	.52	.13	.60
	Pattern Analysis	.22	.21	.79	.21	.76
	Copying	.42	.42	.27	.07	.43
	Quantitative	.38	.54	.36	-.35	.69
	Bead Memory	.21	.09	.11	.80	.71
	Memory for Sentences	.70	.17	.15	.06	.55
Sum of Squared Loadings		5.38	3.46	2.73	1.86	
% of Total Variance		26.9	17.3	13.7	9.3	
Cummulative % Var.		26.9	44.2	57.9	67.2	

[a] The rotated principal components listed in this table are based on the entire sample for this age group (n=46). Missing data were estimated through the regression procedure outlined previously.

Table 45

Varimax Rotated Five-Component Solution for the Joint Analysis of the Stanford-Binet IV and WISC-R: Ages 6-0 to 8-11 Years[a]

Age 6-0 to 8-11	Tests	Varimax Rotated Principal Components					
		I	II	III	IV	V	h^2
WISC-R Tests	Information	.65	.39	.22	.20	-.08	.68
	Similarities	.77	.29	.06	.24	.10	.75
	Arithmetic	.47	.64	.12	.22	.13	.71
	Vocabulary	.72	.37	.15	.11	.18	.72
	Comprehension	.64	.34	-.11	.07	.47	.76
	Digit Span	.30	.46	-.16	.65	.22	.80
	Picture Completion	-.01	.27	.21	.07	.69	.60
	Picture Arrangement	.49	.37	.20	.33	.38	.67
	Block Design	.36	-.01	.44	.74	.03	.87
	Object Assembly	.07	.40	.16	.74	-.24	.80
	Coding	.18	.72	.20	.14	.20	.65
	Mazes	.14	.13	.86	.09	.10	.79
Stanford-Binet IV Tests	Vocabulary	.75	.35	.26	.12	.06	.77
	Comprehension	.73	.19	.46	-.13	.18	.83
	Absurdities	.37	.43	.48	.19	.14	.61
	Pattern Analysis	.19	.25	.77	.15	.21	.76
	Copying	.32	.59	.21	.09	.10	.52
	Quantitative	.30	.56	.31	.31	-.33	.70
	Bead Memory	.24	-.07	.13	.18	.79	.74
	Memory for Sentences	.77	-.09	.16	.39	.03	.77
Sum of Squared Loadings		4.79	3.10	2.49	2.20	1.91	
% of Total Variance		24.0	15.5	12.5	11.0	9.6	
Cummulative % Var.		24.0	39.5	52.0	63.0	72.6	

[a] The rotated principal components listed in this table are based on the entire sample for this age group (n=46). Missing data were estimated through the regression procedure outlined previously.

In the five-component solution (Table 45) the first component is Verbal, with loadings from Information, Similarities, W.Vocabulary[47], W.Comprehension, Picture Arrangement, B.Vocabulary, B.Comprehension, and Memory For Sentences. The second component is Numerical / Concentration, and has loadings from Arithmetic, Digit Span, Coding, Copying, and Quantitative.[48] The third component is Visual Planning / Ability to See Relationships, and has strong loadings from Mazes and Pattern Analysis, and moderate loadings from Block Design, Absurdities, and B.Comprehension. The fourth component is Perceptual Organization as typically found on the WISC-R; Block Design, Object Assembly, and Digit Span have high loadings on this component. The fifth component is Working Memory / Conceptual Categorization; Picture Completion and Bead Memory have high loadings on this compenent, while W.Comprehension has a moderate loading.

In the five-component solution, 72.6% of the total subtest variance is accounted for by the five components; the Verbal component is slightly larger than the other components, accounting for 24.0% of the total subtest variance; the remaining components are similar in size, and range from 9.6-15.5% of the total subtest variance.

Examination of the three-, and four-component solutions also indicates that the first component is fairly large and that the remaining components are quite similar in size. However, in the two-component solution both components are reasonably large, together accounting for 53.7% of the total subtest variance.

When the two-component solution is inspected, the first component appears to be Verbal-Educational / Crystallized Intelligence while the second is Spatial-Mechanical / Fluid Intelligence. In the three-component solution the first two components are very similar to those in the two-component solution; the third component involves both Sequencing and

---

47 Because some subtests on the Stanford-Binet IV and the WISC-R have the same names, a capitalized letter "W" precedes the WISC-R subtests while a "B" precedes that Stanford-Binet IV subtests; for example, W.Vocabulary means that this is the Vocabulary subtest from the WISC-R.

48 The presence of loadings from Arithmetic, Digit Span, and Coding suggest that this component could also be conceived of as "Freedom from Distractibility".

Gestalt Perception, and has high loadings from Bead Memory, Picture Completion, Comprehension, Digit Span, Picture Arrangement, and Object Assembly.

In the four-component solution the nature of the first two components has changed as compared to the two-, and three-components solutions. The first component is Verbal, with high loadings from Information, Similarities, Arithmetic, W.Vocabulary, W.Comprehension, B.Vocabulary, B.Comprehension, and Memory For Sentences. The second component involves Perception of Relationships & Numerical Ability, and has high loadings from Arithmetic, Digit Span, Block Design, Object Assembly, Coding, and Quantitative. The third component is Nonverbal Reasoning & Planning, and has loadings from Block Design, Mazes, B.Comprehension, Absurdities, Pattern Analysis, and Quantitative. The fourth component is Working Memory / Conceptual Categorization, and has high loadings from Bead Memory, Picture Completion, and W.Comprehension.

To this point, only the varimax rotated components for the joint analysis of the Stanford-Binet IV and WISC-R for the youngest age group have been reported. In Table 46, the five-component quartimax rotated solution for this age group is presented. Examination of the first component reveals that it can be considered a general component, similar to *g* ; however, it is worth noting that this component has highest loadings from subtests with a verbal content, while other subtests have moderate loadings. Three subtests have weak loadings on this general component: Picture Completion, Mazes, and Bead Memory. This first component accounts for 43.9% of the total subtest variance, and dwarfs all other components. Each of the remaining four components accounts for a small amount of the total subtest variance, ranging from 5.5-8.5%.

Table 46
Quartimax Rotated Five-Component Solution for the Joint Analysis of the Stanford-Binet IV and WISC-R: Ages 6-0 to 8-11 Years[a]

Age 6-0 to 8-11		Quartimax Rotated Principal Components					
	Tests	I	II	III	IV	V	h^2
WISC-R Tests	Information	.80	.04	-.18	-.02	.10	.68
	Similarities	.82	-.14	.01	-.02	.22	.75
	Arithmetic	.81	-.04	-.01	.13	-.20	.71
	Vocabulary	.83	-.05	.08	-.11	.08	.72
	Comprehension	.73	-.28	.37	-.09	.00	.76
	Digit Span	.62	-.26	.11	.58	-.02	.80
	Picture Completion	.30	.18	.63	.11	-.26	.60
	Picture Arrangement	.75	.05	.28	.18	.03	.67
	Block Design	.53	.36	.00	.49	.46	.87
	Object Assembly	.51	.12	.15	.71	-.03	.80
	Coding	.64	.09	.06	.18	-.44	.65
	Mazes	.40	.79	.06	-.03	.02	.79
Stanford-Binet IV Tests	Vocabulary	.86	.05	-.04	-.13	.13	.77
	Comprehension	.76	.26	.11	-.39	.16	.83
	Absurdities	.69	.35	.04	.07	-.09	.61
	Pattern Analysis	.52	.68	.15	.05	-.04	.76
	Copying	.66	.09	-.02	.05	-.28	.52
	Quantitative	.63	.20	-.44	.23	-.13	.70
	Bead Memory	.32	.06	.78	.07	.15	.74
	Memory for Sentences	.66	-.01	.01	.03	.58	.77
Sum of Squared Loadings		8.77	1.69	1.53	1.43	1.09	
% of Total Variance		43.9	8.5	7.7	7.2	5.5	
Cummulative % Var.		43.9	52.4	60.1	67.3	72.8	

a The rotated principal components listed in this table are based on the entire sample for this age group (n=56). Missing data were estimated through the regression procedure outlined previously.

Comparison of the quartimax and varimax five-component solutions reveals that there is much similarity between the second, third, fourth, and fifth components, though they emerge in different orders and vary somewhat in composition. In the quartimax solution, the second component is Visual Planning / Ability to See Relationships, and has its highest loadings from Pattern Analysis, Mazes, Block Design, and Absurdities. The third component is Working Memory / Conceptual Categorization, and has its highest loadings from Bead Memory, Picture Completion, and W.Comprehension; it is interesting to note that of these three subtests, both Bead Memory and Picture Completion have relatively low loadings on the General component. The fourth component has high loadings from Object Assembly, Digit Span, and Block Design; this appears to be some type of Perceptual Organization component. The fifth component is difficult to interpret and is bipolar, with one pole being Semantic Memory and the other being Non-Meaningful Visual Memory.

Principal Components From the Joint Analysis of the Stanford-Binet IV and the WISC-R for the Age Group 9-0 to 12-11 Years

Examination of Table 41 reveals that for the age group of 9-0 to 12-11 years, four components have eigenvalues of one or greater. However, a fifth component has an eigenvalue of .99. This is interesting in that the younger group also had five components with eigenvalues of one or greater. Thus, by a strict interpretation of the eigenvalue-one criterion only four components would be rotated for this age group, but a more lenient approach might also consider a five-component solution given that .99 ≈ 1.0 .

To further collaborate the number of factors to rotate, the successive eigenvalues for this age group were plotted in Figure 19 so that the scree test could be used. In Figure 19 several breaks in the scree are evident. The most feasible break occurs between the fifth and sixth components, suggesting that the five-component solution is most appropriate. As a result, taking the scree test and eigenvalue-one criteria together, the five-component solution is the optimum choice.

The two-, and three-component solutions are presented in Table 47, the four-component solution is given in Table 48; and the five-component solution is reported in Table 49. Because the five-component solution is indicated by both the eigenvalue-one and scree test criteria, it is given the most emphasis, is discussed first, and is treated in more depth than the other solutions.

Figure 19
Scree Plot for the Joint Analysis of
the Stanford-Binet IV and the WISC-R for Ages 9-0 to 12-11 Years

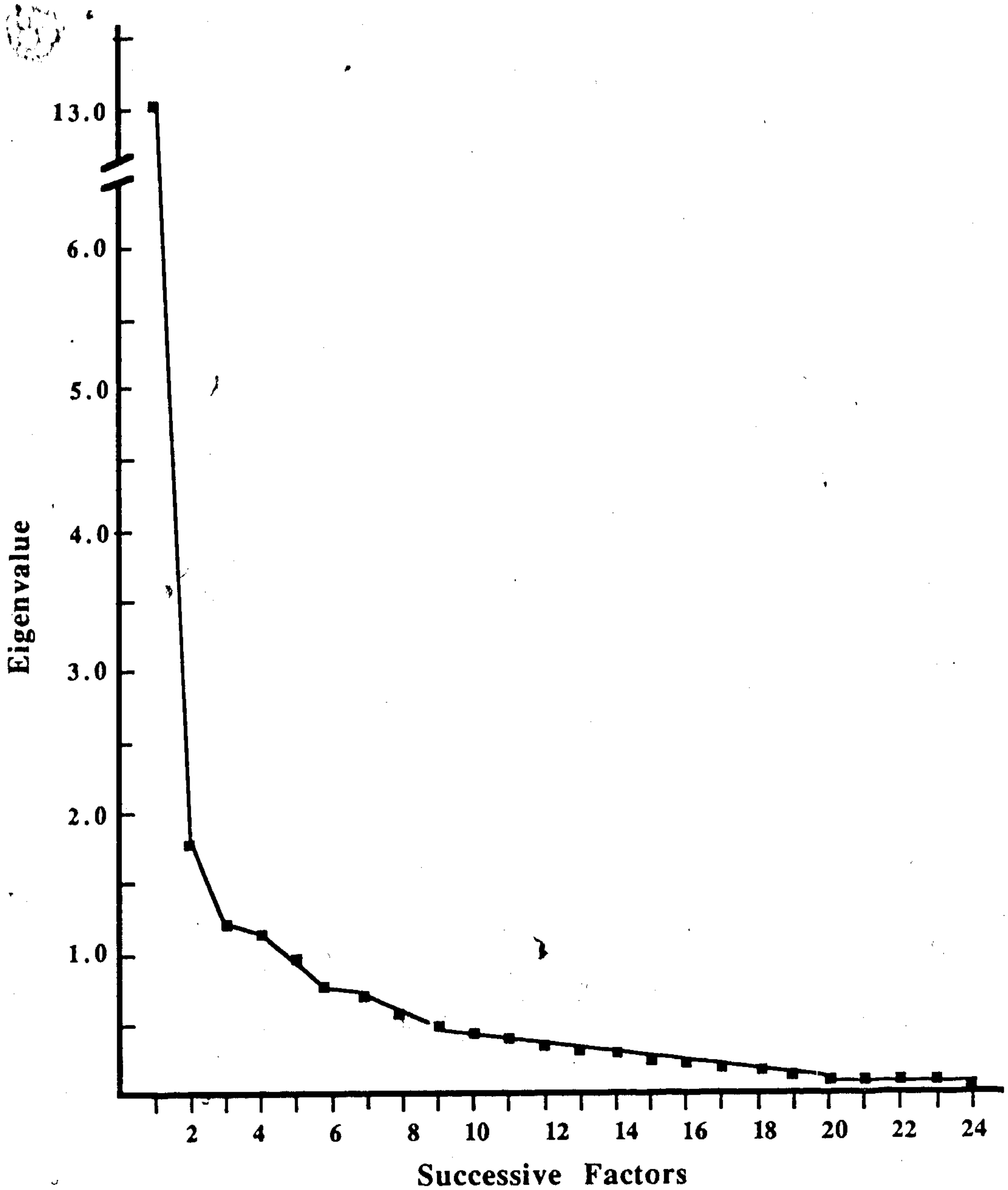


Table 47

Varimax Rotated Two- and Three-Component Solutions for the Joint Analysis of the Stanford-Binet IV and WISC-R: Ages 9-0 to 12-11 Years[a]

	Age 9-0 to 12-11 Tests	Two Components Extracted I	II	h^2	Three Components Extracted I	II	III	h^2
WISC-R Tests	Information	.81	.38	.80	.78	.36	.25	.81
	Similarities	.67	.48	.68	.64	.54	.20	.74
	Arithmetic	.72	.36	.65	.69	.29	.28	.65
	Vocabulary	.77	.41	.76	.74	.36	.29	.76
	Comprehension	.69	.52	.75	.66	.50	.30	.77
	Digit Span	.82	.11	.69	.81	.03	.23	.71
	Picture Completion	.31	.69	.57	.26	.62	.38	.60
	Picture Arrangement	.15	.44	.22	.13	.60	.01	.38
	Block Design	.43	.75	.75	.37	.44	.67	.78
	Object Assembly	.20	.82	.71	.14	.61	.57	.71
	Coding	.31	.33	.20	.27	-.08	.60	.43
	Mazes	.04	.62	.38	.00	.54	.33	.40
Stanford-Binet IV Tests	Vocabulary	.77	.43	.78	.74	.39	.29	.79
	Comprehension	.68	.45	.67	.65	.45	.25	.69
	Absurdities	.58	.63	.74	.53	.52	.43	.74
	Pattern Analysis	.43	.66	.63	.37	.36	.64	.67
	Copying	.25	.72	.58	.19	.42	.62	.60
	Matrices	.60	.60	.72	.56	.50	.40	.72
	Quantitative	.56	.42	.49	.54	.59	.04	.64
	Number Series	.67	.52	.72	.64	.51	.29	.75
	Bead Memory	.34	.66	.55	.27	.21	.79	.74
	Memory for Sentences	.86	.16	.76	.84	.07	.25	.77
	Memory for Digits	.80	.11	.65	.78	-.10	.37	.75
	Memory for Objects	.40	.49	.40	.35	.20	.54	.46
Sum of Squared Loadings		8.22	6.61		7.40	4.38	4.28	
% of Total Variance		34.3	27.5		30.8	18.3	17.8	
Cummulative % Var.		34.3	61.8		30.8	49.1	66.9	

[a] The rotated principal components listed in this table are based on the entire sample for this age group (n=66). Missing data were estimated through the regression procedure outlined previously.

Table 48
Varimax Rotated Four-Component Solution for the Joint Analysis of the Stanford-Binet IV and WISC-R: Ages 9-0 to 12-11 Years[a]

Age 9-0 to 12-11		Varimax Rotated Principal Components				
	Tests	I	II	III	IV	h^2
WISC-R Tests	Information	.83	.33	.09	.05	.82
	Similarities	.72	.38	.00	.26	.74
	Arithmetic	.71	.30	.20	.11	.65
	Vocabulary	.78	.34	.16	.11	.76
	Comprehension	.74	.47	.00	.11	.79
	Digit Span	.75	.05	.41	.04	.74
	Picture Completion	.36	.59	.04	.36	.61
	Picture Arrangement	.17	.16	.17	.88	.85
	Block Design	.43	.77	.17	-.01	.80
	Object Assembly	.23	.74	.13	.32	.73
	Coding	.13	.25	.80	.22	.76
	Mazes	.11	.57	-.09	.23	.40
Stanford-Binet IV Tests	Vocabulary	.79	.37	.13	.12	.79
	Comprehension	.70	.34	.14	.27	.70
	Absurdities	.61	.58	.10	.16	.74
	Pattern Analysis	.43	.71	.15	-.11	.72
	Copying	.23	.68	.26	.18	.61
	Matrices	.64	.55	.07	.13	.73
	Quantitative	.67	.34	-.25	.21	.67
	Number Series	.74	.48	-.05	.08	.79
	Bead Memory	.30	.77	.24	-.29	.83
	Memory for Sentences	.81	.13	.33	-.04	.77
	Memory for Digits	.69	.11	.51	-.11	.77
	Memory for Objects	.35	.47	.35	.04	.47
Sum of Squared Loadings		8.33	5.63	1.72	1.54	
% of Total Variance		34.7	23.5	7.2	6.4	
Cummulative % Var.		34.7	58.2	65.4	71.8	

[a] The rotated principal components listed in this table are based on the entire sample for this age group (n=66). Missing data were estimated through the regression procedure outlined previously.

Table 49
Varimax Rotated Five-Component Solution for the Joint Analysis of the Stanford-Binet IV and WISC-R: Ages 9-0 to 12-11 Years[a]

Age 9-0 to 12-11		Varimax Rotated Principal Components					
	Tests	I	II	III	IV	V	h^2
WISC-R Tests	Information	.65	.29	.57	.00	.04	.83
	Similarities	.77	.22	.28	.11	.15	.76
	Arithmetic	.56	.23	.50	.15	.08	.65
	Vocabulary	.76	.16	.38	.26	.00	.82
	Comprehension	.80	.29	.27	.14	-.01	.83
	Digit Span	.29	.16	.85	.09	.13	.86
	Picture Completion	.66	.35	-.11	.32	.19	.72
	Picture Arrangement	.24	.16	.07	.17	.87	.87
	Block Design	.44	.72	.26	.19	-.04	.82
	Object Assembly	.39	.67	.02	.24	.27	.73
	Coding	.07	.15	.27	.82	.18	.81
	Mazes	.05	.74	.16	-.25	.32	.74
Stanford-Binet IV Tests	Vocabulary	.73	.23	.43	.17	.04	.80
	Comprehension	.71	.19	.33	.22	.18	.73
	Absurdities	.63	.48	.30	.15	.10	.74
	Pattern Analysis	.43	.65	.26	.17	-.14	.73
	Copying	.27	.65	.16	.28	.16	.63
	Matrices	.70	.40	.25	.18	.03	.74
	Quantitative	.71	.25	.21	-.19	.14	.67
	Number Series	.67	.43	.40	-.07	.04	.80
	Bead Memory	.38	.65	.14	.35	-.37	.85
	Memory for Sentences	.48	.11	.71	.16	-.02	.78
	Memory for Digits	.24	.18	.83	.23	-.03	.83
	Memory for Objects	.38	.34	.20	.44	-.03	.49
Sum of Squared Loadings		7.19	4.16	3.77	1.76	1.30	
% of Total Variance		30.0	17.3	15.7	7.3	5.4	
Cummulative % Var.		30.0	47.3	63.0	70.3	75.7	

[a] The rotated principal components listed in this table are based on the entire sample for this age group (n=66). Missing data were estimated through the regression procedure outlined previously.

In the five-component varimax solution (Table 49) the first component is Verbal-Educational / Crystallized Intelligence; it has highest loadings from Information, Similarities, W.Vocabulary, W.Comprehension, Picture Completion, Matrices, B.Vocabulary, B.Comprehension, Absurdities,, Quantitative, Number Series, and Arithmetic; it also has small loadings from Block Design, Pattern Analysis, and Memory for Sentences. The second component is Spatial-Mechanical / Fluid Intelligence; it has high loadings from Block Design, Object Assembly, Mazes, Pattern Analysis, Copying, and Bead Memory; it also has small loadings from Absurdities, Matrices, and Number Series. The third component is Memory or Concentration; it has high loadings from Information, Arithmetic, Digit Span, Memory For Sentences, and Memory For Digits. The fourth component is defined primarily by Coding though Bead Memory and Memory for Objects have small loadings; this component involves Visual Memory. The fifth component is largely a singlet, with only Picture Arrangement having a high loading and with Bead Memory having a small negative loading; as such, this component appears to be bipolar, with Sequencing of Situations (Picture Arrangement) and Sequencing of Beads (Bead Memory) as the two poles.

The four-component solution also has the first component as Verbal-Educational / Crystallized Intelligence; high loadings include Information, Similarities, Arithmetic, W.Vocabulary, W.Comprehension, Digit Span, B.Vocabulary, B.Comprehension, Absurdities, Matrices, Quantitative, Number Series, Memory For Sentences, and Memory For Digits. The second component is once again Spatial-Mechanical / Fluid Intelligence; it has high loadings from Picture Completion, Block Design, Object Assembly, Mazes, Absurdities, Pattern Analysis, Copying, Matrices, and Bead Memory; it also has small loadings from Comprehension, Number Series, and Memory For Objects. The third component is Memory or Concentration; the subtests with strongest loadings on this component are Digit Span, Coding, and Memory For Digits. The fourth component is once again primarily defined by Picture Arrangement, although Picture Completion does a a weak loading; this component deals with Awareness of the Environment.

In both the two-, and three-component solutions, the component pattern is complex for several subtests; their loadings on more than one factor are moderately strong. In both of these solutions that first two components could be conceived of as Crystallized and Fluid Intelligence respectively. However, the complexity of many of the subtests suggests that too much variance may have been compressed onto the first two or three components when these solutions are rotated.

To allow for a general component to emerge, quartimax rotations were also used. The five-component quartimax solution is presented in Table 50. Indeed, the first component is a general one, with high loadings from all subtests except Picture Arrangement, Coding, and Mazes. It is worth noting that the Verbal subtests seem to have the highest loading on the general component; this may account for the fact that no Verbal component emerges from the quartimax rotations. Also, as would be expected, both Matrices and Number Series have very strong loadings on this first component. The second component is Auditory Short-Term Memory, with high loadings from Digit Span, Memory For Digits, and Memory For Sentences. The third component is Visual Planning, and the strongest loadings on this component come from Mazes, Copying, Block Design, Object Assembly, and Pattern Analysis. On the fourth component, only Coding has a strong loading, while Memory For Objects has a small loading, and Quantitative has a small negative loading; as such, this component may be bipolar, with one pole being Passive Copying of Numbers (Coding) and the other pole being Active Numerical Processing (Quantitative). The fifth component is also bipolar, with a strong loading from Picture Arrangement and a moderate negative loading from Bead Memory; this component has Sequencing of Situations (Picture Arrangement) and Sequencing of Beads (Bead Memory) as the two poles.

Table 50
Quartimax Rotated Five-Component Solution for the Joint Analysis of the Stanford-Binet IV and WISC-R: Ages 9-0 to 12-11 Years[a]

Age 9-0 to 12-11		Quartimax Rotated Principal Components					
	Tests	I	II	III	IV	V	h^2
WISC-R Tests	Information	.86	.26	-.01	-.15	.00	.83
	Similarities	.84	-.05	-.11	-.09	.15	.76
	Arithmetic	.77	.21	-.05	.01	.05	.65
	Vocabulary	.87	.03	-.23	.04	.01	.82
	Comprehension	.89	-.08	-.11	-.08	-.02	.83
	Digit Span	.65	.65	.02	.05	.09	.86
	Picture Completion	.71	-.42	.01	.14	.16	.72
	Picture Arrangement	.36	-.06	.20	.15	.82	.87
	Block Design	.79	-.05	.39	.13	-.17	.82
	Object Assembly	.66	-.25	.42	.19	.13	.73
	Coding	.40	.10	-.06	.79	.14	.81
	Mazes	.38	.03	.74	-.15	.14	.74
Stanford-Binet IV Tests	Vocabulary	.88	.09	-.14	-.03	.03	.80
	Comprehension	.82	.01	-.14	.03	.17	.73
	Absurdities	.85	-.03	.14	.01	.03	.74
	Pattern Analysis	.74	-.04	.31	.10	-.26	.73
	Copying	.62	-.09	.42	.26	.02	.63
	Matrices	.86	-.09	.02	.00	-.01	.74
	Quantitative	.73	-.06	.00	-.36	.12	.67
	Number Series	.86	.09	.12	-.21	-.02	.80
	Bead Memory	.68	-.16	.23	.27	-.48	.85
	Memory for Sentences	.74	.46	-.15	.04	-.03	.78
	Memory for Digits	.64	.62	-.02	.19	-.07	.83
	Memory for Objects	.61	-.06	.03	.34	-.09	.49
Sum of Squared Loadings		12.91	1.46	1.42	1.26	1.17	
% of Total Variance		53.8	6.1	5.9	5.3	4.9	
Cummulative % Var.		53.8	59.9	65.8	71.1	76.0	

a The rotated principal components listed in this table are based on the entire sample for this age group (n=66). Missing data were estimated through the regression procedure outlined previously.

Principal Components From the Joint Analysis of the Stanford-Binet IV and the WISC-R for the Age Group 13-0 to 16-11 Years

Returning to Table 41, it can be seen that for the age group of 13-0 to 16-11 years, only two components have eigenvalues above one. As such, the eigenvalue-one criteria suggests that only two components be rotated. This atypical sample, and the unusually high correlations for this age group result in an artifactually strong first component, which accounts for 73.4% of the total subtest variance. Consequently, as in other analyses for this age group, the results from the joint analysis of the Stanford-Binet IV and WISC-R must be interpreted cautiously as the true number of components may be obscured in this age group. Hence, the discussion of the result from this group will be minimal.

A plot of the successive components for this age group is provided in Figure 20. Using this plot, it is relatively clear that a four-component solution is appropriate. Therefore, two-, three-, and four-component solutions for this age group are reported in Table 51 and Table 52.

In the two-component solution the first component has relatively strong loadings from most subtests; however, if only subtests with loadings above .60 are considered in defining a component[49], the first component represents Crystallized Intelligence. If similarly criteria are used, the second component represents Fluid Intelligence.

In the three-component solution, the first two component once again turn out to be Crystallized and Fluid Intelligence when subtests that have loadings of near .60 on the respective components are used to identify the components. The first component, Crystallized Intelligence, has loadings from Information, Similarities, Arithmetic, W.Vocabulary, W.Comprehension, B.Vocabulary, B.Comprehension, Matrices, Quantitative, Number Series, and Memory For Sentences. Fluid Intelligence, the second component, has loadings from Picture Completion, Picture Arrangement, Block Design, Object Assembly, Mazes, Pattern Analysis, Matrices, and Bead Memory. Finally, the third

[49] This stringent of a loading may be necessary at this level because all loadings are relatively high because of the heterogeneity of this age group.

component is Short-Term Memory, with loadings from Digit Span, Object Assembly, Memory for Sentences, Memory for Digits, and Memory for Objects.

Figure 20

Scree Plot for the Joint Analysis of the Stanford-Binet IV and the WISC-R for Ages 13-0 to 16-11 Years

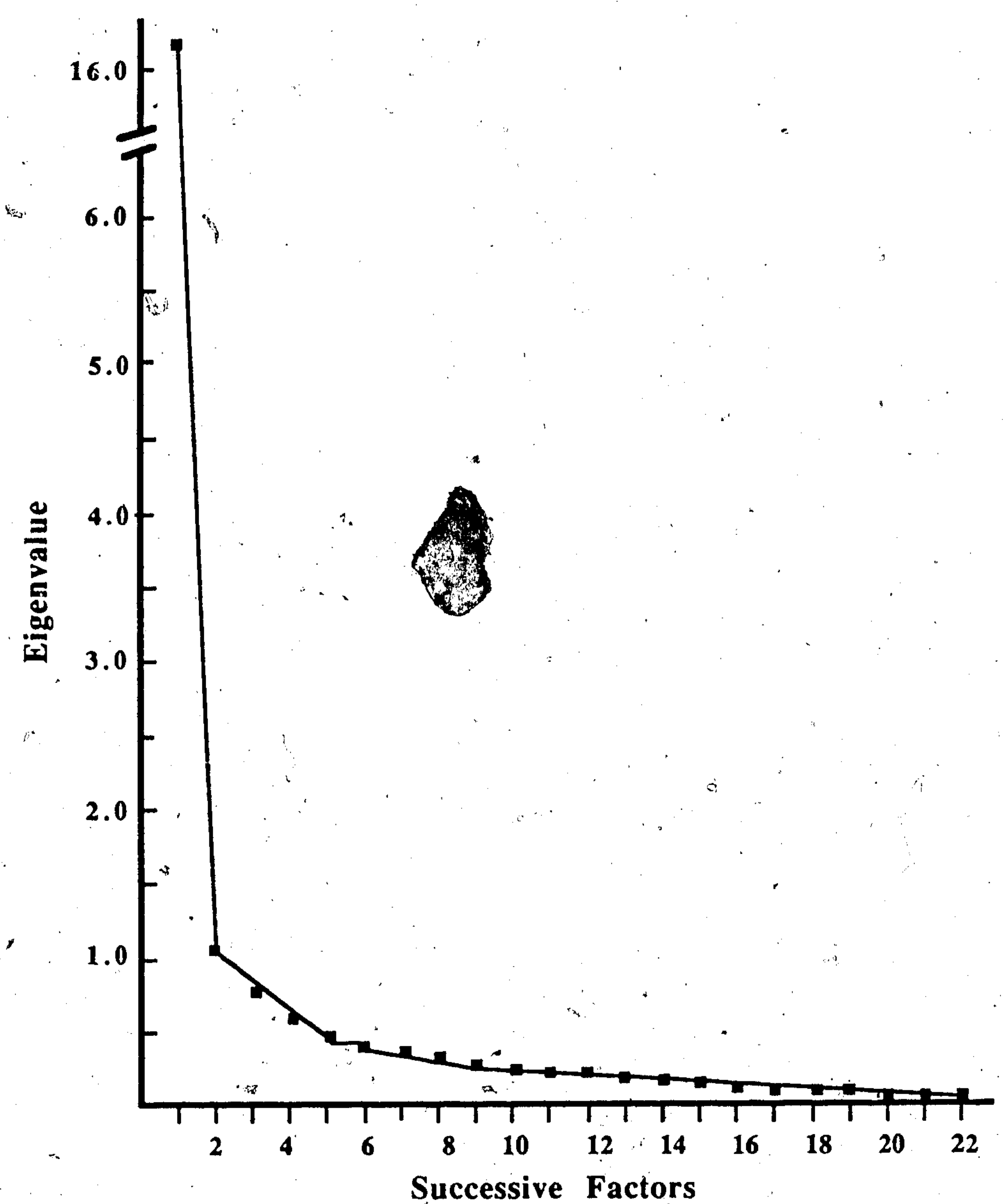


Table 51
Varimax Rotated Two-, and Three-Component Solutions for the Joint Analysis of the Stanford-Binet IV and WISC-R: Ages 13-0 to 16-11 Years[a]

Age 13-0 to 16-11		Varimax Rotated Principal Components						
		Two Components Extracted			Three Components Extracted			
	Tests	I	II	h^2	I	II	III	h^2
WISC-R Tests	Information	.84	.40	.86	.78	.33	.39	.87
	Similarities	.76	.53	.87	.74	.49	.31	.89
	Arithmetic	.69	.54	.77	.68	.50	.28	.79
	Vocabulary	.90	.29	.89	.82	.23	.41	.90
	Comprehension	.81	.40	.81	.79	.36	.29	.84
	Digit Span	.77	.41	.76	.57	.30	.63	.82
	Picture Completion	.30	.81	.74	.37	.81	.09	.80
	Picture Arrangement	.41	.73	.70	.37	.69	.30	.71
	Block Design	.61	.63	.77	.58	.59	.31	.78
	Object Assembly	.50	.68	.71	.30	.58	.61	.80
	Coding	.54	.54	.58	.43	.47	.43	.59
	Mazes	.25	.85	.79	.16	.80	.37	.80
Stanford-Binet IV Tests	Vocabulary	.85	.35	.84	.83	.30	.31	.87
	Comprehension	.80	.50	.88	.77	.45	.32	.90
	Pattern Analysis	.62	.65	.82	.57	[illegible]	.36	.82
	Matrices	.66	.64	.84	.60	.58	.40	.84
	Quantitative	.78	.50	.86	.71	.43	.41	.86
	Number Series	.77	.53	.88	.73	.48	.36	.89
	Bead Memory	.56	.66	.76	.53	.61	.33	.76
	Memory for Sentences	.83	.35	.81	.64	.24	.61	.84
	Memory for Digits	.69	.48	.71	.46	.37	.68	.81
	Memory for Objects	.63	.40	.55	.33	.26	.79	.80
Sum of Squared Loadings		10.31	6.89		8.17	5.59	4.22	
% of Total Variance		46.9	31.3		37.1	25.4	19.2	
Cummulative % Var.		46.9	78.2		37.1	62.5	81.7	

[a] The rotated principal components listed in this table are based on the entire sample for this age group (n=56). Missing data were estimated through the regression procedure outlined previously.

Table 52
Varimax Rotated Four-Component Solution for the Joint Analysis of the Stanford-Binet IV and WISC-R: Ages 13-0 to 16-11 Years[a]

Age 13-0 to 16-11		Varimax Rotated Principal Components				
	Tests	I	II	III	IV	h^2
WISC-R Tests	Information	.78	.30	.35	.23	.87
	Similarities	.73	.39	.27	.35	.89
	Arithmetic	.66	.57	.26	.16	.85
	Vocabulary	.82	.20	.38	.19	.90
	Comprehension	.79	.26	.25	.30	.84
	Digit Span	.58	.22	.59	.28	.82
	Picture Completion	.35	.51	.04	.68	.84
	Picture Arrangement	.34	.79	.30	.17	.86
	Block Design	.56	.53	.28	.34	.79
	Object Assembly	.29	.50	.59	.35	.81
	Coding	.44	.04	.36	.73	.85
	Mazes	.15	.52	.33	.65	.82
Stanford-Binet IV Tests	Vocabulary	.82	.29	.28	.18	.87
	Comprehension	.77	.35	.28	.35	.91
	Pattern Analysis	.56	.46	.32	.43	.82
	Matrices	.58	.46	.36	.40	.84
	Quantitative	.70	.38	.38	.29	.86
	Number Series	.72	.42	.32	.30	.89
	Bead Memory	.50	.57	.30	.32	.77
	Memory for Sentences	.66	.12	.57	.30	.86
	Memory for Digits	.46	.29	.65	.28	.81
	Memory for Objects	.34	.28	.78	.12	.82
Sum of Squared Loadings		8.00	3.86	3.66	3.05	
% of Total Variance		36.4	17.5	16.6	13.9	
Cummulative % Var.		36.4	53.9	70.5	84.4	

[a] The rotated principal components listed in this table are based on the entire sample for this age group (n=56). Missing data were estimated through the regression procedure outlined previously.

Examination of the four-component solution in Table 52 reveals that the first component is still Crystallized Intelligence, with loadings from Information, Similarities, Arithmetic, W.Vocabulary, W.Comprehension, Digit Span, Block Design, B.Vocabulary, B.Comprehension, Pattern Analysis, Matrices, Quantitative, Number Series, and Memory For Sentences. The second component is Perceptual Organization, with loadings from Picture Completion, Picture Arrangement, Block Design, Object Assembly, Mazes, and Bead Memory. Again, as in the three-component solution, the third component is Short-Term Memory, with loadings from Digit Span, Object Assembly, Memory For Sentences, Memory For Digits, and Memory For Objects. Finally, the fourth component is Visual Scan; this component is defined by Picture Completion, Coding, and Mazes.

To allow for comparison with the varimax solution, the quartimax rotated four-component solution for the oldest age group is reported in Table 53. As expected, all the subtests have strong loadings on the general component, which accounts for 73.3% of the total subtest variance. The second component is Visual Scan, and has highest loadings from Picture Completion and Mazes. The third component is Short-Term Memory, and has loadings from Memory For Objects, Memory For Digits, and Object Assembly. The four component is bipolar, with one pole being Formation of Abstract Associations (Coding) and the other pole being Formation of Social Associations (Picture Arrangement). It is further important to note that the second, third, and fourth components are relatively small and each of these being primarily a doublet.

Table 53

Quartimax Rotated Four-Component Solution for the Joint Analysis of the Stanford-Binet IV and WISC-R: Ages 13-0 to 16-11 Years[a]

	Age 13-0 to 16-11 Tests	Quartimax Rotated Principal Components I	II	III	IV	h^2
WISC-R Tests	Information	.91	-.18	-.05	.03	.87
	Similarities	.93	-.03	-.13	.01	.89
	Arithmetic	.88	-.05	-.09	-.25	.85
	Vocabulary	.90	-.28	-.03	.10	.90
	Comprehension	.89	-.14	-.15	.09	.84
	Digit Span	.86	-.09	.24	.13	.82
	Picture Completion	.73	.48	-.26	.04	.84
	Picture Arrangement	.77	.24	.06	-.45	.86
	Block Design	.87	.12	-.07	-.12	.79
	Object Assembly	.80	.25	.32	-.09	.81
	Coding	.75	.25	.00	.48	.85
	Mazes	.70	.57	.08	.02	.82
Stanford-Binet IV Tests	Vocabulary	.89	-.24	-.12	.01	.87
	Comprehension	.94	-.07	-.14	.06	.91
	Pattern Analysis	.89	.16	-.04	-.01	.82
	Matrices	[illegible]	.12	-.01	-.02	.84
	Quantitative	.93	-.07	-.01	.00	.86
	Number Series	.94	-.05	-.07	-.03	.89
	Bead Memory	.85	.16	-.02	-.17	.77
	Memory for Sentences	.87	-.16	.19	.22	.86
	Memory for Digits	.83	.01	.33	.06	.81
	Memory for Objects	.73	-.04	.53	-.02	.82
Sum of Squared Loadings		16.13	1.06	0.76	0.64	
% of Total Variance		73.3	4.8	3.5	2.9	
Cummulative % Var.		73.3	78.1	81.6	84.5	

[a] The rotated principal components listed in this table are based on the entire sample for this age group (n=56). Missing data were estimated through the regression procedure outlined previously.

CHAPTER V

DISCUSSION

Before launching into a discussion of the results, it is appropriate to review some of the fundamental rationales for this thesis. The Stanford-Binet IV, like all other psychological tests, needs to be validated before clinicians can place whole-hearted faith in it. Unfortunately, little research has yet been done on this new scale. Although there have been some studies that have looked at the concurrent validity between the Stanford-Binet IV and the WISC-R, most of these studies were done by Thorndike et al. (1986) and were reported in the *Technical Manual* for the Stanford-Binet IV. Hence, there is a great need for further independent examination of the concurrent validity of the Stanford-Binet IV.

An even greater [illegible] is assessment of the construct validity of the Stanford-Binet IV. Because the confirmatory factor analyses done by Thorndike et al. do not clearly support the theoretical structure of the test, it seems appropriate to use exploratory factor analyses so that the data can "speak" for themselves. Even so, the factoring procedure chosen should allow the emergence of both a general factor as well as group factors; consequently, the principal components method of factoring, with quartimax rotation of factors was undertaken. Further, in order to facilitate comparison with the literature on the WISC-R, there was a need to retain some similarity in factoring procedures normally used on that test (e.g., using varimax rotations). Finally, the joint factoring of the Stanford-Binet IV and the WISC-R was undertaken to provide a better understanding of the Stanford-Binet IV and to add to the body of theoretical information on intelligence.

A few qualifications about the findings from this thesis are in order at this point. First, though three age groups (6-0 to 8-11 years, 9-0 to 12-11 years, and 13-0 to 16-11 years) were used, the heterogeneity of the oldest age group resulted in atypically high correlations not only between the subtests on the WISC-R but also on the Stanford-Binet IV; further, these extremely high correlations complicated the factor analyses at this age level and resulted in an artifactually high general component for the WISC-R, the Stanford-Binet IV, and the joint analysis of the WISC-R and Stanford-Binet IV. Consequently, given the distortion of the WISC-R factor structure for this 13-0 to 16-11 age group, it was decided that the results for this group would not be discussed. Hence, this discussion is limited to the results from the two younger age groups.

It must also be remembered that the sample sizes for these two age groups are small (n=46 and n=66). Though Baggaley (1982) suggests that high correlations between subtests allows the use of fewer subjects, these small sample sizes also reduce the stability of the correlations (Loo, 1983). Hence, the results from this thesis must be treated as tentative given that the small sample sizes make it slightly more difficult to determine the correct number of components, and also introduce more error variance into the correlations. Another qualification of the current results has to do with the fact that a clinical sample of children was used and there was a wide range of abilities; hence, correlations may have been slightly elevated by this heterogeneity and may result in slightly higher factor loadings than would be expected if a more "normal" sample had been used. However, despite these limitations, the general trends found in this thesis should be reasonably stable, especially for the larger components.

The intention of this section is not to review all the results, but to highlight the basic objectives and hypotheses that were laid out in Chapter One; also, the contents of this chapter provide a discussion of the results and their implications for use of the Stanford-Binet IV, and a few suggestions for further research that is needed on this new test.

A. Descriptive Statistics

In order to determine the concurrent validity of the Stanford-Binet IV, means, standard deviations, and correlations between subtests, Areas, and the Composite were computed. These were compared to the same statistics from the WISC-R.

First, the mean for the Stanford-Binet IV Composite score is approximately 2 points lower than the mean for the WISC-R Full Scale IQ; this finding is similar to that found by Thorndike et al. (1986). Displayed graphically (Figure 5), the relationship between these two scales is linear. Taken as a whole, this suggests that the Stanford-Binet IV Composite score and the WISC-R FSIQ are similar enough to be used almost interchangeably, though it must be remembered that the Stanford-Binet IV Composite score will be about 2 or so points lower than the WISC-R FSIQ. Similarly, for the younger and middle age groups the correlations between these two scores are .87 and .91 respectively. Thus, the Stanford-Binet IV Composite score has good concurrent validity with the WISC-R FSIQ.

Moving to the other major scales on the Stanford-Binet IV and WISC-R, the mean on the Verbal Reasoning Area of the Stanford-Binet IV and the mean for the Verbal Scale IQ on the WISC-R are fairly similar. As a general rule, the mean scores on the Verbal Reasoning Area is 2.3 to 3.7 points higher than the comparable scores on the WISC-R VIQ. For the youngest age group these two scales correlate .81 while for the middle age group they correlate .89. Hence, it can be concluded that these two scores are considerably related and that there is a good deal of concurrent validity between them.

Next, it is logical to compare the Abstract/Visual Reasoning Area with the WISC-R Performance Scale score. For the youngest age group the mean for the Abstract / Visual Reasoning Area is 7 points lower than the WISC-R PIQ, and the correlation between the two is .69. Meanwhile, for the middle age group the WISC-R PIQ is only 3.2 points lower than the Abstract / Visual Reasoning Area score and the correlation between the two was .81. Thus, for the two different age groups the relationship between these two scores is different. For the middle age group there is fairly good concurrent validity between the Abstract / Visual Reasoning Area and the WISC-R PIQ. However, for the youngest age group, the concurrent validity for the Abstract / Visual Reasoning Area is questionable; this is not unexplainable given that for the youngest age group only two subtests are included on the Abstract / Visual Area; or that Thorndike et al. (1986) found that for their 2 through 6 age group, and 7 through 11 age group, that the Abstract / Visual factor was not well defined.

For the Quantitative and Short-Term Memory Areas on the Stanford-Binet IV, there are no equivalent scales on the WISC-R that can provide an index of concurrent validity. Yet, the relationship between these two areas and the three major scales on the WISC-R can be seen by examining Tables 23 and 24. For the younger group, both the Quantitative and Short-Term Memory Areas have reasonably strong correlations with the WISC-R VIQ, PIQ, and FSIQ and range from .50 to .71.[3] The relationship between these two Areas and the three WISC-R scales is stronger for the middle age group with the correlations ranging from .61 to .81. Of particular theoretical importance, the Quantitative Reasoning Area had a correlation of .81 with the WISC-R VIQ for the middle age group. This is particularly interesting given that both Vernon (1961) and Cattell (1971) suggest that quantitative ability belongs to a Verbal-Educational / Crystallized Intelligence factor. This finding lends some indirect support to Thorndike et al.'s (1986) suggestion that the Verbal and Quantitative Reasoning Areas are lower level factors stemming from Crystallized Intelligence.

Before moving on, a final set of correlations between the major scales on the two tests needs to be discussed for the middle age group. From Table 24 it is seen that the Stanford-Binet IV Composite score correlates .91, .78, and .91 with the WISC-R VIQ, PIQ, and FSIQ respectively. The implication is that the Stanford-Binet IV Composite score may reflect a highly Verbal bias at this age[50] ; in contrast, for the younger group the Composite may not be as Verbally biased.

On the Stanford-Binet IV, the means and standard deviations on the different subtests tend to be similar, though there are a few exceptions at different age levels. For example, in the 6-0 to 8-11 year age group, the Copying subtest has a lower mean but a comparable standard deviation to other subtests; in contrast, the Quantitative subtest has a comparable mean, but lower standard deviation than other subtests. These problems may result from standardization difficulties, or from a restricted range of difficulty at this age. For the 9-0 to 12-11 age group, the mean scores on the Quantitative, Bead Memory, and Memory for Sentences subtests are slightly lower than other subtest means.

On the WISC-R, the subtest means are quite similar, though the means for Information, Arithmetic, Digit Span, and Coding tend to be slightly lower than on other subtests. As such, this implies the presence of the third factor for the WISC-R as is commonly found in clinical populations. The question that this raises is whether or not the low means on the given Stanford-Binet IV subtests occur because these subtests load on the same factor or factors. From the current results a conclusion of this nature is not clearly warranted; further studies need to replication this finding before the reason for these low means is known[51].

A particularly surprising finding was that the standard deviations for the Stanford-Binet IV Composite and Area scores are lower than the standard deviations for the WISC-R FSIQ, VIQ, and PIQ by two or more points. Taking into consideration that the major scales on the Stanford-Binet are designed with a standard deviation of 16, and the WISC-R with 15, this finding is disturbing. Though it is possible that this finding is idiosyncratic

50 It could alternately be argued that the WISC-R PIQ does not measure "intelligence" as well as does the WISC-R VIQ, but given the perennial difficulty of developing tests that are not biased towards measuring verbal abilities, this alternate argument seems less defensible than concluding that the Composite score may be biased to measuring Verbal ability.

51 This is particularly true given that no tests of significance were used to determine whether these variations in the means were significant or due to chance.

to this sample, it is noteworthy that Thorndike et al. (1986) reported the same trend in some samples of non-exceptional and learning disabled children. Until further large scale studies address this concern, clinicians need to be aware that standard scores (z-scores) and hence percentiles earned on the Stanford-Binet IV may contain more error than suggested by Thorndike et al.; this may be particularly true of the Quantitative Reasoning Area for children under 13 years of age.

The correlations between the Stanford-Binet IV subtests, Areas, and Composite that are found in the two younger age groups were very similar to those reported by Thorndike et al. (1986). This was encouraging because it meant that the factoring done in this thesis started from a pool of data that was quite similar to that used by Thorndike et al.. It also suggests that the correlations between subtests, Areas, and the Composite are not that different between "normal" and clinically referred children. Finally, the current correlations tend to validate those found by Thorndike et al. and to those found in Carvajal and Weyand's (1986) limited study.

B. Factor Analyses of the WISC-R

The purpose of factoring the WISC-R was two-fold: to validate the current sample and age groups as typical of other clinical samples; and to provide a reference structure against which the Stanford-Binet IV and the joint factoring of the Stanford-Binet IV and the WISC-R could be compared. Because the majority of the factoring done on the WISC-R has used varimax rotations, the current component analyses on the WISC-R also used varimax rotations.

For the 6-0 to 8-11 age group, the eigenvalue-one criterion indicates that a three-component solution was most appropriate; examination of two-, and four-component solutions also suggested that a three-component solution was the best choice. Verbal Comprehension emerged as the first component, accounting for 32.1% of the total subtest variance; it has high loadings from Information, Similarities, Arithmetic, Vocabulary, and Comprehension; Digit Span, Picture Arrangement and Coding also has moderate loadings .

on this component. Perceptual Organization emerged as the second component, accounting for 21.6% of the total subtest variance; it has high loadings from Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Digit Span. The third component, accounting for 13.6% of the total subtest variance, was different than the usual third factor on the WISC-R. This component had strong loadings from Block Design and Mazes, and a moderate loading from Information; thus, this component required Visual Planning and Reasoning.

Taken as a whole, the WISC-R component structure for the youngest age group is quite similar to that found in other clinical and "normal" populations (e.g., Kaufman, 1975) though there are some slight variations in the first two components, and the third component is unique. The later findings are not unexpected given that this was a clinical sample. However, there is sufficient similarity between the component structure that emerged and the prototypical factor structure (Kaufman, 1975) to accept the youngest age group as a valid and representative clinical sample.

In the 9-0 to 12-11 age group, both the eigenvalue-one and scree test criteria suggests that a two-component solution is best. The first component emerges as Verbal Comprehension, accounting for 39.9% of the total subtest variance; Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Digit Span have high loadings on this component while Picture Completion and Block Design have moderate loadings. Perceptual Organization emerges as the second component, accounting for 23.3% of the total subtest variance; Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Mazes have high loadings on this component. Overall, the component structure that emerged for this age group is indeed representative of the prototypic factor structure found in other samples (e.g., Kaufman, 1975). Therefore, this age group was accepted as a valid and representative clinical sample.

In the 13-0 to 16-11 age group, only one component has an eigenvalue greater than one. Because of the extreme heterogeneity of the scores in this age group, the first unrotated component dwarfs all other components. Consequently, the typical WISC-R factor structure does not emerge in two-, three-, or four-component solutions. The extreme diversity of the abilities of children in this age group results in a distortion of the magnitude of loadings on the components, increased complexity of the subtests, an artifactual general component, and obscuration of the number of components to extract. It was for these reasons that the decision was made not to discuss the results from this age group; although general patterns and directions of loadings may be correct, there is too much opaqueness in the results from this age group to warrant further discussion, especially given the limited size (n=56) of this age group.

C. Factor Analyses of the Stanford-Binet IV

The purpose behind factoring the Stanford-Binet IV was to examine its construct validity in the various age groups. Given the inconclusive and disappointing nature of the confirmatory factor analyses done by Thorndike et al. (1986), it was decided that exploratory factor analyses were needed to let the data speak for themselves and determine the underlying factor structure. Principal components analyses were done and components were rotated both by the quartimax and varimax criteria; the quartimax criterion proceeds very much like the factoring done by Thorndike et al.; the varimax criterion was included so that the results from the Stanford-Binet IV could be compared to the results normally found on the WISC-R. The important point to remember is that very different conclusions are sometimes drawn from the same data, depending on the type of rotations that are used (e.g., Thurstone (1938) and Eysenck (1939) or Paden (1981); or Kaufman (1975) and Wallbrown et al., (1975)).

To illustrate the blindness of some researchers, remember that Slate (1986) criticized the inclusion of group factors on the Stanford-Binet IV given that Thorndike et al. (1986) found the general factor to be so strong. Slate's erroneously compared the results from Thorndike et al.'s confirmatory factoring with Kaufman's (1975) varimax rotated principal axes. Even Vernon (1987) indirectly compared the factoring done on the Stanford-Binet IV with the factoring done on the WISC-R and WAIS-R and implied that one can directly compare the results. In criticizing the inclusion of group factors on the Fourth Edition, Vernon referred to the three group factors normally found on the Wechsler scales; however, he failed to clarify that these are found using varimax rotations--when factoring methods more comparable with those used on the Stanford-Binet IV are used on the WISC-R (e.g., Wallbrown et al., 1976; Blaha & Wallbrown, 1984) then a large general factor emerges (36% of total variance) while the group factors of *v:ed* and *k:m* are small (accounting for 5-6% of the total variance respectively). This later finding is quite comparable to the results found by Thorndike et al. (1986) on the Stanford-Binet IV. As Osberg (1986) wisely advises, more factoring of the Stanford-Binet IV is needed using "alternate techniques"--techniques that will allow more direct comparison between the results found for the WISC-R and those found for the Stanford-Binet IV. It was for this reason that both quartimax and varimax rotations were used in the factoring of the Stanford-Binet IV done in this thesis.

As a reminder, in Chapter One it was suggested that a strong general component was expected on the Stanford-Binet IV when quartimax rotations are used (or alternately the first unrotated component); when varimax rotations are used, three components of Crystallized Intelligence / Verbal-Educational, Fluid Intelligence / Spatial-Mechanical-Practical , and Short-Term Memory were expected. However, the number of factors to extract at each level was determined primarily by the eigenvalue-one and scree test criteria.

In the 6-0 to 8-11 age group, both the eigenvalue-one and scree test criteria suggests that a two-component solution is appropriate; however, the three-component solution seems more psychologically meaningful, and is more consistent with the findings by Thorndike et al. (1986). Therefore, the three-component solution is the one discussed here. For the quartimax rotations, the first component represents *g* and accounts for

43.9% of the total subtest variance; all subtests except Bead Memory (.21) have strong loadings on this first component. The second component represents Verbal ability, and has loadings from Vocabulary, Comprehension, and Memory for Sentences; this component accounts for 14.1% of the total subtest variance. Finally, the third component represents Abstract / Visual skills, has its highest loadings from Bead Memory and Pattern Analysis, and accounts for 13.9% of the total subtest variance. In many ways, the three-component quartimax solution for the youngest age group is very similar to the results from the confirmatory factor analysis that Thorndike et al. (1986) did on their 2 through 6 age group.

When the three-component varimax solution for the 6-0 to 8-11 age group is examined, the first component is Fluid Intelligence / Spatial-Mechanical ability, and accounts for 30.4% of the total subtest variance; Absurdities, Pattern Analysis, Copying, and Quantitative have loadings on this component. The second component that emerges is Verbal in nature, accounts for 26.0% of the total subtest variance, and has loadings from Vocabulary, Comprehension, and Memory for Sentences. As in the quartimax solution, the third component is identified by a very strong loading from Bead Memory, and a moderate loading from Pattern Analysis; though this component may represent Abstract / Visual ability, it could also be considered as Ability to Perceive and Organize Material; regardless of the name applied to this component, it accounts for 15.6% of the total subtest variance.

A key finding for the youngest age group is that neither a Quantitative nor Short-Term Memory component emerges. As such, for this age level these two Areas on the Stanford-Binet IV are not supported. Failure of a Memory component to emerge confirms the same finding by Thorndike et al. in their youngest age group; this is not surprising given that Memory for Sentences is Verbal in Nature, that Bead Memory requires both simultaneous processing and ability to perceive and organize materials (Thorndike, personal communication June 2, 1987), and the inappropriateness of Memory for Digits at this age level. Of further interest, the Bead Memory subtest is not found to load on the general component in this age group; as such, its presence in the battery for children in the 6-0 to 8-11 year age range may be questionable in that its relationship to intelligence is poorly defined.

Moving on to the 9-0 to 12-11 year age group, the eigenvalue-one criteria suggests that only one component should be extracted though the second component might be considered (eigenvalue of .97); meanwhile, the scree test suggests that three components should be extracted and rotated; similarly, the three-component solution seems the most meaningful. Consequently, only the three-component quartimax and varimax rotated solutions are discussed here. It is clearly recognized that other researchers may choose one of the other number-of-component solutions as being more appropriate based on theoretical or technical reasons; however, given the small sample size, the difficulty of determining the "true" number of components, and attempts to find psychological and theoretical meaningfulness, the three-component solution was chosen.

In the three-component quartimax rotated solution for the 9-0 to 12-11 age group, the first component represents *g* . This component accounts for 58.8% of the total subtest variance and is nearly identical to the first unrotated component. All twelve subtests have strong loadings on *g* with the range being from .61 to .89; the subtests with the lowest loadings are the four subtests from the Short-Term Memory Area, and Copying; the subtest with the highest loadings are Matrices, Number Series, and the Verbal Reasoning Area subtests. The second component emerges as Auditory Short-Term Memory. This component has loadings from Memory for Digits and Memory for Sentences, and accounts for 7.8% of the total subtest variance. The third component, accounting for 7.7% of the total subtest variance, is Visual / Conceptual Ability; it had its highest loadings from Bead Memory, Memory for Objects, and Copying; Quantitative has a moderate negative loading; it should be noted that Pattern Analysis has a weak loading (.28) on the third component,--

if this is taken into consideration, then one might leniently consider this third component as Abstract / Visual in nature.

In the varimax rotated three-component solution for the 9-0 to 12-11 age group, the first component that emerges is Crystallized Intelligence. This component accounts for 29.6% of the total subtest variance and has high loadings from Quantitative, Number Series, Vocabulary, Comprehension, Absurdities, and Matrices. The second component is Fluid Intelligence. It accounts for 26.1% of the total subtest variance and has high loadings from Pattern Analysis, Copying, Bead Memory, Memory for Objects, Absurdities, and Matrices. Several subtests are complex, with loadings on both Crystallized and Fluid Intelligence. Finally, the third component turns out to be Auditory Short-Term Memory. This component accounts for 18.6% of the total subtest variance and has high loadings from Memory for Digits, Memory for Sentences, and Vocabulary.

Noteworthy findings in the 9-0 to 12-11 age group include the failure of a Verbal component to emerge in the quartimax rotated solution; this is in contrast to Thorndike et al. (1986) who found a Verbal factor in their 7 through 11 age group. The fact that Vocabulary, Comprehension, and Absurdities have high loadings on *g* may be responsible for the fact that no Verbal factor emerges in the quartimax three-component solution. Similarly, it may explain why the Stanford-Binet IV Composite Score correlates so highly with the WISC-R VIQ (.91) for this age group. Therefore, the variance that would normally have been attributed to a verbal component seems to have been added to the general component; this may partially explain why *g* in this middle age group accounts for 58.8% of the total subtest variance rather than 42% as found by Thorndike et al.. Next, the emergence of a Short-Term Memory component lends some credence to Thorndike et al.'s finding of the same, and to the inclusion of this Area on the Stanford-Binet IV for this age level; however, the nature of the Short-Term Memory component in this age group was more restricted than Thorndike et al.'s Memory factor, and was limited to auditory material (i.e., Memory for Digits and Memory for Sentences).

As a whole, the quartimax solution in the middle age group supports a strong general component. It does not, however, support the retention of the Verbal Reasoning Area or the Quantitative Reasoning Area on the Stanford-Binet IV at this age level. Though the Abstract / Visual Reasoning Area receives some support[52], it should be duly noted that both Bead Memory and Memory for Objects seem to be related to this factor rather than to Short-Term Memory. Likewise, the Short-Term Memory Area for this age group should be restricted to Memory for Digits and Memory for Sentences.

The results from the varimax rotations for this age group are also heartening in terms of the theoretical structure of the Stanford-Binet IV. First, Fluid and Crystallized Intelligence did emerge, as did a Short-Term Memory component. Further, on the Crystallized Intelligence component all three of the Verbal Reasoning Area, and both the Quantitative Reasoning Area subtests have reasonably strong loadings, just as would be expected according to Thorndike et al.'s theoretical structure. Unfortunately, the varimax solution does not deal with the general component and thus only partially supports the theoretical structure of the Stanford-Binet IV.

D. Joint Factor Analyses of the Stanford-Binet IV and the WISC-R

The purpose of jointly factoring the Stanford-Binet IV and the WISC-R was to clarify which subtests from the two tests would load on the same components. As Kieth (1987) has suggested, jointly factoring a new test with an older well understood test can help researchers evaluate the merits of the new test. By jointly factoring the subtests from these two tests, four objectives were considered: (a) to see if both tests are measuring the same construct of intelligence; that is, which subtests from the two tests will load on the same general component which represents *g* ; (b) to provide a better understanding of which subtests from the two tests load on the same components; (c) to determine whether

52 Given that Pattern Analysis only has a weak loading on this component, the "Abstract" label may be more questionable; a name of Visual / Conceptual may be more fitting for this component.

the component structure of the WISC-R remains constant when the subtests from the Stanford-Binet IV are added, or whether new components are defined given that more subtests are available to define such possible components; and (d) to offer some suggestions for clinicians wanting to supplement subtests from either the WISC-R or Stanford-Binet IV when using only one of these tests to assess children.

In terms of general expectations, it was believed that a strong general component would emerge with most of the subtests from both instruments having strong loadings. Second, that three components would emerge similar to those found in the factoring of the Stanford-Binet IV in each of the two age groups. As in the factoring done on the Stanford-Binet IV in this thesis, principal component analyses were done followed by varimax and quartimax rotations. Although the eigenvalue-one and scree test criteria were used to guide the selection of the correct number of components, the ultimate choice of the number of factors was based on which solution gave the most psychological meaningfulness. Therefore, others may find a different number of components to be more desirable--and they are free to so choose from the results section. For both age groups, the five-component solution seems most appropriate when consideration is jointly given to the scree test and eigenvalue-one criteria.

For the 6-0 to 8-11 year old age group, the five-component quartimax rotated solution suggests that a general component synonymous with *g* exists, and that it accounts for 43.9% of the total subtest variance. Most of the subtests from both the WISC-R and Stanford-Binet IV load on this component; subtests from the Verbal Scale on the WISC-R and from the Verbal Reasoning Area on the Stanford-Binet IV have the highest loadings on *g*. In contrast, the subtests with the lowest loadings on *g* are Picture Completion, Mazes, and Bead Memory. As a rule, the first unrotated component for this age group is very similar to the first quartimax rotated component.

The second component in the five-component quartimax solution is Visual Planning / Ability to See Relationships and this accounts for 8.5% of the total subtest variance. This component has high loadings from Pattern Analysis, Mazes, Block Design, and Absurdities. The third component accounts for 7.7% of the total subtest variance and represents Working Memory / Conceptual Categorization. It has high loadings from Bead Memory, Picture Completion, a moderate loading from W.Comprehension, and a moderate negative loading from Quantitative; of some interest, the two subtests that define this component also have the lowest loadings on *g* . The fourth component is defined primarily by Object Assembly, Block Design, and Digit Span, and appears to be some form of Perceptual Organization component, accounting for 7.2% of the total subtest variance. The fifth component is difficult to interpret in that it is bipolar. One pole is defined by Memory for Sentences and represents Semantic Memory, while the other pole is defined by Coding and represents Non-Meaningful Visual Memory. This fifth component accounts for 5.5% of the total subtest variance.

Looking at the five-component varimax rotated solution for the youngest age group, the component structure is a little clearer. The first component, accounting for 24.0% of the total subtest variance, is Verbal in nature. It has high loadings from Information, Similarities, W.Vocabulary, Picture Arrangement, B.Vocabulary, B.Comprehension, and Memory for Sentences. The second component accounts for 15.5% of the total subtest variance and represents Numerical Ability / Concentration. Subtests with high loadings on this component include Arithmetic, Digit Span, Coding, Copying, and Quantitative. Visual Planning / Ability to See Relationships emerges as the third component, accounting for 12.5% of the total subtest variance. This component is defined by Mazes, Pattern Analysis, Block Design, Absurdities, and B.Comprehension. The next component to emerge is Perceptual Organization, accounting for 11.0% of the total subtest variance. This component has its highest loadings from Block Design, Object Assembly, and Digit Span. The fifth component turns out to be Working Memory / Conceptual Categorization and accounts for 9.6% of the total subtest variance; this component has high loadings from Picture Completion and Bead Memory.

Comparing the quartimax and varimax five-component solutions for the youngest age group, there appears to be much similarity between the second, third, fourth, and fifth components, though they emerge in different orders and the quartimax components tend to be of smaller magnitude. The major difference is that the first component in the quartimax solution represents *g* whereas in the varimax solution it represents Verbal ability. Given that most of the subtests from both the WISC-R and the Stanford-Binet IV load on the same general component (whether the first unrotated principal component or the first quartimax rotated component) it seems safe to conclude that both tests are measuring the same construct of intelligence.

The results from the joint analysis of the Stanford-Binet IV and the WISC-R for the 6-0 to 8-11 year age group suggest that when the two tests are factored together that the factor structure seen on either test individually changes; that is, the addition of more subtest allows more components to be defined. Unfortunately, these components are narrower and more difficult to explain. As such, it is difficult to make suggestions to clinicians regarding which subtests to use as supplemental to administration of only one of the two tests. As a tentative suggestion, more work needs to be done to see how Bead Memory and Picture Completion fit together--especially given that both of these subtests tended to have little in common with the general component. There also appears to be an interesting, though unclear, relationship between Mazes and Pattern Analysis. As a general warning, it is worth noting that Block Design and Pattern Analysis may not be exactly the same; nor are W.Comprehension and B.Comprehension. B.Vocabulary and W.Vocabulary, on the other hand, seem to be nearly identical in terms of factorial composition.

Turning to the the 9-0 to 12-11 age group, the five-component quartimax solution is also a challenge to interpret. The first component clearly represents *g* and accounts for 53.8% of the total subtest variance. All subtests from both the Stanford-Binet IV and the WISC-R load on this component, though Picture Arrangement (.36), Coding (.40), and Mazes (.38) only have weak loadings. As in the factoring of the Stanford-Binet IV for this age group, the general component has its highest loadings from subtests that are from the Verbal scales on the respective tests; conversely, no verbal component emerges when the factoring is done. Of further importance, both Matrices and Number Series have very strong loadings on *g* .

The second component in the quartimax five-component solution represents Auditory Short-Term Memory, and accounts for 6.1% of the total subtest variance. As in the factoring of the Stanford-Binet IV, Memory for Digits and Memory for Sentences load on this factor, and are joined by the Digit Span subtest from the WISC-R. The third component is Visual Planing and accounts for 5.9% of the total subtest variance--this component may have an Abstract / Visual composition though the emphasis appears to be on the Visual element. In an encouraging finding, the subtests that load on this component include Mazes, Copying, Block Design, Object Assembly, and Pattern Analysis. The fourth and fifth components that emerge are bipolar and account for 5.3 and 4.9% of the total subtest variance respectively. The fourth component represents Passive Copying of Numbers (Coding) on one pole and Active Numerical Processing (Quantitative) on the other. The fifth component has one pole of Sequencing of Situations and the other pole of Sequencing of Beads.

Examination of the five-component varimax rotated solution for the 9-0 to 12-11 age group provides a slightly different perspective than the quartimax solution. The first component in the varimax solution is Crystallized Intelligence / Verbal-Educational ability, and accounts for 30.0% of the total subtest variance. This component has high loadings from Information, Similarities, W.Vocabulary, W.Comprehension, Picture Completion, Matrices, B.Vocabulary, B.Comprehension, Absurdities, Quantitative, Number Series, and Arithmetic. The second component represents Fluid Intelligence / Spatial-Mechanical ability, and accounts for 17.3% of the total subtest variance. It has strong loadings from Block Design, Object Assembly, Mazes, Pattern Analysis, Copying, and Bead Memory. The next component to emerge is Memory / Concentration which accounts for 15.7% of the total subtest variance. This component is defined by Information, Arithmetic, Digit Span, Memory for Sentences, and Memory for Digits. Fourth, a Visual Memory component emerges. It accounts for 7.3% of the total subtest variance, and has a high loading from Coding and small loadings from Bead Memory and Memory for Objects. The fifth component is small and basically a singlet with only Picture Arrangement loading on it, though Bead Memory does have a small negative loading. As such, this factor represents Sequencing of Situations.

Comparing the varimax and quartimax five-component solution for the 9-0 to 12-11 age group indicates that the two rotational methods result in different component structures. The varimax solution produces components of Crystallized Intelligence, Fluid Intelligence, Memory / Concentration, Visual Memory, and Sequencing of Situations. In contrast, the quartimax solution produces components of *g* , Auditory Short-Term Memory, Visual Planning, Numerical (bipolar), and Sequencing (bipolar). Worth noting, the component structure that emerges from the joint analysis bears more similarity to the component structure found for the Stanford-Binet IV than for the WISC-R. However, with the larger number of subtests, more components were able to be defined than when the Stanford-Binet IV was factored by itself.

Important findings from the joint factoring of the Stanford-Binet IV and the WISC-R for the middle age group include the finding that nearly all the subtests from both the Stanford-Binet IV and WISC-R load strongly on *g* . This supports the contention that the two tests are measuring the same construct that we call intelligence. In terms of the Areas on the Stanford-Binet IV, the joint analysis partially gives support to Short-Term Memory, but also suggests that this area may need to be broken down into two divisions: one for auditory and the other for visual short-term memory. Though the Quantitative Area receives some support from the emergence of a Numerical component in the quartimax solution, this component is not well defined and hence the inclusion of the Quantitative Reasoning Area on the Stanford-Binet IV at this age seems questionable. Similarly, once the variance accounted for by *g* is removed, the presence of a Verbal component for this age is questionable--though *g* itself does seem to be quite closely related to Verbal ability. Thus, even with the addition of more subtests, the four Reasoning Areas on the Stanford-Binet IV still fail to emerge as group components once the variance attributable to *g* is removed; in the same fashion, removing the *g* variance results in the loss of the typical three-factor solution found for the WISC-R (Kaufman, 1975).

E. Implications For Further Research

The factoring that has been done on the Stanford-Binet IV in this thesis is a beginning point for further research. Given the unclear findings in this thesis, much work still needs to be done on determining the construct validity of the Stanford-Binet IV; though its construct validity in terms of measuring *g* seems undisputable, validation of the theoretical structure of this test has yet to substantiate the division of the test into group factors which Thorndike et al. (1986) refer to as Reasoning Areas.

Though the problem of missing data will continue to be a problem for those factoring the Stanford-Binet IV because of the adaptive testing format, this problem can be handled by using regression estimates to replace missing data (Timm, 1970). However, a more hidden problem has to do with choosing appropriate age groups to base factor analyses on. Researchers are warned that Thorndike et al.'s (1986) procedure of using median correlations for factor analyses has a serious potential problem; namely, that this allows for large discrepancies in the sample sizes for the various correlations (Gorsuch, 1983).

Future research using a truly hierarchical factoring procedure is needed as is more research using exploratory factoring with both varimax and quartimax rotations. Likewise, further factoring using principal factoring and the use of squared multiple correlation as initial communality estimates is needed. This will allow for determination of the specific variance accounted for by each subtest on the Stanford-Binet IV, and can help to suggest which subtests merit interpretation in profile analyses and which do not. Likewise, once a stable factor pattern has been found for the Stanford-Binet IV, it appears that much work will need to be done on developing factor scores and deviation quotients to replace the four Reasoning Areas that to date have received little validation.

Researcher should be careful not to fall into the trap of comparing factor analyses done on the WISC-R using varimax rotations and those done on the Stanford-Binet IV with other rotational methods. What is needed is research that concurrently factors the WISC-R and Stanford-Binet IV on the same sample of children, using the same factoring procedures. Further, much theoretical information can be gained by jointly factoring the WISC-R and Stanford-Binet IV; unlike the joint factoring of the K-ABC and WISC-R (Kaufman & McLean, 1986, 1987) where the typical three-factor WISC-R structure remained intact, it appears that the joint factoring of the WISC-R and Stanford-Binet IV results in a new factor structure with several new factors emerging.

F. Summary

In this study, the primary focus was on validating the Stanford-Binet IV. As a whole, Composite was shown to have good concurrent validity with the WISC-R Full Scale IQ. However, the construct validity of the Stanford-Binet IV was only partially substantiated. Most notably, the emergence of a strong general component of intelligence (*g*) validates the Composite score on the Stanford-Binet IV; the joint analysis of the Stanford-Binet IV and the WISC-R further confirms that these two tests are measuring the same construct of general intelligence. Unfortunately, though the Stanford-Binet IV is validated as a measure of general intelligence, its utility in measuring group factors such as

Verbal Reasoning, Quantitative Reasoning, Abstract / Visual Reasoning, and Short-Term Memory is not clearly validated from the current component analyses. In fact, not only is the division of the Stanford-Binet IV into four Reasoning Areas open to question, but the placement of various subtests on these Areas is also highly suspect. For example, Bead Memory does not appear to measure Short-Term Memory but rather some other Visual / Conceptual skill. Much more research is urgently needed on the factor structure of the Stanford-Binet IV before clinicians can confidently assert that it measures not only general intelligence but also Verbal, Quantitative, Abstract / Visual, and Short-Term Memory abilities.

REFERENCES

Alcamoni, L. M. (1976). The relation of sample size to the number of variables in using factor analysis techniques. *Educational and Psychological Measurement, 36*, 879-883.

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

Anastasi, Anne. (1982). *Psychological Testing*, (5th ed). New York: Macmillan's Publishing Co., Inc..

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

Baggaley, A. R. (1982). Deciding on the ratio of number of subjects to number of variables in factor analysis. *Multivariate Experimental Clinical Research, 6*, 81-85

Barnard, J. (1987). Consideration regarding the use of the new Stanford-Binet. *The Alberta School Psychologist*, 7 (1), 16.

Binet, A., & Simon, T. (1916) The development of intelligence in children (the Binet-Simon Scale) (E. S. Kite, Trans.). Vineland, New Jersey: Publications of the Training School at Vineland New Jersey (No. 11, May, 1916), Department of Research. (Original work published 1905). [pp. 9-181 in translator's work].

Binet, A., & Simon, T (1916). The development of intelligence in the child (E. S. Kite, Trans.). Vineland, New Jersey: Publications of the Training School at Vineland New Jersey (No. 11, May, 1916), Department of Research. (Original work published 1908). [pp. 182-337 in translator's work].

Blaha, J., & Wallbrown, F. H. (1984). Hierarchical analyses of the WISC and WISC-R: Synthesis and clinical implications. *Journal of Clinical Psychology, 40*, 556-571.

Blaha, J., & Vance, H. (1979). The hierarchical factor structure of the WISC-R for learning disabled children. *Learning Disability Quarterly, 2,* 71-75.

Bloom, A. S., Reese, A., Altshuler, L., Meckler, C. L., & Raskin, L. M. (1983). IQ discrepancies between the Binet and WISC-R in children with developmental problems. *Journal of Clinical Psychology, 39,* 600-603.

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

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

Carlson, L., Reynolds, C., & Gutkin, T. B. (1983). Consistency of the factorial validity of the WISC-R for Upper and Lower SES groups. *Journal of School Psychology, 21,* 319-326.

Carroll, J. B. (1978). How shall we study individual differences in cognitive abilities?--Methodological and theoretical perspectives. *Intelligence, 2,* 87-115.

Carroll, J. B. (1980). Remarks on Sternberg's "factor theories of intelligence are all right almost". *Educational Researcher, 9,* 14-18.

Carroll, J. B. (1984). Raymond B. Cattell's contribution to the theory of cognitive abilities. *Multivariate Behavioral Research, 19,* 300-306.

Carroll, J. B. (1985). Exploratory factor analysis: A tutorial. In D. K. Detterman (Ed.), *Current topics in human intelligence : Research Methodology* (Vol. 1) (25-58). Norwood, NJ: Ablex Publishing Corporation.

Carvajal, H., & Weyand, K. (1986). Relationships between scores on Stanford-Binet IV and Wechsler Intelligence Scale For Children-Revised. *Psychological Reports, 59,* 963-966.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton Mifflin Company.

Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement, 15,* 139-164.

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

Chi, M. T. H. (1976). Short-term memory limitations in children: Capacity or processing deficits? *Memory and Cognition, 4* , 559-572.

Clampit, M. K., Adair, J., & Strenio, J. (1983). Frequency of discrepancies between deviation quotients on the WISC-R: A table for clinicians. *Journal of Consulting and Clinical Psychology, 51,* 795-796.

Clampit, M. K., & Silver, S. J. (1986). Four tables for statistical interprestation of factor scores on the Wechsler Intelligence Scale for Children-Revised (WISC-R). *Journal of School Psychology, 24,* 395-404.

Comrey, A. L. (1978). Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology, 46,* 648-659.

Covin, T. M., & Sattler, J. M. (1985). A longitudinal study of the Stanford-Binet and WISC-R with special education students. *Psychology in the Schools, 22,* 274-276.

Cronbach, L. J. (1960). *Essentials of psychological testing* (2nd ed.). New York: Harper & Row, Publishers.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450-466.

Davis, E. E., & Rowland, T. (1974). A replacement for the venerable Stanford-Binet? *Journal of Clinical Psychology, 30,* 517-521.

DeHorn, A., & Klinge, V. (1978). Correlations and factor analysis of the WISC-R and Peabody Picture Vocabulary Tests for an adolescent psychiatric sample. *Journal of Consulting and Clinical Psychology, 46,* 1160-1161.

Detterman, D. K. (1982). Does "g" exist? *Intelligence, 6,* 99-108.

Dinnan, J. A. (1977). Questioning the testing of the intellectual "process". *Education, 97,* 217-279.

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

Eysenck, H. J. (1939). Critical note of 'primary mental abilities' by L. L. Thurstone. *British Journal of Educational Psychology, 9,* 270-275.

Eysenck, H. J. (1979). *The structure and measurement of intelligence.* Berlin: Springer-Verlag.

Fagan, T. (1986). Stanford-Binet draws board attention. *Communique, 15* (3), 1-2.

Fancher, R. E. (1985a) Spearman's original computation of *g* : A model for Burt? *British Journal of Psychology, 76,* 341-352.

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

Flynn, J. R. (1984). IQ gains and the Binet decrements. *Journal of Educational Measurement, 21,* 283-290.

Freeman, F. S. (1955). *Theory and Practice of Pscychological Testing* (revised edition). New York: Henry Holt and Company.

French, L. A. (1986). MR testing and evaluation: New dimensions and old concerns. *Psychology in the Schools, 23,* 64-76.

Friedes, D. (1972). Review of the Stanford-Binet Intelligence Scale, third revision. In O. K. Buros (Ed.) *Seventh Mental Measurements Yearbook.* Highland Park, N. J.: Gryphon Press, 772-773.

Garfinkel, R., & Thorndike, R. L. (1976). Binet item difficulty then and now. *Child Development, 47,* 959-965.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Philadelphia: W.B. Saunders.

Groff, M., & Hubble, L. (1982). WISC-R factor structure of younger and older youth with low IQs. *Journal of Consulting and Clinical Psychology, 50,* 141-142.

Guertin, A. Z., Guertin, W. H., & Ware, W. B. (1981). Distortion as a function of the number of factors rotated under varying levels of common variance and error. *Educational and Psychological Measurement, 41,* 1-9.

Guilford, J. P. (1977). The invariance problem in factor analysis. *Educational and Psychological Measurement, 37,* 11-19.

Gustafsson, J-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8,* 179-203.

Gutkin, T. B. (1978). Some useful statistics for the interpretation of the WISC-R. *Journal of Consulting and Clinical Psychology, 46,* 1561-1563.

Gutkin, T. B. (1979). The WISC-R verbal comprehension, perceptual organization, and freedom from distractibility deviation quotients: Data for practitioners. *Psychology in the Schools, 16,* 359-360.

Gutkin, T. B. (1982). WISC-R deviation quotients vs. traditional IQs: an examination of the standardization sample and some implications for score interpretation. *Journal of Clinical Psychology, 38,* 179-182.

Hakstian, A. R., Rogers, W. T., & Cattell, R. B. (1982). The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research, 17,* 193-219.

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

Hargreaves, D. J. (1974). Psychological testing: Current perspectives and future developments. *Educational Review, 27,* 26-33.

Harman, H. H. (1967). *Modern factor analysis*. Chicago: The University of Chicago Press.

Hertzog, C. (1985) Applications of confirmatory faotor analysis to the study of Intelligence. In D. K. Detterman (Ed.), *Current topics in human intelligence : Research Methodology* (Vol. 1) (pp. 59-97). Norwood, NJ: Ablex Publishing Corporation.

Hill, T. D., Reddon, J. R., & Jackson, D. N. (1985). The factor structure of the Wechsler scales: A brief review. *Clinical Psychology Review, 5,* 287-306.

Hodges, K. (1982). Factor structure of the WISC-R for a psychiatric sample. *Journal of Consulting and Clinical Pscyhology, 50,* 141-142.

Hollinger, C. L., & Kosek, S. (1986). Beyond the use of full scale IQ scores. *Gifted Child Quarterly, 30,* 74-77.

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

Horn, J. L. (1976). Human abilities: A reveiw of research and theory in the early 1970s. *Annual Review of Psychology, 27,* 437-485.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57,* 253-270.

Horn, J. L., & Cattell, R. B. (1982). Whimsy and misunderstandings of Gf-Gc theory: A comment on Guilford. *Psychological Bulletin, 91,* 623-633.

Huttenlocher, J., & Burke, D. (1976). Why does memory span increase with age? *Cognitive Psychology, 8* , 1-31.

Jensen, A. R. (1979). The nature of intelligence and its relation to learning. *Journal of Research and Development in Education, 12,* 79-95.

Jensen, A. R. (1983). The definition of intelligence and factor-score indeterminancy. *Behavioral & Brain Sciences, 6* , 313-315.

Johnston, W. T., & Bolen, L. M. (1984). A comparison of the factor structure of the WISC-R for blacks and whites. *Psychology in the Schools, 21,* 42-44.

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

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

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23,* 187-200.

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

Karnes, F. A., & Brown, K. E. (1980). Factor analysis of the WISC-R for the gifted. *Journal of Educational Psychology, 72,* 197-199.

Kaufman, A. S. (1975). Factor analysis of the WISC-R at 11 age levels between 6 1/2 and 16 1/2 years. *Journal of Consulting and Clinical Psychology, 43,* 135-147.

Kaufman, A. S. (1979). WISC-R research: Implications for interpretation. *School Psychology Digest, 8,* 5-27.

Kaufman, A. S. (1981). The WISC-R and learning disabilities assessment: State of the art. *Journal of Learning Disabilities, 14,* 520-526.

Kaufman, A. S., & McLean, J. E. (1986). K-ABC / WISC-R factor analysis for a learning disabled population. *Journal of Learning Disabilities, 19,* 145-153.

Kaufman, A. S., & McLean, J. E. (1987). Joint factor analysis of the K-ABC and the WISC-R with normal children. *Journal of School Psychology, 25,* 105-118.

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,* 72-78.

Kerlinger, F. N. (1979). *Behavioral research: A conceptual approach.* New York: Holt, Rinehart and Winston.

Kieth, T. Z. (1986). Factor structure of the K-ABC for referred school children. *Psychology in the Schools, 23,* 241-246.

Kieth, T. Z. (1987). Assesment research: An assessment and recommended interventions. *School Psychology Review, 16,* 276-289.

Lee, H. B., & Comrey, A. L. (1979). Distortions in a commonly used factor analytic procedure. *Multivariate Behavioral Research, 14,* 301-321.

Lombard, T. J., & Riedel, R. G. (1978). An analysis of the factor structure of the WISC-R and the effect of color on the coding subtest. *Psychology in the Schools, 15,* 176-179.

Loo, R. (1979). The orthogonal rotation of factors in clinical research: A critical note. *Journal of Clincial Psychology, 35,* 762-765.

Loo, R. (1983). Caveat on sample size in factor analysis. *Perceptual and Motor Skills, 56,* 371-374.

MacCallum, R. (1983). A comparison of factor analysis programs in SPSS, BMDP, and SAS. *Psychometrika, 48,* 223-231.

McCallum, R. S., Karnes, F. A., & Edwards, R. P. (1984). The test of choice for assessment of gifted children: A comparison of the K-ABC, WISC-R, and Stanford-Binet. *Journal of Psychoeducational Assessment, 2,* 57-63.

McDonald, R. P., & Mulaik, S. A. (1979). Determinacy of common factors: A nontechnical review. *Psychological Bulletin, 86,* 297-306.

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

McNemar, Q. (1964). Lost: Our intelligence? Why? *American Psychologist, 19,* 871-882

Merrill, M. A. (1938). The significance of IQs on the Revised Stanford-Binet Scales. *The Journal of Educational Psychology, 29,* 641-651.

Mulaik, S. A. (1972). *The foundations of factor analysis.* New York: McGraw-Hill Book Company.

Mulaik, S. A. (1986). Factor analysis and Psychometrika: Major developments. *Psychometrika, 51,* 23-33.

Naglieri, J. A. (1981). Factor structure of the WISC-R for children identified as learning disabled. *Psychological Reports, 49,* 891-895.

Nicholls, J. G., Patashnick, M., & Mettetal, G. (1986). Conceptions of ability and intelligence. *Child Development, 57,* 636-645.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw Hill Book Company.

Nunnally, J. C. (1970). *Introduction to psychological measurement.*. New York: McGraw-Hill Book Company.

Osberg, D. (1986). New tests: New options. *Communique, 15* (3), 3.

Ownby, R. L., & Matthews, C. G. (1985). On the meaning of the WISC-R third factor: Relations to selected neuropsychological measures. *Journal of Consulting and Clinical Psychology, 53,* 531-534.

Paden, P. A. (1981). Hierarchical structure of abilities: Factorial validation of higher order constructs in Thurstone's primary mental abilities. *Educational and Psychological Measurement, 41,* 803-813.

Petersen, C., R., & Hart, D. H. (1979). Factor structure of the WISC-R for a clinic-referred population and specific subgroups. *Journal of Consulting and Clinical Psychology, 47,* 643-645.

Pollane, L. P., & Schnittjer, C. J. (1977). Principal-axis factor analysis: A comparison of four selected computer programs. *Educational and Psychological Measurement, 37,* 207-211.

Popplestone, J. A., & McPherson, M. W. (1974). The prolonged avoidance of intellectual behavior. *The Psychological Record, 24,* 549-557.

Quattrocchi, M., & Sherrets, S. (1980). WISC-R: The first five years. *Psychology in the Schools, 17,* 297-312.

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

Reynolds, C. R., & Gutkin, T. B. (1980). Stability of the WISC-R factor structure across sex at two age levels. *Journal of Clinical Psychology, 36,* 775-777.

Salvia, J., Ysseldyke, J. E., & Lee, M. (1975). 1972 revision of the Stanford-Binet: A farewell to mental age. *Psychology in the Schools, 12,* 421-422.

Sapp, G. L., Chissom, B., & Graham, E. (1985). Factor analysis of the WISC-R for gifted students: Atreplication and comparison. *Psychological Reports, 57,* 947-951.

Sattler, J. M. (1965). Analysis of functions of the 1960 Stanford-Binet Intelligence Scale, Form L-M. *Journal of Clinical Psychology, 21,* 173-179.

Sattler, J. M. (1974). *Assessment of children's intelligence.* Philadelphia: W. B. Saunders.

Sattler, J. M. (1982). *Assessment of children's intelligence and special abilities,* (2nd ed.). Boston: Allyn and Bacon, Inc.

Schooler, D., L., Beebe, M. C., & Keopke, T. (1978). Factor analysis of WISC-R scores for children identified as learning disabled, educable mentally impaired, and emotionally impaired. *Psychology in the Schools, 15,* 478-485.

Sewell, T., & Manni, J. (1977). Comparison of scores of normal children on the WISC-R and Stanfrod-Binet, Form LM, 1972. *Perceptual and Motor Skills, 45,* 1057-1058.

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

Silverstein, A. B. (1969). Analysis of two item-classification schemes for the Stanford-Binet. *Psychological Reports, 24,* 503-505.

Silverstein, A. B. (1974). Structure-Of-Intellect categories in the Stanford-Binet Performance. *American Journal of Mental Deficiency, 78,* 762-764.

Silverstein, A. B. (1980). Estimating the general factor in the WISC-R. *Psychological Reports, 47,* 1185-1186.

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

Slate, J. (1986). New does not necessarily mean better. *Communique, 15* (1), 2-3.

Sobotka, K., & Black, F. (1978). A procedure for the rapid computation of WISC-R factor scores. *Journal of Clinical Psychology, 34,* 117-119.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201-203.

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

Steiger, J. H. (1979). Factor indeterminacy in the 1930's and the 1970's some interesting parallels. *Psychometrika, 44,* 157-167.

Sternberg, R. J. (1981). A componential theory of intellectual giftedness. *Gifted Child Quarterly, 25,* 86-93.

Sternberg, R. J. (1986). The future of intelligence testing. *Educational Measurement: Issues and Practice, 5,* 19-22.

Sutter, E. G., & Bishop, P. C. (1986). Factor structure of the WISC-R and ITPA for learning-disabled emotionally disturbed, and control children. *Journal of Clinical Psychology, 42,* 975-978.

Swerdlik, M. E., & Schweitzer, J. (1978). A comparison of factor structures of the WISC and WISC-R. *Psychology in the Schools, 15,* 166-172.

Telzrow, C. (1987). New Stanford-Binet and preschoolers. *The Alberta School Psychologist, 7* (1), 22.

Terman, L. W., & Merrill, M. A. (1937). *Measuring intelligence: A guide to the administration of the new revised Stanford-Binet Test of Intelligence.* Cambridge, Massachusetts: The Riverside Press.

Terman, L. W., & Merrill, M. A. (1953) . B. 1937 Stanford-Binet Scales (p. 510-521). In Arthur Weider (ed.) *Contributions toward medical psychology: Theory and psychodiagnostic methods, Vol. II*. New York: The Ronald Press Company.

Terman, L. W., & Merrill, M. A. (1973) *Stanford-Binet Intelligence Scale: Manual for the third revsion form L-M: 1972 Norms Edition*. Boston: Houghton Mifflin Company.

Thompson, J. E. (1984). An empirical versus conceptual model of the Stanford-Binet for interpretive purposes. Unpublished masters degree, University of Alberta, Edmonton.

Thorndike, R. L. (1975). Mr. Binet's test 70 years later. *Educational Researcher, 4* (5), 3-7.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford-Binet Intelligence Scale: Fouth edition*. Chicago: The Riverside Publishing Company.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs,* No. 1. Chicago: University of Chicago Press.

Thurstone, L. L., & Thurstone, T. G. (1941). Factorial studies of intelligence. *Psychometric Monographs,* No. 2. Chicago: University of Chicago Press.

Thurstone, T. G. (1941). Primary mental abilities of children. *Educational and Psychological Measurment, 1,* 105-116.

Timm, N. H. (1970). The estimation of variance-covariance and correlation matrices from incomplete data. *Psychometrika, 35,* 417-437.

Undheim, J. O. (1976). Ability structure in 10-11-year-old children and the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 4,* 411-423.

Undheim, J. O. (1981a). On intelligence I: Broad ability factors in 15-year-old children and Cattell's theory of fluid and crystallized intelligence. *Scandinavian Journal of Psychology, 22,* 171-179.

Undheim, J. O. (1981b). On intelligence II: A neo-Spearman model to replace Cattell's theory of fluid and crystallized intelligence. *Scandinavian Journal of Psychology, 22,* 181-187.

Van Hagen, J., & Kaufman, A. S. (1975). Factor analysis of the WISC-R for a group of mentally retarded children and adolescents. *Journal of Consulting and Clinical Psychology, 43,* 661-667.

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

Vance, H., & Wallbrown, F. H. (1977). Hierarchical factor structure of the WISC-R for referred children and adolescents. *Psychological Reports, 41,* 699-702.

Vance, H. B., & Wallbrown, F. H. (1978). The structure of intelligence for black children: A hierarchical approach. *Psychological Record, 28,* 31-39.

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

Vernon, P. E. (1961). *The structure of human abilities, 2nd Ed.* London: Methuen & Co. Ltd..

Vernon, P. E. (1965). Ability factors and environmental influences. *American Pschologist, 20,* 723-733.

Vernon, P. E. (1978). *Intelligence testing 1928-1978: What next?* Edinburg: The Scottish Council for Research in Education.

Vernon, P. E. (1987). The demise of the Stanford-Binet Scale. *Canadian Psychology, 28,* 251-258.

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

Walkey, F. H. (1983). Simple versus complex factor analyses of responses to multiple scale questionnaires. *Multivariate Behavioral Research, 18,* 401-421.

Wallbrown, F. H., & Blaha, J. (1979). The hierarchical factor structure of the WISC-R for reading disabled children. *Multivariate Experimental Clinical Research, 4*, 73-80.

Wallbrown, F. H., Blaha, J., Wallbrown, J. D., & Engin, W. (1975). The hierarchical factor structure of the Wechsler Intelligence Scale for Children-Revised. *Journal of Psychology, 89*, 223-235.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale For Children-Revised.* New York: The Psychological Corporation.

Whitworth, J. R., & Sutton, D. L. (1982). *Stanford-Binet form L-M compilation: What to do now that you know the score.* Novato, California: Academic Therapy Publications.

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

Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research, 17*, 253-269.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442.

Appendix A
The Essentials of Factor Analysis

Many psychologists and researchers regard factor analysis with a degree of suspicion. Some even suggest that factor analysis has a magical or mystical nature. It is not surprising, however, that many regard factor analysis with such trepidation--throughout history unknown and poorly understood ideas have been given an askance treatment, and often denounced. Although the popularization of factor analysis has progressed significantly, it still has a long way to go in debunking misconceptions about it. As early as 1960 it was obvious that factoring procedures were poorly understood. Consider a comment may by Kaiser (1960) in a lecture that he was delivering: "Factor analysis will eventually come out of the realm of a strange, mystical, ad hoc, half-art, half-science sort of numerology into the camp of reputable methodologies..." (p. 148). However, even as late as 1985 an undercurrent of disregard towards factor analysis was noted by Hill, Reddon, and Jackson (1985): "Yet, some researchers have criticized the use of factor analysis and recommended avoiding its application because of 'subjective' decisions, indeterminacy, ancillary variables, and reification. Topics contributing to this controversy include the validity of a Spearman-like 'g' factor, alternative axis orientations, and the number of factors" (p. 288).

The question that can rightfully be asked is whether factor analysis has indeed come into the camp of reputable methodologies. Tentatively, one could say that factor analysis has indeed become more widely accepted; the availability of high speed computers along with the development and implementation of principal axes factoring and analytic rotation methods like the varimax criterion presented by Kaiser (1958) have resulted in an increasing appearance of factor analytic studies in the psychometric literature. However, it seems that many still regard factor analysis as half-art and half-science. This notion may even have some substance, as many procedures in factor analysis do involve a degree of subjective judgment (e.g., the number of factors to rotate, the rotational method to use, and the naming of factors). Because the results of this thesis rest primarily on factoring procedures, an attempt will be made to unshroud some of the mysteries of factor analysis.

As a starting point, it is wise to begin with an elementary discussion of the purpose of factor analysis. Kerlinger (1979) does an admirable job of describing one of the main purposes of factor analysis:

> *Factor analysis* is an analytic method for determining the number and nature of the variables that underlie larger numbers of variables or measures. It tells the researcher, in effect, what tests or measures belong together--which ones virtually measure the same thing, in other words, and how much they do so. The 'underlying variables' in this definition are called 'factors' (p. 180).

Thus, one of the major objectives of factor analysis is to simplify data by reducing the number of variables or dimensions; to find the smallest number of variables that can account for the common variance of subtests or items; to summarize the interrelationship between variables.

A key point to remember about factor analysis and factors has to do with the difference between "real" and hypothetical constructs. A factor is not a real entity but rather a hypothetical construct--an unobserved variable that is believed to underlie and explain observed measures. This point is elaborated by Kerlinger (1979): "Let us say here that the only 'reality' factors have inheres in their accounting for the variance of observed variables, as revealed through the correlations among the variables" (pp. 180-181). Basically, what this boils down to is the fact that factors cannot be measured directly but must be inferred from the relationship between other measures. Nonetheless, a major contribution of factor analysis to empirical work has been the explication of construct and the resulting gains in theoretical knowledge (Nunnally, 1967).

Because factors are hypothetical or "latent" constructs, it is wise to remember that interpretation of factors may be flawed by one of several reasons. Kerlinger (1979) points out three such possible problems that should be kept in mind: first, a salient factor loading may occur by chance; second, a researcher may mistakenly deduce the nature of the factor from the high loadings; and third, factor analysis may be flawed by one of several technical difficulties (e.g., how many factors to extract). Despite these potential pitfalls, the researcher should not discard factoring procedures as worthless. Rather, he should be aware of potential problems and strive to minimize them.

Given the potential problems related to specific methods used in factor analysis, the researcher should also emphasize the need for replication of results. When the factor structure of a domain consistently emerges as invariant across different samples regardless of the method of factoring used, more weight can be put on the existence [illegible] hypothetical constructs that we call factors. A good example of this type [illegible] lation of knowledge has been the work on the factor structure of the WISC-R. Th[illegible] icular factor structure has proven to be relatively invariant across different samples as well as across different factoring methods (e.g., both orthogonal and oblique rotations). Consequently, when a new domain is being factored, findings should be regarded as tentative; only after repeated verification should firm conclusions be made about the underlying nature of the factor structure. Replication of the factor structure of variables in a given domain, by repeated studies, is essential.

Although factor analysis has been used in several areas of psychology, it has been developed primarily through the work of theorists concerned with the study of intelligence (e.g., Spearman, Thurstone, Cattell, and Burt). Generally, the application of factor analysis to the study of intelligence has had one major focus: the reduction of the number of dimensions needed to adequately explain the variables in the battery. This type of research falls into the category of exploratory factor analysis. The assumption being that the correlations between the various tests are sufficient to elucidate the latent abilities that are tapped by each test. The factoring done in this thesis is also of an exploratory nature; that is, the basic premise is that the correlations between the subtests in the battery will be sufficient to determine the latent factor structure underlying the battery. A simple overview of factor analysis as it applies to the domain of intelligence tests has been provided by Kieth (1987):

> Mathematically, factor analysis is nothing more than a reduction technique, a method of reducing many measures into a few. Psychologically, however, factor analysis helps us understand what it is those many instruments do measure. It is, essentially, an efficient method of determining discriminant and convergent validity; those tests which measure something similar form a factor, whereas those tests which measure something different form separate factors... (p. 281).

One remaining problem deserves some consideration. This has to do with the apparent contradictions that arise when different factoring procedures are used; for example, Thurstone (1938) found several primary abilities but no general factor. In contrast, Spearman (1904, 1923) found only a general factor. What must be realized is the fact that mathematically it is possible to arrive at a large number of "correct" though different solutions even when using the same data (e.g., compare Kaufman's (1975) vs Wallbrown et al.'s (1975) factoring of the WISC-R standardization sample). To a large extent, different results occur because different methods of factoring have been used. Usually, the decision of which method to use is guided by the theoretical orientation of the researcher. Even though one solution may not be technically better than another, some do seem to be more meaningful than others. Ultimately, the preferred method of factoring is one that is both technically correct and yields factors that are psychologically meaningful.

In the following review of factor analysis, several issues pertinent to this thesis will be examined and embedded within an elementary description of the factoring procedure. For example, topics will cover issues such as the impact of sample size and composition on factor structure, methods of handling missing data, rules for deciding the number of factors to rotate, and rotational methods. Though the discussion of these topics is not intended to be complete, it should provide at least a working knowledge of these issues. Those desiring further elaboration can consult one of several texts on factor analysis; Gorsuch's (1983) text, which is often cited in this thesis, is highly recommended in that it does not assume that der is mathematically sophisticated. As applicable to this thesis, emphasis will be place on the principal component model of factoring, and on rotation of factors using the varimax and quartimax criteria. However, other relevant methods that have been used by other researchers will be discussed briefly to allow the reader to make some comparisons between the results found in this thesis and those found by other researchers.

A Few Definitions of Common Terms in Factor Analysis

For the sake of clarity and consensus, a few of the more important terms used in this thesis will be defined. Once again, the purpose is not to provide a detailed account at this time, but to provide a foundation of common terminology that will help to reduce confusion.

Correlation Matrix

The correlation matrix provides the correlations between the variables in the battery. This is the starting point from which factor analysis proceeds. As such, the first step in a factor analysis is the calculation of the correlation matrix, usually using Pearson product-moment correlations. An important point to note is that calculation of correlation coefficients implicitly normalized the data matrix and results in an index that is in standardized form (Gorsuch, 1983). Thus there is no need to bring each variable to a comparative unit before the correlations are calculated. As noted by Gorsuch (1983) "Shifts in the normative scaling of variables will not affect the factor analysis if correlations are computed" (p. 299). Thus, calculation of correlations eliminates the influence caused by different means and variances and results in an index that is largely scale free (Gorsuch, 1983).

Factor Structure & Factor Pattern Matrices

For orthogonal rotations the factor structure and factor pattern matrices are the same (Loo, 1979). However, in oblique rotations the two matrices are different. The pattern matrix gives the weights by which the factors reproduce the variables. Meanwhile, the factor structure matrix gives the correlation between the variables and the factors. This distinction will be clarified in greater detail when oblique rotations are discussed. For the time being, it is worth noting that in orthogonal rotations these two matrices are the same, and therefore are referred to simply as the "factor matrix"--a table of factor loadings that expresses the relationship between the subtests and the underlying factors. This factor matrix simultaneously gives both the factor-variable correlations and the weights by which the factors reproduce the variables.

Factor Loadings

Factor loadings are the correlations between the variables and the factors and ar found in the factor matrix. Like correlation coefficients, factor loadings range from -1.0 to +1.0. These factor loadings express how much a test or observed variable is 'loaded' or 'saturated' on a factor" (Kerlinger, 1979, p. 181). Thus, a factor loading reflects a quantitative relationship between the variable and the factor. The higher the loading, the more one can generalize from the factor to the variable (Gorsuch, 1983). By examining the factor loadings, one can better understand the nature of each factor. Because a factor loading is like a correlation, the square of a factor loading tells the proportion of the variable's variance that is explained by the factor. For example, if a variable has a factor loading of .5 on the second factor, it follows that the second factor accounts for 25 percent of the variance of that variable. For any given variable, the amount of variance accounted for by the factors can be obtained by taking the sum of the squared factor loadings in the row that represents the variable--in orthogonal rotations this is known as the communality of the variable. Similarly, by taking the sum of the squared factor loadings in a column of the factor matrix, the amount of variance accounted for by that factor is obtained; when this is divided by the number of variables in the matrix, it provides the proportion of total variance of the entire group of variables that is accounted for by the factor.

Communality

As has already been mentioned, the communality of a variable is the proportion of its variance that is accounted for by the common factors identified in the data set. That is, what the variable has in common with other variables. The communality of a variable, referred to as h^2, is the sum of the squared factor loadings on the row representing that variable. For example, if a subtest has a *g* loading of .8, and a loading of .5 on a verbal factor, the communality for this subtest would be $(.8)^2 + (.5)^2$ which equals $.64 + .25 = .89$. The communality give the proportion of variance from the variable that is explained by the factors.

Eigenvalue

Also known as the characteristic roots or latent roots. These characteristic roots are equal to the sum of the squared loadings on that principal factor. As such, it is a direct index of how much variance is accounted for by each factor. The size of these characteristic roots is one important criteria for determining how many factors to extract (Gorsuch, 1983).

Salient Loading

A salient loading is one that is sufficiently high to assume that a relationship exists between the variable and factor and that the loading did not happen by chance. Though some consider any loading above .30 to be salient, others use a more intuitive approach and consider only very high loadings a salient. For the purpose of this thesis a salient loading will be consider one which is .50 or greater.

Simplicity and Complexity

The complexity of a variable has to do with how many factors it loads on. A complex variable is one that has significant loadings (non-zero loadings) on more than one factor. In contrast, a simple variable has only one major loading. In terms of interpretability, simple variables are generally preferred to complex ones. The issue of simplicity will be further discussed when rotation of factors is dealt with (i.e., rotation to simple structure).

With some of the major terms used in factor analysis now defined, this review will delve into some of the "mysteries" and complexities that surround factor analysis. Though many issues not included in this review may be of theoretical interest to psychometrician, the emphasis in this review is on issues that are highly pertinent to this thesis.

Impact of Sample Size

No discussion of factor analysis would be complete without some mention of the effect that sample size has on the resulting factor structure. On the one hand, the larger the sample size the better; on the other hand, practicalities often result in sample sizes that are not as large as one would ideally desired. As an absolute minimum, the number of observations should not be smaller than the number of variables (Aleamoni, 1976). Though this criteria may be of interest to mathematicians dealing with a very large number of variables, it is not relevant to most researchers doing applied work. Both Carroll (1985) and Gorsuch (1983) suggests that any sample with less than 100 observations should be viewed cautiously, as the small number of cases may yield artifactual results that are due to chance. The reason for insisting on a relatively large sample size is to prevent capitalization of sampling error--which may lead to errors in interpreting factors. By seeking large samples, the researcher limits the probability that factors and loadings are attributable to chance. Therefore, because the factor loadings are more stable, the researcher can more confidently assume that observed differences in the loadings are real and not just chance variations (Aleamoni, 1976).

Sample size may have some impact on the number of factors that are extractable and readily interpretable. As noted by Carroll (1985), "A rule of thumb is that the sample size, N, needed to support the extraction of m common and interpretable factors must be as least equal to $2m + 2^m$, or preferably much larger if m is small" (pp. 30-31). By this formula, a sample size of 48 would be needed to support the extraction of three factors; a sample size of 128 would be needed to support the extraction of four factors; and a sample size of 320 would be needed to support extraction of five factors. Although this formula may be overly stringent, it does stress the very important point that the more factors that are to be extracted, the larger the sample size should be.

Another fundamental rule is that the number of individuals should be approximately five times the number of variables (Gorsuch, 1983). However, this rule has been challenged by Baggaley (1982). A key point that Baggaley makes is that the number of individuals needed per variable depends on the strength of the intercorrelations between the variables. When the intercorrelations are higher, fewer subjects to variables are needed. He provided a general rule of thumb for determining the number of subjects to the number of variables with the strength of intercorrelation between variables being taken into consideration: "For variables expected to intercorrelate about .30, use at least twice as many subjects as variables. With an expected intercorrelational level of approximately .20, use at least three times as many subjects as variables, and with intercorrelations averaging about .15, the ratio should be at least four to one" (Baggaley, 1982, p. 83).

The above rule of thumb that Baggaley (1982) suggested must be interpreted cautiously when the number of variables is less than 20 (which is the number used by Baggaley). As the number of variables decreases, the number of subjects needed increases somewhat. In terms of this thesis, the expected correlations between the Stanford-Binet IV subtests are mostly above .3; therefore, the ratio of subjects to variables should be about 2 or 3 to one. For the joint factoring of the Stanford-Binet IV and the WISC-R, the number of variables will be around 20 for all three groups; with 20 variables and a mean correlation of .5 between them, Baggaley suggests that the number of subjects should be a minimum of 1.54 times the number of variables. Thus, based on the strength of intercorrelations, the size of the sample needed will vary: the higher the intercorrelations between variables, the fewer subjects to variables that are needed.

Loo (1983), on the other hand, argued that it is not the ratio of variables to subjects that is important, but rather the stability of the correlations. He lists three factors that contribute to the lack of stability of correlations: (a) outliers--the best procedure is to ensure that the observations cover the full range of the values of the variable (or at least the range that is of interest); (b) errors in sampling--this should be minimized, and increased strength of correlation between variables tends to minimize this as can be seen in the formula below; and (c) the heterogeneity of the sample--pooling groups can reduce the stability of the correlations.

$$\sigma_r = 1 - \rho^2 / \sqrt{N-1}$$

Where σ_r is the standard error of the correlation r and ρ^2 is the population correlation squared ($N > 30$)

Thus when the correlation for the variable is strong, the standard error of the correlations is smaller. For example, with a population correlation of .3 and an N of 50, the standard error of the correlation is .13; when the population correlation is .8 and an N of 50, the standard error of the correlation is .05. Therefore, when the population correlations are strong, fewer number of subjects is needed in order to maintain any given standard error of the correlation. So, if the correlations are strong they also tend to be more stable.

Impact of the Nature of the Sample

The nature of the sample can have an impact on the findings that emerge. For example, Thurstone's 1938 sample was a highly select and highly homogeneous group and a general factor did not emerge in the centroid analysis. It is well known that when the range of scores is restricted (homogeneous sample) that lower correlations are found. The opposite is true as well; when the range is extreme (highly heterogeneous) the correlations are much higher than would be expected if a more "normal" range of scores was used.

Comrey (1978) discusses the impact that extreme heterogeneity has on correlations. When a few subjects have rather extreme high or low scores, the effect is to "elongate the scatter plot ellipses, generating very high correlations for those pairs of variables and injecting a great deal of spurious common factor variance into the matrix" (p. 650). Because the sample anticipated for this thesis involves not only gifted children but also those with very low levels of intellectual functioning, this caution mentioned by Comrey is appropriate to note: correlations and common factor variance in this thesis may be somewhat higher than would be found in a more "normal" population. In essence, the use of a clinical population may tend to slightly increase correlations because of the extreme heterogeneity of the population.

The researcher has to strike a balance. On the one hand, performing a factor analysis on homogeneous groups (e.g., groups that have been selected on the basis of high or low IQ) causes a restriction in range and correlations are lowered. On the other hand, factoring done on a group that is extremely heterogeneous can cause elongation of the range and correlations will be inflated. In recognition of this problem, Carroll (1978) recommends the use of a "single group pooled from the several strata, even if the single group is not completely representative of some population because of gaps in the distribution, as where for example, 'high' and 'low' tails of some stratification variable are pooled..." (p. 95).

Although it is desirable to have a full range of scores, Carroll (1985) noted that samples which contain the two tails of a distribution may be appropriate for preliminary evaluation of the data. Carroll (1985) noted that this type of situation may involve a sample that has two groups--one for children with school problems and one for gifted placement: "Sometimes it is convenient to select samples in such a way that they represent high and low portions of a distribution with respect to some ability or abilities, the middle portion being excluded; such samples are suspect to some extent, but acceptable if due account is taken of the possible inflating of correlations and factor loadings" (p. 31). Thus, it is important for the researcher to carefully consider both the range and distribution of scores for the sample used. By doing so, qualifications may have to be placed on the findings if the range is found to be either extremely homogeneous or extremely heterogeneous, or if there are gaps in the range covered by the sample.

Impact of Variables Included and Their Distribution

In designing factorial studies from the ground up, Carroll (1985) suggests that "Each primary factor should be represented by *at least* three variables expected to have high or 'salient' loadings on it and no other factor" (p.29, emphasis in original). Similarly, Zwick and Velicer (1986) indicate that for principal components analysis, components that emerge must contain at least two substantial loadings; conversely, using a principal factoring method (common factor model) a factor should have at least three salient loadings. Thus, for a component to be emerge, at least two variables should be expected to have salient loadings on that component. Therefore, in doing factor analysis, it is essential that each anticipated factor have a minimum of two variables loading on it or else the factor should not be included. Thus, to an extent, the variables included in the analysis will partially determine the factors that emerge.

In choosing variables for inclusion in a factor analysis, Carroll (1985) points out that a basic assumption is that the underlying relationships between variables are linear. Also, it is desirable for the variables to have approximately similar distribution shapes. For ease of interpretability, it is also helpful if variables are scored so that higher levels represent more ability--this makes it easier to evaluate the degree to which "positive manifold" exists (the absence of salient negative loadings). Finally, variables that are combinations of other variables (e.g., A+B or A-B when A and B are already in the analysis) should not be used in the analysis, as they may create artifactual factors (Carroll, 1985).

Handling Missing Data

The problem of missing data can be problematic for those doing factor analytic work. In the factoring of tests like the WISC-R there is little problem with missing data because each person takes the same subtest (e.g., the ten mandatory subtests). However, on a test like the Stanford-Binet IV which uses an adaptive testing format, not all children take the same subtests, depending on their age and ability level. As such, factoring of tests like the Stanford-Binet IV can present serious challenges. The researcher is faced with two possible solutions: (a) eliminate all cases for which variables are not present, or (b) use some type of procedure to estimate or handle the missing data.

The older school of thought has generally favored the exclusion of all cases that do not have scores on each variable to be included in the factor analysis. Carroll (1985) represents the traditional view that individuals with missing data should be dropped from the analysis:

> Sporadic missing data points may not cause significant problems, but in general it is best to eliminate such cases completely unless sample size is thereby drastically reduced. The presence of missing data points can create artifactual results, particularly if the presence or absence of data points is correlated significantly with ability factors. FA studies in which there are substantial missing data problem are to be held suspect; it is my belief that no missing data correlation routine adequately handles the problem (p.32).

Thus, Carroll suggests that missing data, especially if the missing data follows a pattern (as it does on the Stanford-Binet IV), is very problematic for factor analysis.

In contrast to the traditional, hard-line approach, more current thinking acknowledges the necessity for finding adequate methods to handle missing data. Gorsuch (1983) summarizes several such approaches. First, the missing score on a variable can be replaced by the mean of all individuals who took that variable; the effect of this estimation is to leave the mean unaffected, but the variance changes. Second, the score of another individual on the variable can be selected at random and used to replace the missing variable; this leaves both the mean and variance unaffected. In both of these two previous cases, the result is that correlations between variables are lowered. Third, pairwise comparisons can be made; that is, each correlation coefficient is computed using all cases which have complete data for the pair of variables being correlated, even if the case has missing values on any other variable. The correlation matrix resulting when this procedure is used should only be factored if the number of individuals is quite similar for each correlation. When the number of cases for each pairwise correlation varies considerably, Gorsuch warns that "the coefficients may be sufficiently incompatible to prevent a factor analysis or distort the results" (1983, p. 303). Fourth, missing data can be estimated through a multiple regression analysis. The regression analysis is run with only those individuals who have scores on all the variables used in the analysis. The resulting regression equation can then be used in conjunction with existing scores for the individual in order to estimate the missing scores.

Deciding which method to use in estimating missing scores was once done by personal preference. However, Timm (1970) has provided some guidance to this problem. He looked at the effect that different estimation procedures had on the variance and correlation matrices. Timm's study used three correlation levels between variables (.2, .5, and .8), three variable levels (2, 5, and 10), three sample sizes (50, 100, and 200), and three levels of missing data (randomly removing 1%, 10%, or 20% of the total data). The methods for estimating the missing data that he compared were: (a) replacing the missing value with the grand mean, (b) replacing the missing value with the variable mean, (c) replacing the missing value using a regression estimate, and (d) using only cases for which all variables were complete. With 1% of the data missing, the best procedure for maintaining the variance and correlation matrix of the original complete data was the regression estimate procedure--regardless of whether the correlations between variables was low, moderate or high. Of special note is the fact that the regression procedure was also better than using only cases for which all variables existed with no missing data. Even when 10% or 20% of the data was missing, the regression procedure for estimating the missing values was still the best for overall relative efficiency in reproducing the variance and correlation matrix.

From Timm's (1970) study, the important conclusion is that regardless of whether 1%, 10% or 20% of the data is missing, the regression procedure for estimating missing data is one of the best. One thing to remember though, is that when the correlations between variables is high, the regression procedure will tend to increase the correlations slightly; when the correlations are low, the regression procedure will tend to reduce the correlations slightly. The second overall best procedure was that of replacing the missing values with the grand mean; however, this was by far only the second choice. Further, elimination of all cases except those containing all variables tended to reduce the average intercorrelation of variables. Hence, for handling missing data, the regression procedure seems to be the best at maintaining the true variance and correlations.

Principal Axes Factoring

The principal axes factoring method is one of the most popular and most commonly used factoring methods. With the development of high-speed computers the obsolete centroid method has given way to the more sophisticated principal axes factoring method. In fact, the centroid method was really designed by Thurstone as a "computational compromise" to approximate the principal axes method (Gorsuch, 1983; Nunnally, 1967). Because of the popularity of the principal axes factoring method, and because it will be used in this thesis, a review of some of its chief characteristics is in order.

The major feature of the principal axes factoring method is that as much variance as possible is extracted with each successive factor. As such, the maximum amount of variance is extracted for the given number of factors. Of further importance, the reproduced correlation matrix resulting from the use of this method is "the best least-squares estimate of the entire correlation matrix including diagonal elements" (Gorsuch, 1983, p. 96). In a more technical manner, Gorsuch (1983) describes the principal axes method:

> The first factor from the correlation matrix consists of that weighted combination of all the variables that will produce the highest squared correlations between the variables and the factor because the squared correlation is a measure of the variance accounted for. One result is that the sum of the squares of the first column of the factor structure will be maximized.
>
> The second factor is extracted so that it is uncorrelated with the first factor. This maximizes the amount of variance extracted from the residual matrix after the first factor has been removed. Each succeeding factor is extracted in like manner, and so a given number of factors accounts for as much of the variance as that number possibly could (pp. 95-96).

An added feature of the principal axes method is that the the characteristic roots (eigenvalues) are equal to the sum of the squared loadings on the principal factor (any factor). Therefore, the eigenvalues represents the amount of the total subtest variance that is accounted for by each factor (i.e., divide the eigenvalue by the number of variables and this gives the percentage of the total variance accounted for by that factor). This feature makes it easy to calculate the percentage of variance that each factor accounts for.

A clarification that needs to be made about the principal axes factoring method is that both the component and common factor models are applicable. The distinction between these two models as applied to the principal axes method may at first seem trivial, but the ramifications are quite different. The difference between the two models has to do with the diagonal of the correlation matrix. When unities (1s) are placed in the diagonal, the principal components model results; when communality estimates (usually R^2s) are placed in the diagonals, the common factor model results (As a reminder, for the sake of brevity, the application of the common factor model to the principal axes method will be referred to as "principal factoring").

This slight difference between the diagonal elements of the correlation matrices of the principal component and principal factoring method results in differential treatment of the variance of the variables. The principal component method analyzes all the variance from each variable; the principal factoring method analyzes only the common factor variance which is represented by the communality of the variable. As such, the principal factoring method, which proceeds from the common factor model, assumes that the variance from each variable can be divided into two parts: that which is common to other variables, and that which is unique to the variable. The principal component method, taken from the component model, makes no such assumption. Rather, by placing 1s in the diagonal of the correlation matrix, the principal components method "concentrates upon analyzing the variables into a linearly independent set of component variables from which the original variables can be derived" (Mulaik, 1972, p. 174). Put more simply, the total variance of each variable is assumed to be relevant in a compo. ent analysis.

To explicate the difference between the full component model and the common factor model, Gorsuch (1983) has provide two equations that algebraically demonstrate the difference. The component model is defined by the following equation:

$$X_{iv} = w_{v1}F_{1i} + w_{v2}F_{2i} + w_{v3}F_{3i} + ... + w_{vf}F_{fi}$$

where

X_{iv} is individual *i* 's score on variable *v* ; w_{vf} is the weight for variable *v* on factor *f* ; and F_{1i} to F_{fi} are subject *i* 's scores on the *f* factors.

This equation for the component model can be compared to the equation for the common factor model; the two equations are identical except that the equation for the common factor model contains a term for a unique factor which is the "noncommon factor variance for each variable" (Gorsuch, 1983,p. 26). Essentially this means that in the common factor model some of the variance for each variable is attributable to unique influences contributing to only that variable. The common factor model is defined by the following equation:

$$X_{iv} = w_{v1}F_{1i} + w_{v2}F_{2i} + w_{v3}F_{3i} + ... + w_{vf}F_{fi} + w_{vu}U_{iv}$$

where

X_{iv} is individual *i* 's score on variable *v* ; w_{vf} is the weight for variable *v* on factor *f* ; and F_{1i} to F_{fi} are subject *i* 's scores on the *f* factors; w_{vu} is the weight given variable *v* 's unique factor; U_{iv} is individual *i* 's unique factor score for variable *v* .

Principal Components Analysis

A few further points concerning principal components analysis need to be made. First, it is necessary to distinguish between a "full component" and a "truncated component model". Technically, in a component analysis where *n* variables are involved, there will be *n* components that must be extracted in order to account for the total variance of the variables (Gorsuch, 1983; Harman, 1967; Mulaik, 1972). However, when *n* components are extracted, the later components are often very small, are usually not replicable, and account for a trivial amount of the total variance. For this reason, the full component model is rarely used. Instead, the number of components that are retained is less than n ; the smaller components (often those with eigenvalues less than one) are dropped. The remaining components form what is known as a "truncated component solution" (Gorsuch, 1983). Consequently, though the full component model is technically more correct than the truncated component model, the latter is almost always used because of practical considerations.

When the truncated component model is used, the size of the communalities is a good indication of how well this model fits the data. The higher the communalities, the more appropriate the truncated component model is; when the resulting communalities are not very high, the truncated component model is usually not appropriate and the data will not be replicated well. This has led Gorsuch (1983) to the conclusion that "real differences" can be expected between the results of a truncated (or full) component model and a common factor model when some of the communalities are low and the number of variables is less than 20 (p. 124).

In terms of other differences between the principal components and principal factoring methods, it is almost always the case that the factor loadings and communalities from the principal components analysis tend to be slightly higher. Lee and Comrey (1979) found that the factor loadings on important variables were increases an average of .08 (range of .04 to .12) when principal components rather than principal factoring was used. The reason for this has to do with the size of the initial communality estimates. As previously noted, in a principal components analysis the communality for a variable is set to one; for a principal factoring analysis, the initial communality is almost always set to less than one (somewhere near R^2). Because the proportion of total variance extracted equals $\Sigma h^2 / v$ (where v is the number of variables), it follows that the sum of the communalities for the principal factoring method will be smaller. Hence, the percentage of variance accounted for by a component model will be higher than the percentage accounted for by a common factor model (Gorsuch, 1983). The result is that the factor loadings and communalities will be lower in the principal factoring method. As a qualification, if the initial communalities in the principal factoring method are reasonably high, then there will be less difference in the size of the loadings and communalities of the two methods.

Principal Factoring (The Common Factor Model)

The principal factoring procedure has been a popular method for extracting factors from a matrix. As previously mentioned, the difference between the principal components and principal factoring approaches has to to with the question of what values should be put in the diagonal to represent the communality of the variable. However, once the communality estimates have been entered into the diagonal, the extraction procedure for the two approaches is identical (Gorsuch, 1983). It has been this problem of estimating communalities that has provided the most difficulty for the principal factoring method, primarily because there is no way to determine the "true" communality of the variable. In fact, this estimation of communalities is so fundamental to the common factor method that it needs elaboration.

Communality Estimations and Iterations

The communality issue in principal factoring is more complex than most realize. The importance of good communality estimates has been admirably summarized by Gorsuch (1983): "Communality estimation can be quite crucial in the case of a small problem. With less than 10 or 20 variables, the interpretation of factors will often shift if the estimates are distinctively different" (p. 108). To start with, it should be recognized that the correlation of a variable with itself is always equal to 1. Therefore, in a correlation matrix calculated from the data, the diagonals are always 1's. When anything other than 1's is placed in the diagonals, you are no longer correlating an actual variable with a linear combination of actual variables. Consequently, factors that emerge are not 'real' but rather hypothetical--because communalities have been estimated and placed in the diagonals (Gorsuch, 1983). Because there is no way to mathematical solve for communalities, they must be estimated. To further complicate issues, communality estimates will change depending on the number of factors that are extracted.

Several methods for estimating communalities do exist. The most popular method involves initial placement of squared multiple correlations (R^2s) in the diagonal and iteration of these to convergence. Before looking at the R^2 method, two other potential methods will be considered: (a) using unities as initial estimates; and (b) using the square of the largest correlation that the variable has with any other variable (regardless of the sign). If method (a) is used, the principal factoring method becomes the principal component method. The rationale for (b), however, holds some merit. Basically, the square of the highest correlation for the given variable represents the amount of variance between the variable and its most related counterpart, and this is taken to be the lower bound for the communality (Gorsuch, 1983). The problem with this method is that "this procedure underestimates a communality when the highest correlation of one variable with other variables is high and overestimates it when the highest correlation is a low one" (Gorsuch, 1983, p. 104).

In general terms, the lower bound for a communality estimate is the squared multiple correlation of the variable with all other variables in the analysis; the highest bound is usually taken as the variable's reliability. For this reason, the use of squared multiple correlations for initial communality estimates has far surpassed other methods. Although the squared multiple correlation may intuitively seem like the same thing as the communality, there is a distinct difference. The squared multiple correlation gives the percentage of variance that the variable has in common with all the other variables in the matrix; in contrast the communality gives the percentage of variance that the variable has in common with underlying factors (Gorsuch, 1983). Despite this subtle difference, the use of squared multiple correlations (SMC) has some distinct advantages: "If, however, one steadfastly refuses to place unities in the diagonal spaces of the correlation matrix, the use of SMCs is the most sensible approach currently available. The SMCs have the advantages of being (1) unique, (2) directly obtainable on computers, and (3) definitive of at least one type of common variance" (Nunnally, 1967, pp. 353-354).

As discussed earlier, the squared multiple correlations are lower bounds for the communalities, and thus act as a conservative estimate of the communalities. Because they are conservative estimates, the squared multiple correlations are usually iterated to convergence. The procedure for doing so requires knowledge of the appropriate number of factors to extract. As a first step, the R^2s are entered in the diagonal and the appropriate number of factors extracted. Next, the observed communality from this factoring is re-entered into the diagonal and the new correlation matrix is refactored. A new set of observed communalities is then calculated from the second factor analysis and these communalities become new estimates to be entered into the diagonal of the correlation matrix. This iteration process of placing estimates into the diagonal and then finding new estimates on the basis of a complete factor analysis is continued until the maximum change between communalities from one estimate to the next is below some arbitrary level (e.g., .001). Although this iteration procedure seems to work reasonably well, Gorsuch's (1983) warning about it should be noted: "Although the communalities do appear to converge towards appropriate values, **no proof has yet been presented to show that they either must converge or that the value to which they converge is either the theoretical communality or less than the variable's reliability** " (p. 107, emphasis in original).

Occasionally the iteration procedure results in a communality which is greater than one. Such a situation is known as a "Heywood case". Because communalities can not theoretically surpass unity, when Heywood cases exist some corrective measure must be taken. Carroll (1985) suggests that the best way to deal with a Heywood case is to either reduce the number of factors by one, or to drop the variable from the analysis and redo the factoring. Gorsuch (1983) on the other hand suggests that the faulty communality is often set to either .99 or 1.0, even though a more theoretical procedure would be to limit the communality to the reliability coefficient. The occurrence of Heywood cases highlights the problems of communality estimates and iterations. In a harsh statement, Nunnally (1967) suggests that use of an iterative approach "...sometimes leads to the conclusion that the communalities for some variables are greater than 1.0, which certainly casts suspicion on all attempts to obtain communalities through iterative approaches" (p. 353).

Calculating Specific Factor Variance

Because the principal factoring method is derived from the common factor model, it follows that the variance of a variable divides into two parts: the common factor variance as represented by the variables communality, and the unique variance that is uncorrelated with other factors and is related only to that particular variable. This unique variance can be further divided into two parts: the specific factor variance and the error variance.

As has been previously discussed, the communality of a variable (h^2) is the proportion of variance that the variable has in common with the factors; it can be obtained by taking the sum of the squared loadings that the variable has on all the factors. The error variance is easily computed by subtracting the the reliability coefficient from unity (1). The remaining variance which is referred to as the specific factor variance is obtained by subtracting the communality from the reliability coefficient; as this suggests, the specific factor variance is not only reliable but also specific to the particular variable. As was seen in the review of the factorial studies of the WISC-R, the specific variance of a variable has important implications for profile analysis, and should not be totally disregarded. This ability of the common factor model to differentiate between the common factor variance, the specific variance, and the error variance, has been one of its major redeeming features.

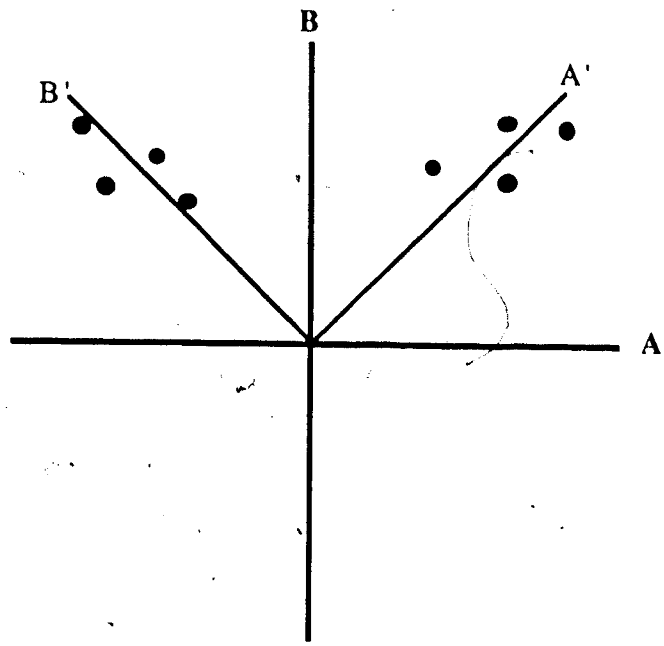
The Indeterminacy Issue

The indeterminacy issue is another complex kick-back to the principal factoring method. Proponents of the common factor model tend to disregard this issue while opponents tend to prey on it. Because of the theoretical complexity, it has been decided that the best way to present the issue is to quote respected factor analysts rather extensively. McDonald and Mulaik (1979) comment on one aspect of the indeterminacy issue: "Factor-score indeterminacy refers to the fact that the common and unique factor scores in the common factor model are not uniquely determined by the observed variables whose correlations they explain, since in general the multiple correlation between a common or unique factor and the observed variables is less than unity" (p. 297). In short, with the common factor model, factor scores do not have unique mathematical solutions. As noted by McDonald and Mulaik, this lack of a unique mathematical definition of factor scores has often led researchers to use some version of a component theory. However, they clearly point out that the problem exist only when one plans on using the factor scores; when the purpose is only to identify and interpret common factors, this problem is not important.

Steiger (1979) reviews the indeterminacy issue as it appeared in the 1930s and also in the 1970s. He notes that many individuals strongly disagree on the importance of factor indeterminacy. Those who believe that it is a major problem often recommend the use of component analysis over the common factor method. The major complaint against factor analysis, as outlined by Steiger (1979) is that factor analysis "does not uniquely define its factors--rather, it identifies a range of random variables which can all be considered factors. This lack of uniqueness raises some interesting dilemmas for those who consider factors to have been 'identified' by a set of factor loadings" (p. 165).

Another aspect of the indeterminacy issue was pointed out by Thurstone (1938). This aspect has to do with rotations and the arbitrary choice of a reference frame for axes. The following diagram adapted from Harman (1967) illustrates his point.

Figure A.1
Graphical Representation of the Indeterminacy Problem



The original axes, A and B, are entirely arbitrary, and can be rotated so that they are in position A' and B'. From a mathematical standpoint, the location of these axes is totally arbitrarily. Thus, this aspect of the indeterminacy issue results from an infinite number of possible reference axes that are equally defensible mathematically (Harman, 1967).

Hill, Reddon, and Jackson (1985) summarize both of the forementioned aspects of the indeterminacy issue: "Position of reference axes are indeterminate because any data set has an infinite number of different orientations for the basis vectors. Indeterminacy also arises due to the problem in specifying the exact length (or communality) of each test vector. Since this length is contingent upon the defining of all the common factors, and this is not possible in practice, the true value is also said to be indeterminant" (p. 297).

In summary, for some individuals the factor indeterminacy problem has remained a fatal flaw in the common factor model; others have disregarded it almost totally. As acknowledged by Mulaik (1986), the factor indeterminacy issue does create some ambiguity in the interpretation of factors, but it does not render the common factor model impotent. Rather, it is a limitation that researchers need to be aware of.

Comparison of Principal Components and Principal Factoring Methods

Choice of the principal components model over the principal factoring model is usually made for one of a few reasons. First, the principal components model is simpler to understand and potentially easier to use. Second, the indeterminacy issue has been a dispute that has made many "would be factor analysts" wary of the common factor model that underlies that principal factoring method. Third, when factor scores are desired, the use of the full component model will enable factor scores to be completely determined whereas the principal factoring method only allows factor scores to be estimated (Velicer, Peacock, & Jackson, 1982).

Although principal components analysis is not truly "factor analysis" in that it does not belong to the common factor method, it does fall under the general "rubric" of factor analysis. In fact, as noted by Mulaik (1972), many factor analysts use component analysis as an approximation to common-factor analysis and "...in return obtain results that often do not differ greatly from those they would have gotten by using a common-factor-analysis model" (p. 174). Furthermore, because principal components and principal factoring both share the principal axes methodology, they are quite similar.

Additional empirical support for the use of principal components analysis as an alternative to principal factoring has been provided by Velicer et al. (1982). In their study they compared the efficiency of principal components, image components, and maximum likelihood factor analysis. To compare these three methods, Velicer et al. constructed a series of contrived matrices each with 36 variables and 6 latent factors. These matrices differed in the degree of complexity and magnitude of loadings. Velicer et al. suggested that "the three methods produce results that, for practical purposes are indistinguishable" (p. 385). Thus, as long as the correct number of factors is known, the three methods are generally equal though the component analyses do produce slightly higher loadings.

Velicer et al.'s (1982) final conclusion was that the common factor method is not superior to the component analysis method; that previous findings suggesting this were problematic in that they were confounded with the number of factors problem in component analysis. As long as the correct number of factors is extracted, the two models produce very similar results.

On the other hand, the principal factoring method is not without some advantages: (a) there are few studies where the variables are thought to be error-free and where all of the variances of the variables can be expected to be predicted from the factors; (b) use of truncated component models results in solutions where the communalities are less than one just like in the principal factoring method; and (c) "a common factor analysis is more likely to result in an actual component analysis when that is the best model than vice versa" (Gorsuch, 1983, p. 124). However, despite these advantages, principal component analyses will probably continue to be used as an alternate to principal factoring because the former is generally simpler, and quicker. For example, if the eigenvalue-one criteria is used to determine the number-of-factors to extract and rotate in the principal factoring method, a principal components analysis must first be done and then a principal factoring solution must be done using the correct number of factors. Thus, by doing only a principal component analysis both computer time and expense can be reduced.

The Principle of Factor Rotation

As was seen in the previous section, the choice of a reference frame or axis for the factor contributes to the indeterminacy issue. The choice of where to locate the axes for factors is one that has proven to be crucial to factor analysis. Although the choice of a reference axis for a factor may be arbitrary from a mathematical standpoint, it is not so from a psychological point of view. When a factor analysis is conducted, the first unrotated principal axes factor usually represents a general factor; however, the remaining unrotated factors are often difficult to interpret as they are often bipolar. By using an analytic rotation method such as the varimax criterion, the bipolar nature of the factors is usually eliminated, and the factors are easier to interpret. Mulaik (1986) has underlined the importance of factor rotations: "In fact, solving of the rotation problem became a kind of Holy Grail for factor analysts to pursue, and many a factor analyst was to make his reputation with a workable analytic scheme of factor rotation" (p. 26). As such, factor rotation is the process of infusing meaning into an otherwise sterile, though correct, mathematical factor matrix. The goal is to meaningfully define the position of the factors in "n-dimensional factor space" (Eysenck, 1979, p. 39).

Because of the pioneering nature of Thurstone's work and his quest for simple structure, his ideas relating to factor rotation has been very influential. The concept of simple structure has been one that has intrigued many factor analysts. Basically the goal is to find the "simplest" matrix which faithfully reproduces the variables; put another way, the ideal goal is to find the factor structure where each variable has only one salient loading and a maximum number of near-zero or insignificant loadings; by so doing, the maximum amount of variance from each variable is confined to a single factor (Vernon, 1961). Unfortunately, although simple structure is a useful concept, it often turns out that variables are complex rather than simple--that is, they have substantial loadings on more than one factor. Hence, it is sometimes very difficult to attain an appropriate analytic rotation when attempting to maintain simple structure (Comrey, 1978).

Deciding on the Number of Factors to Extract and Rotate

A major problem that must be faced concurrently with rotation of factors is deciding on the number of factors to extract and rotate. The number of factors to extract and rotated is very important in terms of reproduction of the data and replicability: "The accuracy with which the data matrix can be reproduced from the factor pattern and factor scores is primarily a function of the number of factors extracted" (Gorsuch, 1983, p. 143). Several points that relate to factor rotations need to be considered before looking at specific methods of determining the number of factors to rotate. In terms of a basic understanding, it should be realized that extraction of as many factors as variables will result in near perfect reproduction of the initial correlations. Unfortunately, when this is done, there are often many "trivial" factors and the factor structure can often be distorted. Thus, the question of how many factors to extract hinges on finding a limited number of factors that contain a maximum amount of information and also that can be interpreted in a meaningful way.

An often unforgotten consideration that indirectly effects the number of factors to extract has to do with sample size. Because the standard error of correlations is related to sample size, larger samples usually have clearer indications of the number of factor; with small samples, sampling errors make it more difficult to clearly identify the number of factors. As a redeeming feature, when communalities are high, the problem of small sample sizes is less of a problem because the random correlations will be given less weight. This relationship between communalities, standard error of correlations, and determination of the number of factors to extract has been elaborated by Gorsuch: "**Therefore, high communality situations can be expected to contain less error not included in the model and the number of factors may be easier to estimate accurately**" (1983, p. 147, emphasis in original). Consequently, as long as communalities are high, the effect of small sample size on determining the number of factors will not be a major handicap.

From a statistical standpoint, the number of factors to extract from a matrix has to do with the size of the residuals after each factor has been extracted. Gorsuch (1983) summarizes the importance of the residual matrix:

> In the statistical procedure for determining the number of factors to extract, the following question is asked: Is the residual matrix after f factors have been extracted statistically significant? If there is statistically significant variance left, then at least one more factor could be extracted with some assurance that non-chance variance is being processed. If there is no significant variance, then the correct number of factors has already been extracted (p. 148).

Although procedures like Bartlett's test have been developed to look at the significance of the residuals, they are not used extensively because they tend to be too dependent on sample size and have not proven to be very accurate (Gorsuch, 1983). However, the concept and importance of the residual matrix are a basic underlying assumption in all number-of-factors solutions; that is, they all assume that when the correct number of factors has been extracted that the residual matrix does not contain a significant amount of non-chance variance.

Although there are several different methods for determining the "correct" number of factors to extract from a matrix, some have been more popular than others. The reason for this popularity stems from three apparent reasons: (a) ease of computation, (b) simple conceptual rationale for the method, and (c) availability on computer programs. By far the most common method involves examination of the contribution that each factor makes in accounting for the total variance of the variables; this is known as the eigenvalue-one criteria. Another popular method is the scree test, which in some ways is related to the eigenvalue-one method though the emphasis is on the distribution of the eigenvalues rather than some minimum cut-off point. Together, these two methods have historically been the most frequently used. However, with the growing sophistication of factor analysis, and with recognition of the importance of the number-of-factors problem, several complex methods have been developed for determining how many factors to extract. Unfortunately, their complexity and unavailability have limited their usage, even though they have demonstrated their potential superiority over both the eigenvalue-one and scree test criteria. Although only the eigenvalue-one and scree test criteria will be used in this thesis, a brief description of other methods is given for those who intend to do further work in factor analysis.

The Eigenvalue-One Criterion

This method is know by several names such as ""The Kaiser-Guttman Unity-eigenvalue Rule" or the "Eigenvalue-one Criterion", and by the "eigenvalues greater than or equal to one" criterion (Carroll, 1985). For simplicity, this method will be referred to as the eigenvalue-one criterion. The name for this rule tells much about its basic premise; the number of common factors to be extracted is taken as the number of principal components that have eigenvalues of one (unity) or greater. Because each variable adds an eigenvalue of 1.0 to the total communality of the entire matrix, any factor that is to be retained should have an eigenvalue of 1.0 or greater; that is, the variance accounted for by the factor should contribute at least as much as the effect of adding one variable. Therefore, the number of factors to rotates is taken as the number of unrotated principal components with eigenvalues that are one or greater (See Comrey, 1978). As pointed out by Guttman (1954) the number of eigenvalues that are greater than one constitutes a 'weak lower bound' on the number of factors; that is, this gives the smallest number of factors that can adequately explain the variance.

Any researcher using the eigenvalue-one method needs to be aware of some basic facts. To start with, the eigenvalue-one method is a rule of thumb, not a mathematical criteria. Because of this, several complications can arise:

(a) as Gorsuch (1983) has pointed out, the eigenvalue-one method works best when there are less than 40 variables, and where the number of factors is expected to be between 1/5 and 1/3 the number of variables (see also Zwick & Velicer, 1986); thus, the number of factors is somewhat a function of the number of variables.

(b) Carroll (1985) has indicated that when variables have substantial correlations this method tends to give too few factors, or conversely when the correlations between variables are low, it gives too many factors.

(c) Comrey (1978) advises that this method be used only when the variables are reliable, continuous measures, strongly intercorrelated, and have high communalities; when these conditions are met he suggest that "the distortion introduced by the eigenvalue-one procedure is less likely to represent a problem" (p. 652).

(d) Hakstian, Rogers, and Cattell (1982) found that when the factor loadings were complex, the eigenvalue-one criterion tends to overestimate the number of factors about half of the time; or conversely when few variables were used and the number of factors was high, the tendency was to underestimate the number of factors.

In summary, the eigenvalue-one method is most applicable when the number of expected factors is 1/3 to 1/5 the number of variables, when less than 40 variables are included, and when variables are reliable and with moderately strong intercorrelations and high communalities. Even then,the eigenvalue-one method only gives an approximate number of factors; it may either overestimate or underestimate the number of factors (Gorsuch, 1983; Zwick & Velicer, 1986).

The Scree Test

Like the eigenvalue-one method, the scree test is based on the size of the consecutive eigenvalues. However, whereas the number of factors to extract according to the eigenvalue-one method is the number of unrotated principal components with eigenvalues ≥ 1, the scree test criterion suggests that the number of factors can be determined by the "break" in the scree line. The scree test comes from Cattell's work, and the basic concept of scree refers to the rubble at the bottom of a cliff. "A basic rationale for the scree test is that the battery of variables is measuring a limited number of factors well and a larger number of trivial, specific, and error factors much less well" (Gorsuch, 1983, p. 166). To use the scree test, the eigenvalues from the successive principal components are plotted (i.e, eigenvalues on the X axes and successive components on the Y axes). Once the eigenvalues are plotted, they are examined to find the point at which the line through the plotted eigenvalues has an elbow (also known as the "break" in the line); that is, the point where there is a substantial change in the slope of the scree line. The basic conception is that the eigenvalues for the correct number of factors will form a relatively straight line with a steep slope while the eigenvalues for the "error" factors will form another more-or-less straight line with a much smaller slope (which represents the 'scree' or rubble at the foot of the mountain). The number of common factors is taken as the number of eigenvalues above the line formed by the scree.

The problem with the scree test is that often two or more elbows or breaks are present on the scree line. Because of the possibility of more than one breaking point, and deciding where lines should be drawn through the eigenvalues (especially the low ones), a degree of subjectivity enters into the scree test. Hakstian et al. (1982) discuss the proper use of the scree test as they conceive of it:

> Some researchers interpret the rule to mean simply finding the first large break in the plot of eigenvalues of **R**, proceeding from the largest downward, whereas others (most notably Cattell) proceed by seeking areas in the plot in which the eigenvalues describe a steady, relatively flat descent, generally beginning from the low end of the plot and working upwards (p. 194).

As with the eigenvalue-one criteria, there are some specific circumstances under which the scree test works best. According to Gorsuch (1983), this method is best when the average communalities are greater than or equal to .6, and when the ration of factors to variables is 1/3. Zwick and Velicer (1986) concur, suggesting that the scree test is best when there are strong components and when the sample size is fairly large. Hakstian et al. (1982) found that in their contrived matrices, the scree test identified the correct number of common factors in all cases where the structure of the matrix was relatively simple; however, when minor common factors existed and communalities were low, the scree test overestimated the number of factors.

Overall, the scree test seems to perform well. When errors exist, it appears that there is a slight tendency to overestimate the number of factors. However, this slight overestimation at times does not appear to be a major problem : "Nevertheless, even given its variability and tendency toward overestimation, the scree test seldom led to the retention of more than two components over the criterion value" (Zwick & Velicer, 1986, p. 440). More importantly, Zwick and Velicer (1982) found that overall the scree test was better than the eigenvalue-one, Bartlett's Test, and the Minimum Average Partial methods. As such, the scree test has been described as having "moderate overall reliability" (Zwick & Velicer, 1986, p. 440)

Using the Eigenvalue-One and Scree Test Together

Hakstian et al. (1982) found that as long as the communalities are high, the factor structure relatively simple, the number of variables not above 30 or so, and the ratio of factors to variables low (around or below 1/3), then both the scree test and the eigenvalue-one criteria were reasonably good. Further, they implied that by using the scree test and the eigenvalue-one criteria jointly, one can adjust for the nuances that each method has. To use these two methods together, Hakstian et al. suggest that the true number of factors probably lies somewhere between the number of factors that each method suggests. For example, if the eigenvalue-one criterion suggests 5 factors, and the scree test suggests 7 factors, then the proper number of factors is probably somewhere between 5 and 7. By using both of these criteria together, more faith can be put in the number of factors to extract.

Bartlett's Test of Significance

Bartlett's test of significance examines the residual matrix to determine if additional factors need to be extracted. Once the remaining eigenvalues are of equal value (the null hypothesis) no more factors are extracted. To determine significance levels, Bartlett's test uses chi-squares to determine whether the residual matrix needs to have a further factor extracted. To determine this, the chi-square of the residual matrix is subtracted from the chi-square of the matrix immediately before the last factor was extracted; the degrees of freedom for this resultant chi-square are taken to be the difference between the two degrees of freedom associated with the original chi-squares (Gorsuch, 1983).

Bartlett's Test is not frequently used for several reasons. First, it is not available on most statistical packages. Second, it gives only an upper bound to the number of factors that "might be significant" (Gorsuch, 1983). Third, it often yields too many factors in that it finds both common and specific factors to be significant (Gorsuch, 1983). Given these problems, it is not surprising that Zwick and Velicer (1986) found the accuracy of Bartlett's Test to be highly variable; in fact, they state that they could not endorse it as a method for determining the number of factors to retain.

The Parallel Analysis Criterion

In the Parallel Analysis criterion, the eigenvalues obtained from the principal axes method (usually principal factoring) are compared to the eigenvalues estimated as most likely to occur in a similar analysis of correlations for random data for the given sample size. The number of common factors is taken to be the number of factors from the principal axes factoring that have eigenvalues greater in magnitude than those for the corresponding random data (Carroll, 1985). Zwick and Velicer (1986) found the Parallel Analysis method to be more accurate than the eigenvalue-one, the scree test, and the Minimum Average Partial methods. Despite showing much promise, this rule needs to be evaluated with caution and is probably best if used in conjunction with other criteria. The major drawbacks to using this method include lack of programs on most statistical packages that will run this method, and the need to generate a large set of random correlation matrices by which to judge the parallel analysis (Zwick & Velicer, 1986). Because of these two drawbacks, the Parallel Analysis criterion has not been used very often to date.

Minimum Average Partial Method

Zwick and Velicer (1982, 1986) report a procedure called the Minimum Average Partial Method. In this method, after each of the first *m* components is partialled out, the average of the squared partial correlations is calculated. When the average of the partial correlations reaches a minimum, the number of components partialled out is the number of components to retain. Zwick and Velicer (1986) found this procedure to one of the best at identifying the number of factors to extract. Unfortunately this procedure requires FORTRAN subroutines and is not available on most statistical packages. Further, its complex nature makes its use foreboding.

Rotating Several Number-of-Factors Solutions Around Those Suggested

Because of problems inherent in most methods of solving the number-of-factors problem, several researchers (Carroll, 1985; Comrey, 1978; Gorsuch, 1983) suggest that one or more of the above criteria be used to identify the number of factors to extract. Once this has been done, the researcher then rotates various numbers of factors around those suggested (e.g., if three factors are suggested by the eigenvalue-one criteria, the researcher may rotate two-, three-, and four-factor solutions and examine each to see which appears to be most appropriate). Comrey (1978) provides an excellent rationale for rotating several different number-of-factors solutions: "There is no defin[illegible]e solution to the problem of determining the correct number of factors. It is up to the [illegible]ator to use all of the information that he can get to reach a conclusion about th[illegible] [illegible]ry to justify his conclusion within several converging lines of evidence instea[illegible] [illegible]lying on a universal rule of thumb" (p. 652). When using this method, only factors with at least two salient loadings should be considered; any factor with less than two salient loadings is considered trivial or insufficiently defined, and therefore one less factor would be suggested (Carroll, 1985; Gorsuch, 1983).

Percentage of Total Variance Accounted For

As we have already seen, when using the eigenvalue-one method, the percentage of variance accounted for by the factor must be equal to or greater than the percentage of variance that a single variable contributes (an eigenvalue of 1.0). An older method for determining how many factors to extract was based on examination of the the cumulative amount of variance that was accounted for by the factors. Typically factoring was stopped after 75, 80 or 85% of the variance was accounted for. "Usually factor extraction is stopped after a large portion of the variance has been extracted and when the next factor would add only a very small percentage to the total variance extracted" (Gorsuch, 1983, p. 165). With more sensitive methods available, this method is no longer practiced by itself; it is, though, sometimes used in conjunction with other methods to decide how many factors to extract.

Number of Salient Loadings, Iterations For Convergence, Over- and Under-Rotation

When using any method of determining the number of factors to extract, a few things must be taken into consideration. Three general tips should always be in mind. The first, as pointed out by Zwick and Velicer (1986) deals with the number of salient loadings on each factor. To accept any principal component, it must have at least two substantial loadings; for principal factoring each factor must contain at least three significant (nonzero) loadings. To distinguish between factors based on the number of salient loadings Zwick and Velicer (1986) have designated components (or alternately factors) as either major, minor, or trivial. (A major component has an eigenvalue ≥ 1 and three or more substantial loadings; a minor component has an eigenvalue ≥ 1, with less than three substantial loadings, or conversely three substantial loadings but an eigenvalue < 1; a trivial components has an eigenvalues < 1, and less than three substantial loadings--according to Zwick and Velicer the latter should never be retained). Second, when an excessive number of iterations (> 50) is required to achieve convergence either in communalities or rotations, then overfactoring is likely to have occurred (Carroll, 1985). Third and finally, over- and under-rotation of factors has some consequences. Several researchers agree (Comrey, 1978; Gorsuch, 1983; Guertin, Guertin, & Ware, 1981) that the extraction of one or two extra factors is less harmful than extraction of one or two too few factors. When too few factors are extracted, the variance from the unextracted factors will be erroneously distributed onto the extracted factors and will distort their "true" appearance. If only one or two extra factors are extracted, these will drain off some of the variance from the true common factors, but usually they do not seriously distort the factor structure of the true factors; however, if an excessive number of factors are extracted, the emergent factors will be highly specific and consequently a loss of the true common factors will occur.

Taking the Hypothesized Number of Factors into Account

Walkey (1983) recommends that in determining the number factors to extract, one carefully consider the interpretability of factors; further, he suggests that it is a good idea to check the hypothesized structure of the matrix as outlined by the developer of the test. For example, in studying the Stanford-Binet IV, one should at least consider extracting the four Areas (Verbal, Abstract/Visual, Quantitative, & Short-Term Memory) as suggested by Thorndike et al. (1986). The appropriateness of this number of factors can be checked against more "objective" criteria by seeing how well the hypothesized number of factors, the objective number of factors, and the test constructor's factors match.

Summary of the Number-of-Factors Problem

To reiterate and summarize, the number-of-factors problem is perhaps one of the most important aspects of factor analysis. Zwick and Velicer (1986) stress the importance of the decision about how many factors to retain:

> The determination of the number of components or factors to retain is likely to be the most important decision a researcher will make. Decisions involving choice of method, type of rotation, and type of score will have relatively less impact because of demonstrated robustness of results across different alternatives in these areas (p. 432).

For all methods of determining the number of factors to extract, the basic principle is that extraction stops when further extraction would not account for further "meaningful" variance. Unfortunately, the determination of "meaningful" variance is not clearly defined. Several methods exist for determining the number of factors to retain and each has some merits as well as disadvantages. By far, the eigenvalue-one criteria has been the most extensively used. The scree test, also working with eigenvalues, has proven to be another popular method of determining the number of factors. Several new approaches like the Parallel Analysis criterion, and the Minimum Average Partial method hold promise for the future, but complex rationales and limited availability hamper their widespread use. The consensus seems to be leaning toward the use of several different criteria at the same time, and then finding the most meaningful number of factors from those suggested (e.g., Gorsuch, 1983; Hakstian et al., 1982).

Orthogonality Versus Obliqueness

Orthogonal and oblique rotations form two broad categories dealing with whether factors are or are not correlated; all rotation of factor axes are either orthogonal or oblique. Orthogonal factors are ones that are not correlated; when plotted on a Cartesian coordinate, the angle between the two factors is 90° which indicates that the two factors are independent. From a geometric viewpoint, when the variables are plotted on these axes, a resulting vector for each variable is formed by connecting the coordinate for the variables to the origin. The length of this vector (when squared) gives the communality for the variable. Oblique factors, on the other hand, are correlated factors. That is, when plotted on a Cartesian coordinate the angle between the factors is less than 90°. Unlike in orthogonal rotations, the placement of one factor axes in the oblique method does not place a theoretical limitation on the second factor axes (though in practice the correlations between factors is not usually allowed to go much above .5 which represents an angle of 30° between the factor axes) (Gorsuch, 1983).

Orthogonal Rotations

The use of orthogonal rather than oblique rotations is approximately three times more common in the literature (Loo, 1979). The main reason for this is that working with orthogonal factors is "far simpler" than working with oblique factors (Nunnally, 1967). A point to remember about orthogonal rotations is that the general factor is often "rotated away" and is represented by small positive loadings in all of the factors (Gustafsson, 1984). Thus, Gustafsson suggests that "It may thus be claimed that orthogonal rotations to simple structure are quite deceptive in the presence of a general factor" (1984, p. 200). Because the variance from the general factor may be spread over the common factors in an orthogonally rotated matrix, the factor loadings on the average tend to be slightly higher than for oblique rotations using the same data.

Varimax Criterion

The varimax method of rotating factors is probably the most common of all factor rotations and is generally considered one of the best orthogonal rotations (Gorsuch, 1983). This method was proposed by Kaiser (1958) as an "analytic criterion" for rotation in that it imposes mathematical conditions such that the "factor matrix is uniquely determined" (p. 187). As a further feature, the varimax criterion does approximate simple structure (Hill et al., 1985; Kaiser, 1958) in that it attempts to minimize the complexity of variables. The varimax criterion is designed to maximize the variance of squared loadings of a factor; that is, it maximizes the variance of the squared factor loadings for the column that represents the factor.

Kaiser's (1958) emphasis on maximizing the squared loadings on the columns rather than the rows was a major break from the older quartimax criterion which sought to maximize the squared loadings across rows. As noted by Kaiser, the quartimax criterion tended to yield a general factor because it emphasized rows; in contrast, the varimax criteria usually does not yield a general factor because it emphasizes rows. Consequently, Gorsuch (1983) warns against the use of the varimax criterion if a general factor is expected: "Maximizing the varimax function means that any tendency toward a general factor in the solution will be minimized. **Varimax is inappropriate if the theoretical expectation suggests a general factor may occur**" (p. 185, emphasis in original). Hill et al. (1985) further clarify the ramifications of spreading the variance from the general factor onto the common factors:

> But the efficacy of varimax as a simple structure criterion is not necessarily support for the practice of distributing the variance contained in the general factor among a number of group factors. The result of this procedure is the confounding of a main effect with interactions since the first unrotated factor or principal component may be interpreted as a main effect due to individual differences, and the remaining unrotated dimensions may be interpreted as interactions between tests and people... (p. 297).

Thus, the varimax criterion has been extremely popular and useful in that it maximizes the variance of a given factor and approximates simple structure. However, those who use this type of rotation need to be cognizant that it disperses the variance from the general factor onto the group / common factors. Therefore, if a general factor is theoretically expected, it may be useful to at least consider using another rotational method alongside of the varimax criterion in order to allow the general factor to emerge. Then the factor structure from the varimax method can be compared to the factor structure from the other method.

Quartimax Criterion

As has already been briefly discussed, the quartimax criterion attempts to maximize the variance of the squared factor loadings for each variable (rows in the factor matrix). The origin of the quartimax criterion can be traced to several theoreticians who arrived at similar solutions independently (see Harman, 1967). In order to maximize the inequalities among the squares of the factor loadings, fourth powers of those loadings are used (Harman, 1967). Because of this, the quartimax solution tends produce a general factor and hence has not been widely accepted according to Gorsuch (1983): "A quartimax solution tends to include one factor with all major loadings and no other major loadings in the rest of the matrix, or have the moderate loadings all retained on the same factor" (p. 184). Put another way, the quartimax method attempts to maximize the large factor loadings and decrease the small ones for each variable in the original factor matrix (Harman, 1967).

Both Nunnally (1967) and Harman (1967) suggests that the quartimax criterion does approximate simple structure though the small values for the factor loadings may not be as close to zero as one would ideally want in simple structure. Further, because it tends to produce a general factor along with one or more group factors, the simple structure is lost to some degree in that a variable will tend to load on both the general factor and also one of the group factors. Despite this apparent departure from simple structure, the quartimax method does appear to have some usefulness. It appears to be relatively similar to the British factoring method used by Vernon (1961) where the general factor is extracted and then the residual matrix factored for group factors.

Oblique Rotations

Oblique solutions are ones in which the angles between factors are no longer orthogonal; however, as noted by Harman (1967), "the conditions set for for an oblique solution do not preclude zero correlations among the factors" (p. 314). Thus, when factors are truly orthogonal, they may emerge so when most oblique rotational methods are used. An added advantage of using oblique rotations is that they may come closer to simple structure (Undheim, 1977). As such, Loo (1979) suggests that if variables load on more than one factor after a varimax rotation that an oblique rotation be tried as it may provide better results in terms of simple structure.

As mentioned previously, working with oblique rotations is more complex than working with orthogonal rotations. Whereas the pattern and structure matrices are identical when orthogonal rotations are used, the two matrices are different when oblique rotations are used. Therefore, when oblique rotations are use, both the pattern and structure matrices must be examined as each provides slightly different information.

The factor pattern matrix provides factor coefficients which are like regression coefficients which can be used for reproducing the original variables (Mulaik, 1972). Or as Mulaik (1972) suggests, "The coefficients of a factor-pattern matrix are weights to be assigned to the common factors in deriving the observed variables as linear combinations of the common and unique factors. In many respects the factor-pattern coefficients are like regression coefficients in predicting a criterion variable from some predictor variables" (p. 01). Similarly, the pattern matrix is used to reproduce the correlations between the variables in order to determine the adequacy of the solution (Harman, 1967).

The factor structure matrix, in contrast, gives the correlations between the variables and the factors; this is useful in identifying the factors (Harman, 1967). Thus, the structure matrix is useful in determining the variance of each variable as jointly accounted for by a particular factor and the interaction effects of that factor with the others (Loo, 1979). As such, the factor structure matrix is the matrix that is usually interpreted in the identification and naming of factors.

Another consideration when using oblique rotations is that the correlations between the oblique factors can be subjected to higher-order factoring in order to find the general factor (Cattell, 1971; Gustafsson, 1984). However, as Gustafsson (1984) warns, the true correlation between factors is usually underestimated because there are always small positive loadings scattered in the matrix. Therefore, the general factor may be slightly underestimated when an oblique method is used.

A further complication from using oblique rotations is that communalities cannot be calculated in a straight-forward way: "Oblique rotations change some of the essential characteristics of the original matrix of factor loadings. The sum of squared loadings in any row would equal h^2 only by chance, and the sum of average squared loadings in the columns of the matrix would equal V only by chance" (Nunnally, 1967, pp. 325-326). Hence, to calculate the communality of oblique factors, one must take into consideration not only the direct contribution of factors, but also the joint contribution of the factors (see Harman, 1967, p. 284 for a formula to calculate h^2 for oblique factors).

A final warning about interpreting oblique rotations is provided by Nunnally (1967). He suggests that interpretation of the oblique factors "...can, and often does, fool the investigator into thinking his data are simpler than they actually are. In most problems, even an approximation to simple structure by this method forces negative correlations among the factor axes..." (p. 331). In interpreting the factors, it is difficult to simultaneously attempt to identify the factors as well as remember the multiplicity of correlations with other factors, especially if they are negative.

At this point, enough warnings and background information on oblique rotations have been given in order to briefly consider a few relevant oblique rotational methods. Coverage is not intended to be complete, but rather elementary; anyone desiring more information should consult an appropriate source such as Gorsuch (1983) or Mulaik (1972).

Oblique Procrustes Rotation

Procrustes rotation is usually used in a confirmatory sense. Thus, a target matrix is specified, and the existing correlations are rotated towards that [illegible] matrix. Use of this technique involves forced rotation of the principal components or factors so as to approximate the hypothesized factor structure as represented by the target matrix. The major problem involves the selection of an appropriate target matrix. It is generally believed that when a hypothesized target matrix is used that the "Procrustean rotations can lead to artifactually confirming a wide range of theories" (Carroll, 1985, p. 47).

Promax Rotation

According to Carroll (1985) the promax method is basically a special form of procrustean rotation. A target matrix is obtained by raising the entries of the varimax matrix to a certain integer power k and preserving the signs of the entries. The purpose of raising the loadings to the k power (usually 4 or 6) is to increase the discrepancy between large and small loadings. By varying the power, both the degree of obliquity and the simplicity of structure can be varied. Gorsuch (1983) states "The orthogonal solution is used as a basis for creating an ideal oblique solution. The unrotated factor matrix is then rotated to the best least-squares fit to the ideal solution by the Procrustes procedure..." (p. 190). In summary, from the varimax solution and ideal matrix is constructed where high loadings from the varimax are increase and low loadings decreased by taking them to the fourth or sixth power.

Direct Oblimin

Whereas most oblique methods use an indirect manner for getting the primary factor pattern, the direct oblimin approach uses a procedure that gets oblimin like solutions directly (Harman, 1967). That is, this method proceeds directly from an initial to a primary-factor pattern. The goal is to make the primary factor pattern approximate simple structure, while limiting the factor intercorrelations which have tended to be objectionably high in some of the indirect oblimin methods (Gorsuch, 1983).

Summary of Orthogonal and Oblique Rotations

The choice of whether to use orthogonal or oblique rotations is left to the researcher. Most often, orthogonal rotations have been used because they tend to be simpler to understand and interpret. This probably accounts for the predominance that has been given to the varimax criterion. It is readily available and easily interpreted. However, as pointed out by Loo (1979) the use of oblique rotations may provide a better fit to simple structure, and thus may be valuable. Especially in hierarchical analyses, the oblique methods of rotation can provide important theoretical information--though complexity increases and interpretability may suffer to some degree.

Confirmatory Factor Analysis

Historically, psychology has relied on exploratory factor analysis. That is, the data is allowed to "speak for themself"--to determine the latent factors and their structure. Carroll (1985) goes on to say that exploratory factor analysis has two advantages over confirmatory factor analysis: "(a) it does not call for the advance specification of any parameters or statistical hypotheses to be tested, and is thus unbiased by any such specifications (which are sometimes difficult to make in confirmatory FA); and (b) its computations are generally simpler and less expensive" (p. 26). Nevertheless, other researchers have duly noted the increasing use of confirmatory factor analysis.

Mulaik (1986) notes that in the late 1960s and early 1970s factor analysts began to to realize that confirmatory factor analysis would be useful in testing hypotheses. Implementation of the desire to test hypotheses was done through the procrustean method:

> Nevertheless the urge to do confirmatory studies remained, giving rise to another avenue of attack, the method of procrustean transformations. Here the researcher sought transformations of the unrotated factors that would

> produce factors within the common factor space that were most like some anticipated hypothetical factors, defined by prespecified factor loadings (Mulaik, 1972, p. 28).

The link between the procrustean transformations and confirmatory factor analysis has been given by Mulaik (1972) : "The term *procrustean* is applied to any harsh or inflexible attempt to force someone or something to fit some preconceived idea or system. The term owes its existence to Procrustes, a highway bandit of ancient Greek mythology who tied his victims to an iron bed and stretched or cut off their legs to make them fit its length" (p. 293). Thus, as outline by Mulaik (1972), a procrustean transformation attempts to fit a given matrix to a preconceived structure using certain restraints. As such, it is clearly a version of confirmatory factor analysis.

In confirmatory factor analysis, the researcher a priori formulates a model of the structure of latent factors and the variables that relate to each. Then a factor analysis is conducted to test the hypotheses that a prespecified number of variables define the anticipated factors (Gorsuch, 1983). As pointed out by Hertzog (1985) "...one uses theoretical conceptions regarding the nature of the constructs and their relations to observed variables to specify (a priori) a model predicting the factor structure of measures" (p. 60). As such, rotations are not needed as the confirmatory analysis gives the solution directly (Gorsuch, 1983).

Although confirmatory factor analysis can substantiate a hypothetical model underlying the observed factor structure, it can not "prove" the model to be true. As Hertzog (1985) notes, all that confirmatory factor analysis can do is empirically verify consistency across samples:

> The term *confirmation* derives in part from a rather different connotation. Before we seriously entertain a factor model as being a useful representation of the factor structure of a set of tests, we wish to demonstrate that the empirical solution is replicable across different samples from the same

> population. However, assume that we replicate the results. We have confirmed that the model produces consistent patterns of parameter estimates, but we have not shown that the theory which produced the model is correct. In short, we must recognize that confirmation in this limited sense refers to empirical consistency and not to the validity of the theory which produced the model (p. 65).

Despite the fact that confirmatory factor analysis cannot "prove" a theory to be correct, it is apparent that it is useful in testing hypotheses and models to see how well they fit the data. The trade-off is not unlike Procrustes' treatment of legs; confirmatory factor analysis may find a good fit by mutilating the data to fit a preconceive notion. A worthwhile suggestion to this effect has been made by Gorsuch (1983). He recommends that exploratory factor analysis be used in areas where no prior analyses have been conducted. After a factor structure has been demonstrated in a given realm, the use of confirmatory factor analysis may be more appropriate in terms of theoretical gains.

Kieth (1987) is a staunch supporter of confirmatory factor analysis. He suggests that confirmatory factor analysis "...provides a much stronger test of the underlying structure of a scale than does exploratory factor analysis, and is especially useful for measures such as the K-ABC and the new Binet which are based on an explicitly underlying theory" (p. 282). He goes on to say that because the researcher can specify the number of factors and their relationship as well as which variables will load on given factors, there is less subjectivity in confirmatory as opposed to exploratory factor analysis.

Unfortunately, confirmatory factor analysis remains a complex and expensive methodology and one that may discourage many researchers. It requires a high degree of sophistication and a strong theoretical commitment to some model that is presumed to underlie the data. Therefore, it has potential usefulness, but should be limited to areas that are well researched through factor analysis, and where the variables included in the analysis have proven in the past to be "marker" variables for the hypothesized factors.

APPENDIX A.1

Number of Subjects Taking Each Subtest As a Function of Age Using Three Age Groupings: 6-0 to 8-11; 9-0 to 12-11; and 13-0 to 16-11

6-0 to 8-11 Years of Age [a]			9-0 to 12-11 Years of Age [b]			13-0 to 16-11 Years of Age [c]		
Subtest	Number Taking Subtest	(%)	Subtest	Number Taking Subtest	(%)	Subtest	Number Taking Subtest	(%)
Vocabulary	46	100%	Vocabulary	66	100%	Vocabulary	56	100%
Comprehension	45	98%	Comprehension	64	97%	Comprehension	55	98%
Absurdities	46	100%	Absurdities	63	95%	Absurdities	44	79%
Verbal Relations	0	0%	Verbal Relations	4	6%	Verbal Relations	23	41%
Pattern Analysis	46	100%	Pattern Analysis	66	100%	Pattern Analysis	56	100%
Copying	45	98%	Copying	52	79%	Copying	31	55%
Matrices	25	54%	Matrices	59	89%	Matrices	50	89%
Paper Folding	0	0%	Paper Folding	5	8%	Paper Folding	29	52%
Quantitative	46	100%	Quantitative	66	100%	Quantitative	56	100%
Number Series	19	41%	Number Series	55	83%	Number Series	43	77%
Equation Building	0	0%	Equation Building	2	3%	Equation Building	18	32%
Bead Memory	46	100%	Bead Memory	66	100%	Bead Memory	56	100%
Memory For Sentences	45	98%	Memory For Sentences	65	98%	Memory For Sentences	56	100%
Memory For Digits	30	65%	Memory For Digits	63	95%	Memory For Digits	53	95%
Memory For Objects	28	61%	Memory For Objects	60	91%	Memory For Objects	53	95%

[a] Sample size for 6-0 to 8-11 year-olds is n=46
[b] Sample size for 9-0 to 12-11 year-olds is n=66
[c] Sample size for 13-0 to 16-11 year-olds is n=56

APPENDIX A.1.1

Number of Cases For Pairwise Comparisons of Subtests For The Age Groups 6-0 Through 8-11Years[a]

	Vocab	Comp	Absurd	Pat Anal	Copying	Matrices	Quant	Num Sers	Bead Mem	MFS	MFD	MFO
Vocabulary	46	45	46	46	45	25	46	19	46	45	30	28
Comprehension	45	45	45	45	44	24	45	19	45	45	29	28
Absurdities	46	45	46	46	45	25	46	19	46	45	30	28
Pattern Analysis	46	45	46	46	45	25	46	19	46	45	30	28
Copying	45	44	45	45	45	24	45	19	45	44	29	27
Matrices	25	24	25	25	24	25	25	19	25	24	25	24
Quantitative	46	45	46	46	45	25	46	19	46	45	30	28
Number Series	19	19	19	19	19	19	19	19	19	19	19	19
Bead Memory	46	45	46	46	45	25	46	19	46	45	30	28
Mem For Sentences	45	45	45	45	44	24	45	19	45	45	29	28
Memory For Digit	30	29	30	30	29	25	30	19	30	29	30	27
Memory For Objects	28	28	28	28	27	24	28	19	28	28	27	28

a Total **n** for this group is 46; minimum number of pairwise cases using twelve variables (excluding Verbal Relations, Paper Folding, and Equation Building) is 19; when only eight variables are used (exluding Verbal Relations, Paper Folding, Equation Building, Matrices, Number Series, Memory For Digits, and Memory For Objects) the minimum number of pairwise cases is 44 (which gives a missing data percentage of 3/368 = .8%)

APPENDIX A.1.2

Number of Cases For Pairwise Comparisons of Subtests For The Age Groups 9-0 Through 12-11Years[a]

	Vocab	Comp	Absurd	Pat Anal	Copying	Matrices	Quant	Num Sers	Bead Mem	MFS	MFD	MFO
Vocabulary	66	64	63	66	52	59	66	55	66	65	63	60
Comprehension	64	64	61	64	51	57	64	54	64	64	62	60
Absurdities	63	61	63	63	51	56	63	52	63	62	60	58
Pattern Analysis	66	64	63	66	52	59	66	55	66	65	63	60
Copying	52	51	51	52	52	45	52	43	52	52	49	48
Matrices	59	57	56	59	45	59	59	53	59	58	57	54
Quantitative	66	64	63	66	52	59	66	55	66	65	63	60
Number Series	55	54	52	55	43	53	55	55	55	54	55	51
Bead Memory	66	64	63	66	52	59	66	55	66	65	63	60
Mem For Sentences	65	64	62	65	52	58	65	54	65	65	62	60
Memory For Digit	63	62	60	63	49	57	63	55	63	62	63	59
Memory For Objects	60	60	58	60	48	54	60	51	60	60	59	60

[a] Total **n** for this group is 66; minimum number of pairwise cases using twelve variables (excluding Verbal Relations, Paper Folding, and Equation Building) is 43 (which gives a missing data percentage of 47/792 = 5.9%) with the mean number of pairwise cases being 59

APPENDIX A.1.3

Number of Cases For Pairwise Comparisons of Subtests For The Age Groups 13-0 Through 16-11 Years[a]

	Vocab	Comp	Absurd	Pat Anal	Copying	Matrices	Quant	Num Sers	Bead Mem	MFS	MFD	MFO
Vocabulary	56	55	44	56	31	50	56	43	56	56	53	53
Comprehension	55	55	43	55	31	49	55	42	55	55	52	52
Absurdities	44	43	44	44	31	38	44	34	44	44	43	43
Pattern Analysis	56	55	44	56	31	50	56	43	56	56	53	53
Copying	31	31	31	31	31	26	31	22	31	31	30	30
Matrices	50	49	38	50	26	50	50	43	50	50	47	47
Quantitative	56	55	44	56	31	50	56	43	56	56	53	53
Number Series	43	42	34	43	22	43	43	43	43	43	42	43
Bead Memory	56	55	44	56	31	50	56	43	56	56	53	53
Mem For Sentences	56	55	44	56	31	50	56	43	56	56	53	53
Memory For Digit	53	52	43	53	30	47	53	42	53	53	53	52
Memory For Objects	53	52	43	53	30	47	53	43	53	53	52	53

a Total **n** for this group is 56; minimum number of pairwise cases using all twelve variables (excluding Verbal Relations, Paper Folding, and Equation Building) is 22; when only ten variables are used (exluding Verbal Relations, Paper Folding, Equation Building, Copying, and Absurdities) the minimum number of pairwise cases is 42 (which gives a missing data percentage of 26/560 = 4.6%) with the mean number of pairwise cases being 51

APPENDIX A.2

Number of Subjects Taking Each Subtest As a Function of Age: Using Thorndike, Hagen, and Sattler's (1986) Age Groupings*

6-0 to 11-11 Years of Age [a]			12-0 to 16-11 Years of Age [b]		
Subtest	Number Taking Subtest	(%)	Subtest	Number Taking Subtest	(%)
Vocabulary	96	100%	Vocabulary	72	100%
Comprehension	93	97%	Comprehension	71	99%
Absurdities	94	98%	Absurdities	59	82%
Verbal Relations	1	1%	Verbal Relations	26	36%
Pattern Analysis	96	100%	Pattern Analysis	72	100%
Copying	85	89%	Copying	43	60%
Matrices	71	74%	Matrices	63	88%
Paper Folding	2	2%	Paper Folding	32	44%
Quantitative	96	100%	Quantitative	72	100%
Number Series	61	64%	Number Series	56	78%
Equation Building	0	0%	Equation Building	19	26%
Bead Memory	96	100%	Bead Memory	72	100%
Memory For Sentences	94	98%	Memory For Sentences	72	100%
Memory For Digits	78	81%	Memory For Digits	68	94%
Memory For Objects	74	77%	Memory For Objects	67	93%

* These figures are presented to demonstrate the effect of using only two age groups; these figures are not used in this thesis.

[a] The sample size for this group would have been 96.

[b] The sample size for this group would have been 72.

APPENDIX A.2.1

Number of Cases For Pairwise Comparisons of Subtests Using Only Two Age Groups* (6-0 to 11-11; and 12-0 to 16-11 years)

Age 6-0 to 11-11 Years[a]

	Vocab	Comp	Absurd	Pat Anal	Copying	Matrices	Quant	Num Sers	Bead Mem	MFS	MFD	MFO
Vocabulary	96	93	94	96	85	71	96	61	96	94	78	74
Comprehension	93	93	91	93	83	68	93	60	93	93	76	74
Absurdities	94	91	94	94	84	69	94	59	94	92	76	73
Pattern Analysis	96	93	94	96	85	71	96	61	96	94	78	74
Copying	85	83	84	85	85	60	85	53	85	84	67	64
Matrices	71	68	69	71	60	71	71	60	71	69	69	66
Quantitative	96	93	94	96	85	71	96	61	96	94	78	74
Number Series	61	60	59	61	53	60	61	61	61	60	61	58
Bead Memory	96	93	94	96	85	71	96	61	96	94	78	74
Mem For Sentences	94	93	92	94	84	69	94	60	94	94	76	74
Memory For Digit	78	76	76	78	67	69	78	61	78	76	78	72
Memory For Objects	74	74	73	74	64	66	74	58	74	74	72	74

a Total **n** for this group would be 96; minimum number of pairwise cases is 53; mean size for pairwise comparisons is 78.9

* *This table is provided* ***only*** *for sake of comparison to Thorndike et al. (1986)--this grouping is not used in this thesis.*

APPENDIX A.2.2

Number of Cases For Pairwise Comparisons of Subtests Using Only Two Age Groups* (6-0 to 11-11; and 12-0 to 16-11 years)

Age 12-0 to 16-11 Years[a]

	Vocab	Comp	Absurd	Pat Anal	Copying	Matrices	Quant	Num Sers	Bead Mem	MFS	MFD	MFO
Vocabulary	72	71	59	72	43	63	72	56	72	72	68	67
Comprehension	71	71	58	71	43	62	71	55	71	71	67	66
Absurdities	59	58	59	59	43	50	59	46	59	59	57	56
Pattern Analysis	72	71	59	72	43	63	72	56	72	72	68	67
Copying	43	43	43	43	43	35	43	31	43	43	41	41
Matrices	63	62	50	63	35	63	63	55	63	63	60	59
Quantitative	72	71	59	72	43	63	72	56	72	72	68	67
Number Series	56	55	46	56	31	55	56	56	56	56	55	55
Bead Memory	72	71	59	72	43	63	72	56	72	72	68	67
Mem For Sentences	72	71	59	72	43	63	72	56	72	72	68	67
Memory For Digit	68	67	57	68	41	60	68	55	68	68	68	66
Memory For Objects	67	66	56	67	41	59	67	55	67	67	66	67

a Total **n** for this group would be 72; minimum number of pairwise cases is 31; mean size for pairwise comparisons is 60.4

* *This table is provided* ***only*** *for sake of comparison to Thorndike et al. (1986)--this grouping is not used in this thesis.*

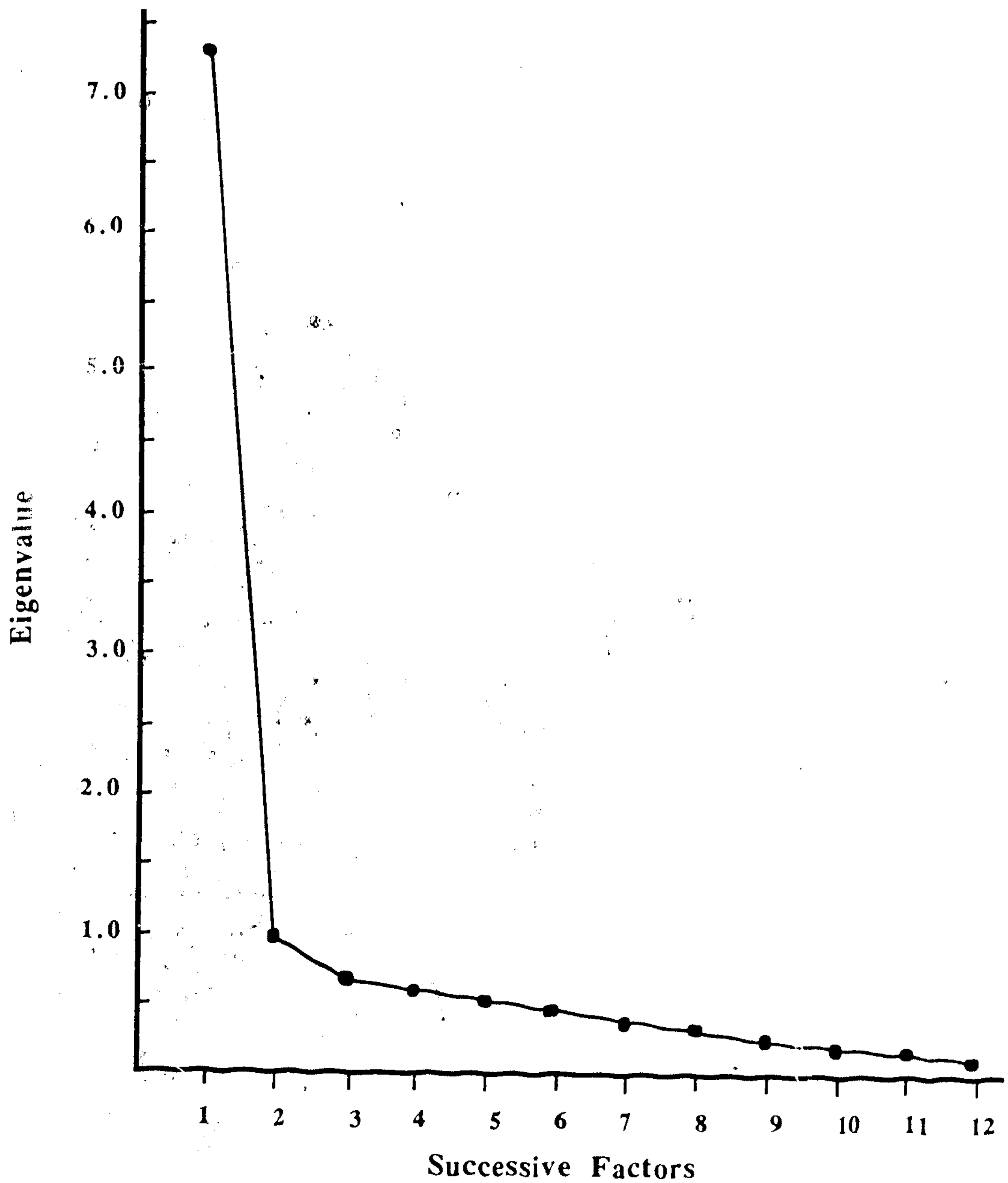
Appendix B.1

First Unrotated Principal Component For The WISC-R: Combined Groups, Ages 6-0 to 16-11 Years (Missing Data Handled Through Pairwise Correlations; n=168)

WISC-R Tests	Unrotated First Component	Component	Eigenvalue	% of Total Variance
	Combined Age Groups (6-0 to 16-11 Years)			
Information	.87	I	7.32	61.0
Similarities	.86	II	0.95	7.9
Arithmetic	.84	III	0.66	5.5
Vocabulary	.86	IV	0.59	4.9
Comprehension	.85	V	0.52	4.3
Digit Span	.78	VI	0.45	3.8
Picture Completion	.70	VII	0.36	3.0
Picture Arrangement	.72	VIII	0.32	2.7
Block Design	.83	IX	0.27	2.2
Object Assembly	.77	X	0.23	1.9
Coding	.64	XI	0.19	1.5
Mazes	.60	XII	0.14	1.2

Appendix B.2

Scree Plot For the WISC-R When All Ages are Combined Into a Single Group: Ages 6-0 to 16-11 Years (Missing Data Handled Pairwise; n=168)



Appendix B.3

Two-, Three-, and Four-Component Varimax Rotated Principal Components For The WISC-R: Combined Groups, Ages 6-0 to 16-11 Years (Missing Data Handled Through Pairwise Correlations; n=168)

WISC-R Combined Groups Age 6-0 to 16-11	Varimax Rotated Principal Components											
	Two Components Extracted			Three Components Extracted				Four Components Extracted				
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Information	.84	.33	.81	.83	.32	.20	.83	.84	.20	.27	.16	.85
Similarities	.78	.38	.76	.81	.37	.11	.80	.78	.38	.20	.09	.81
Arithmetic	.77	.36	.73	.74	.34	.26	.73	.74	.23	.27	.23	.74
Vocabulary	.86	.28	.81	.84	.26	.23	.82	.81	.35	.10	.21	.83
Comprehension	.80	.34	.76	.81	.33	.15	.79	.78	.37	.16	.14	.79
Digit Span	.78	.26	.68	.67	.22	.46	.71	.72	.02	.28	.41	.77
Picture Completion	.39	.65	.57	.40	.64	.09	.58	.29	.83	.20	.13	.84
Picture Arrangement	.42	.64	.58	.39	.63	.19	.58	.34	.57	.37	.20	.62
Block Design	.57	.63	.72	.56	.61	.19	.7[illegible]	.58	.31	.55	.16	.75
Object Assembly	.42	.73	.71	.37	.71	.26	.71	.36	.46	.55	.26	.71
Coding	.57	.29	.42	.28	.24	.89	.93	.29	.22	.13	.90	.95
Mazes	.13	.83	.71	.12	.83	.11	.71	.18	.19	.89	.07	.86
Sum of Squared Loadings	5.07	3.18		4.56	3.01	1.36		4.43	1.90	1.89	1.28	
% of Total Variance	42.3	26.5		38.0	25.1	11.3		36.9	15.8	15.8	10.7	
Cummulative % Var.	42.3	68.8		38.0	63.1	74.4		36.9	52.7	68.5	79.2	

Appendix C.1

Unrotated Principal Components For The Stanford-Binet IV: Ages 6-0 to 8-11 Years (Missing Data Handled Through Pairwise Correlations; n=46)

Stanford-Binet IV Age 6-0 to 8-11	Unrotated Principal Components							
Tests	I	II	III	IV	V	VI	VII	VIII
Vocabulary	.85	-.03	-.32	-.19	-.06	-.18	-.07	-.32
Comprehension	.81	.14	-.16	-.08	-.36	-.29	.15	.24
Absurdities	.78	-.17	.20	.35	.03	-.10	-.44	.07
Pattern Analysis	.72	.13	.40	-.07	-.39	.37	.05	.08
Copying	.67	-.16	.12	-.59	.36	.08	-.06	.12
Quantitative	.63	-.55	.21	.29	.20	-.07	.35	-.04
Bead Memory	.37	.81	.26	.12	.31	-.14	.09	-.05
Memory for Sentences	.68	.15	-.54	.25	.17	.35	.03	.07
Eigenvalue	3.96	1.07	0.74	0.68	0.58	0.42	0.35	0.19
% of Total Variance	49.5	13.4	9.3	8.5	7.3	5.2	4.4	2.4
Cummulative % Var.	49.5	62.9	72.2	80.7	88.0	93.2	97.6	100.0

Appendix C.2

Two-, Three-, and Four-Component Varimax Rotated Principal Components For The Stanford-Binet IV; Ages 6-0 to 8-11 Years (Missing Data Handled Through Pairwise Correlations; n=46)

Stanford-Binet IV Age 6-0 to 8-11	Varimax Rotated Principal Components											
	Two Components Extracted			Three Components Extracted				Four Components Extracted				
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Vocabulary	.80	.28	.72	.45	.78	.09	.82	.70	.25	.54	.06	.86
Comprehension	.70	.43	.68	.42	.66	.29	.70	.61	.27	.43	.28	.71
Absurdities	.79	.13	.63	.73	.33	.17	.67	.33	.77	.18	.23	.79
Pattern Analysis	.62	.39	.54	.65	.18	.49	.70	.12	.46	.48	.49	.70
Copying	.69	.09	.48	.62	.32	.11	.50	.17	.19	.88	.03	.85
Quantitative	.79	-.28	.70	.82	.17	-.20	.75	.17	.86	.20	-.14	.83
Bead Memory	.05	.89	.79	.00	.16	.91	.86	.17	.03	.03	.92	.88
Memory for Sentences	.58	.39	.49	.13	.86	.10	.78	.88	.20	.03	.14	.84
Sum of Squared Loadings	3.58	1.46		2.41	2.08	1.25		1.85	1.76	1.56	1.26	
% of Total Variance	44.8	18.3		30.1	26.0	15.6		23.1	22.0	19.5	15.8	
Cummulative % Var.	44.8	63.1		30.1	56.1	71.7		23.1	45.1	64.6	80.4	

Appendix C.3

Unrotated Principal Components For The Stanford-Binet IV: Ages 9-0 to 12-11 Years (Missing Data Handled Through Pairwise Correlations; n=66)

Stanford-Binet IV Age 9-0 to 12-11	Unrotated Principal Components											
Tests	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Vocabulary	.85	-.29	-.05	-.19	-.14	.00	.04	.08	.19	-.21	.07	-.18
Comprehension	.80	-.30	-.20	-.19	.14	-.18	.31	-.11	-.02	.05	.13	.14
Absurdities	.89	.02	-.20	.00	.02	-.19	.12	.05	-.25	.07	-.19	-.14
Pattern Analysis	.78	.31	-.01	-.04	-.25	.37	.03	-.15	-.23	-.06	.13	-.01
Copying	.67	.46	.07	.44	-.03	-.31	.04	.14	.02	-.13	.09	.04
Matrices	.83	.00	-.12	-.23	.09	-.13	-.42	.11	-.05	.14	.13	.00
Quantitative	.72	-.13	-.34	.42	.09	.32	.05	.16	.13	.16	.00	.00
Number Series	.83	-.12	-.23	[illegible]	[illegible]	-.05	-.22	-.32	.09	-.11	-.14	.08
Bead Memory	.72	.44	.07	-.36	-.28	.01	.09	.07	.20	.12	-.11	.07
Memory for Sentences	.77	-.36	.39	-.01	-.01	.13	-.04	.24	-.12	-.12	-.09	.14
Memory for Digits	.68	-.19	.60	.22	-.07	-.07	.01	-.18	.05	.19	.04	-.08
Memory for Objects	.65	.31	.18	-.13	.63	.14	.01	-.08	.06	-.08	-.04	-.03
Eigenvalue	7.08	0.96	0.82	0.72	0.61	0.46	0.35	0.30	0.24	0.20	0.14	0.11
% of Total Variance	59.0	8.0	6.8	6.0	5.1	3.8	2.9	2.5	2.0	1.7	1.2	0.9
Cummulative % Var.	59.0	67.0	73.9	79.9	84.9	88.8	91.7	94.2	96.2	97.9	99.1	100

Appendix C.4

Two-, Three-, and Four-Component Varimax Rotated Principal Components For The Stanford-Binet IV: Ages 9-0 to 12-11 Years (Missing Data Handled Through Pairwise Correlations; n=66)

Stanford-Binet IV Age 9-0 to 12-11	Varimax Rotated Principal Components													
	Two Components Extracted			Three Components Extracted				Four Components Extracted						
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2		
Vocabulary	.84	.32	.81	.74	.25	.46	.82	.77	.33	.38	.08	.85		
Comprehension	.81	.27	.73	.79	.19	.32	.76	.8[illegible]	.26	.24	.09	.80		
Absurdities	.67	.58	.78	.73	.50	.22	.82	.67	.45	.19	.39	.82		
Pattern Analysis	.41	.74	.71	.43	.70	.19	.71	.36	.64	.18	.36	.71		
Copying	.22	.78	.65	.23	.76	.14	.66	.06	.48	.23	.76	.86		
Matrices	.63	.53	.68	.64	.47	.26	.70	.65	.52	.19	.15	.75		
Quantitative	.63	.36	.53	.75	.27	.09	.65	.61	.04	.13	.66	.82		
Number Series	.72	.44	.71	.76	.36	.23	.76	.68	.24	.22	.48	.79		
Bead Memory	.28	.79	.70	.28	.77	.18	.71	.28	.86	.12	.10	.83		
Memory for Sentences	.82	.21	.72	.45	.20	.79	.87	.49	.22	.75	.08	.87		
Memory for Digits	.65	.29	.50	.19	.32	.85	.86	.19	.22	.88	.24	.91		
Memory for Objects	.31	.65	.52	.22	.64	.29	.55	.20	.64	.27	.20	.56		
Sum of Squared Loadings	4.61	3.45		3.86	2.98	2.01		3.50	2.59	1.85	1.68			
% of Total Variance	38.4	28.8		32.2	24.8	16.8		29.2	21.6	15.4	14.0			
Cummulative % Var.	38.4	67.2		32.2	57.0	73.8		29.2	50.8	66.2	80.2			

Appendix C.5

Five-Component Varimax Rotated Principal Components For The Stanford-Binet IV: Ages 9-0 to 12-11 Years (Missing Data Handled Through Pairwise Correlations; n=66)

Stanford-Binet IV Age 9-0 to 12-11	Varimax Rotated Principal Components					
Tests	**I**	**II**	**III**	**IV**	**V**	**h^2**
Vocabulary	.75	.37	.41	.05	.05	.87
Comprehension	.82	.19	.24	.04	.24	.82
Absurdities	.68	.38	.19	.36	.24	.82
Pattern Analysis	.36	.66	.21	.38	.12	[illegible]
Copying	.10	.37	[illegible]	.78	.23	.86
Matrices	.65	.42	[illegible]	.12	.33	.76
Quantitative	.66	-.02	[illegible]	.61	.08	.83
Number Series	.70	.25	[illegible]	.44	.03	.80
Bead Memory	.25	.88	.16	.13	.19	.91
Memory for Sentences	.48	.17	.76	.07	.16	.87
Memory for Digits	.18	.17	.88	.26	.14	.91
Memory for Objects	.24	.22	.19	.20	.88	.96
Sum of Squared Loadings	3.57	2.01	1.88	1.60	1.16	
% of Total Variance	29.8	16.8	15.7	13.3	9.7	
Cummulative % Var.	29.8	46.6	62.3	75.6	85.3	

Appendix C.6

Unrotated Principal Components For The Stanford-Binet IV: Ages 13-0 to 16-11 Years (Missing Data Handled Through Pairwise Correlations; n=56)

Stanford-Binet IV Age 13-0 to 16-11	Unrotated Principal Components									
Tests	I	II	III	IV	V	VI	VII	VIII	IX	X
Vocabulary	.89	-.05	-.24	.25	-.12	.02	-.25	.04	.11	-.02
Comprehension	.93	-.17	-.16	.06	.03	.00	.06	.11	-.14	.19
Pattern Analysis	.90	-.05	.20	.00	-.14	-.30	-.07	-.17	-.07	.02
Matrices	.90	-.13	.09	-.04	.24	.22	-.06	-.21	.02	.04
Quantitative	.93	-.13	-.07	-.07	.18	-.05	[illegible]	.09	-.15	-.19
Number Series	.89	-.31	.04	-.03	.11	-.13	[illegible]	.06	.21	.00
Bead Memory	.86	-.10	.36	.08	-.25	.20	.06	.09	-.02	-.04
Memory for Sentences	.87	.18	-.33	-.03	-.18	.06	.20	-.13	-.01	-.06
Memory for Digits	.83	.30	.01	-.45	-.05	.01	-.10	.09	.06	.04
Memory for Objects	.73	.59	.14	.25	.18	-.05	.05	.04	.02	[illegible]
Eigenvalue	7.67	0.64	0.39	0.34	0.26	0.20	0.16	0.13	0.10	0.08
% of Total Variance	76.7	6.4	3.9	3.4	2.6	2.0	1.6	1.3	1.0	0.8
Cummulative % Var.	76.7	83.2	87.1	90.5	93.1	95.2	96.8	98.1	99.2	100.0

Appendix C.7

Two-, Three-, and Four-Component Varimax Rotated Principal Components For The Stanford-Binet IV: Ages 13-0 to 16-11 Years (Missing Data Handled Through Pairwise Correlations; n=56)

Stanford-Binet IV Age 13-0 to 16-11	Varimax Rotated Principal Components											
	Two Components Extracted			Three Components Extracted				Four Components Extracted				
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Vocabulary	.76	.47	.79	.46	.72	.33	.85	.46	.75	.34	.14	.90
Comprehension	.86	.40	.90	.59	.71	.27	.92	.59	.69	.19	.26	.93
Pattern Analysis	.77	.47	.82	.72	.39	.42	.85	.72	.36	.33	.31	.85
Matrices	.82	.41	.83	.70	.49	.33	.84	.70	.44	.23	.33	.84
Quantitative	.84	.43	.89	.63	.63	.32	.89	.62	.58	.19	.38	.90
Number Series	.91	.25	.89	.75	.55	.16	.90	.75	.52	.08	.25	.90
Bead Memory	.76	.41	.75	.82	.24	.39	.88	.81	.24	.36	.20	.89
Memory for Sentences	[illegible]	.65	.79	.28	.75	.51	.90	.28	.70	.36	.46	.90
Memory for Digits	.51	.72	.78	.39	.44	.65	.78	.38	.29	.33	.80	.97
Memory for Objects	.27	.90	.88	.27	.24	.88	.90	.27	.27	.87	.26	.96
Sum of Squared Loadings	5.40	2.94		3.50	2.99	2.20		3.47	2.67	1.48	1.46	
% of Total Variance	54.0	29.4		35.0	29.9	22.0		34.7	26.7	14.8	14.6	
Cummulative % Var.	54.0	83.4		35.0	64.9	86.9		34.7	61.4	76.2	90.8	

Appendix D.1

Pattern Matrix[a] For Two-, Three-, and Four-Component Oblimin Rotated Principal Components For The Stanford-Binet IV: Ages 6-0 to 8-11 Years[b]

Stanford-Binet IV Age 6-0 to 8-11	Oblimin Rotated Principal Components											
	Two-Factors Rotated			Three Factors Rotated				Four Factors Rotated				
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Vocabulary	.84	.05	.72	.26	-.04	-.75	.67	.04	-.06	-.66	-.45	.86
Comprehension	.77	.22	.67	.25	.18	-.61	.70	.09	.18	-.54	-.31	.71
Absurdities	.80	-.09	.63	.71	.08	-.15	.67	.79	.17	-.15	.05	.79
Pattern Analysis	[illegible]	.20	.54	.66	.44	.04	.70	.40	.46	.09	-.36	.70
Copying	.70	.10	.48	.61	.04	-.15	.50	.01	-.03	-.03	-.92	.85
Quantitative	.74	-.49	.70	.87	-.28	.01	.74	.92	-.19	.00	-.01	.83
Bead Memory	[illegible]	.85	.79	-.06	.92	-.07	.86	-.09	.94	-.06	.07	.88
Memory for Sentences	.64	.22	.49	-.13	-.02	-.95	.78	.06	.02	-.92	.14	.84

[a] Unlike orthogonal rotations where the pattern matrix and structure matrix are the same, in oblique rotations the pattern matrix and stuctire matrices are different. The factor pattern matrix gives the weights by which factors reproduce the variables.

[b] Missing data has been replaced with regression estimates (n=46).

Appendix D.2

Structure Matrix[a] and Intercorrelations Between Components For Two-, Three-, and Four-Component Oblimin Rotated Principal Components For The Stanford-Binet IV: Ages 6-0 to 8-11 Years[b]

Stanford-Binet IV Age 6-0 to 8-11	Oblimin Rotated Principal Components												
	Two Factors Rotated			Three Factors Rotated				Four Factors Rotated					
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Vocabulary	.85	.15	.72	.64	.21	-.87	.82	.52	.24	-.82	-.69	.86
Comprehension	.79	.31	.67	.60	.39	-.79	.70	.49	.43	-.74	-.58	.71
Absurdities	.79	.00	.63	.80	.24	-.53	.67	.86	.35	-.52	-.39	.79
Pattern Analysis	.71	.28	.54	.72	.54	-.42	.70	.61	.59	-.34	-.61	.70
Copying	.68	-.02	.48	.69	.19	-.47	.50	.43	.18	-.34	-.92	.85
Quantitative	.68	-.40	.70	.82	-.14	-.36	.74	.89	-.03	-.33	-.38	.83
Bead Memory	.30	.87	.79	.12	.93	-.28	.86	.07	.93	-.28	-.12	.88
Memory for Sentences	.66	.30	.49	.35	.22	-.88	.78	.39	.28	-.91	-.21	.84

Correlation Among Factors (Age 6-0 to 8-11 Above)

Factor	Two Factors Rotated		Three Factors Rotated			Four Factors Rotated			
	I	II	I	II	III	I	II	III	IV
I	1.00	.12	1.00	.17	-.51	1.00	.18	-.42	-.45
II	.12	1.00	.17	1.00	-.27	.18	1.00	-.29	-.22
III			-.51	-.27	1.00	-.42	-.29	1.0	.35
IV						-.45	-.22	.35	1.00

[a] Unlike orthogonal rotations where the pattern matrix and structure matrix are the same, in oblique rotations the pattern matrix and structure matrices are different. The structure matrix gives the variable-factor correlations, and is the matrix that is interpreted when oblique rotations are used.

[b] Missing data has been replaced with regression estimates (n=46).

Appendix D.3

Pattern Matrix[a] For Two-, Three-, and Four-Component Oblimin Rotated Principal Components For The Stanford-Binet IV: Ages 9-0 to 12-11 Years[b]

Oblimin Rotated Principal Components

Two Factors Rotated

Stanford-Binet IV Age 9-0 to 12-11 — Tests	I	II	h^2
Vocabulary	.85	.06	.80
Comprehension	.79	.08	.70
Absurdities	.47	.50	.78
Pattern Analysis	.19	.69	.68
Copying	-.07	.85	.65
Matrices	.50	.42	.70
Quantitative	.66	.09	.53
Number Series	.69	.24	.75
Bead Memory	-.05	.88	.71
Memory for Sentences	.96	-.15	.75
Memory for Digits	.80	-.07	.57
Memory for Objects	.08	.66	.51

Three Factors Rotated

Tests	I	II	III	h^2
Vocabulary	.60	.05	.38	.80
Comprehension	.72	.02	.21	.73
Absurdities	.55	.43	.04	.81
Pattern Analysis	.21	.66	.05	.68
Copying	.13	.79	-.15	.65
Matrices	.51	.36	.10	.71
Quantitative	.91	-.05	-.08	.71
Number Series	.75	.15	.09	.80
Bead Memory	-.08	.88	.05	.75
Memory for Sentences	.23	-.03	.81	.87
Memory for Digits	-.04	.10	.88	.83
Memory for Objects	-.08	.69	.18	.56

Four Factors Rotated

Tests	I	II	III	IV	h^2
Vocabulary	.80	-.18	.24	-.18	.87
Comprehension	.87	-.13	.09	-.09	.78
Absurdities	.67	.29	.03	-.12	.81
Pattern Analysis	[illegible]	.42	.06	-.27	.69
Copying	.07	.85	.05	[illegible]	[illegible]
Matrices	.69	.14	.04	[illegible]	[illegible]
Quantitative	.75	.25	-.01	.45	.83
Number Series	.76	.19	.10	.13	.81
Bead Memory	.27	.37	-.01	-.65	.87
Memory for Sentences	.28	-.10	.78	-.01	.87
Memory for Digits	-.08	.12	.95	.06	.89
Memory for Objects	-.02	.56	.29	-.18	.58

[a] Unlike orthogonal rotations where the pattern matrix and structure matrix are the same, in oblique rotations the pattern matrix and structure matrices are different. The factor pattern matrix gives the weights by which factors reproduce the variables.

[b] Missing data has been replaced with regression estimates (n=66).

Appendix D.4
Structure Matrix[a] and Intercorrelations Between Components For Two-, Three-, and Four-Component Oblimin Rotated Principal Components For The Stanford-Binet IV: Ages 9-0 to 12-11 Years[b]

Stanford-Binet IV Age 9-0 to 12-11	Oblimin Rotated Principal Components											
	Two Factors Rotated			Three Factors Rotated				Four Factors* Rotated				
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Vocabulary	.89	.62	.80	.83	.59	.71	.80	.88	.34	.67	-.36	.87
Comprehension	.84	.59	.70	.84	.54	.59	.73	.87	.35	.55	-.26	.78
Absurdities	.80	.81	.78	.82	.77	.53	.81	.85	.65	.53	-.31	.81
Pattern Analysis	.64	.81	.68	.62	.80	.48	.68	.67	.68	.48	-.44	.69
Copying	.49	.81	.65	.52	.80	.31	.65	.51	.89	.38	-.14	.80
Matrices	.78	.75	.70	.78	.71	.54	.71	.82	.53	.52	-.38	.74
Quantitative	.72	.53	.53	.84	.45	.36	.71	.78	.54	.40	.26	.83
Number Series	.85	.69	.75	.88	.63	.55	.80	.88	.57	.56	-.08	.81
Bead Memory	.53	.84	.71	.47	.86	.44	.75	.57	.61	.41	-.76	.87
Memory for Sentences	.86	.48	.75	.62	.50	.91	.87	.68	.32	.91	-.22	.87
Memory for Digits	.75	.46	.57	.46	.51	.91	.83	.51	.41	.94	-.15	.89
Memory for Objects	.51	.71	.51	.42	.73	.48	.56	.46	.68	.52	-.34	.58

* The four-component oblimin solution required 42 iterati[illegible]onverge; this suggests that rotating four factors obliquely may involve over-factoring of the correl[illegible] matrix.

Correlation Among Factors (Age 9-0 to 12-11 Above)

Factor	Two Factors Rotated		Three Factors Rotated			Four Factors Rotated			
	I	II	I	II	III	I	II	III	IV
I	1.00	.66	1.00	.59	.51	1.00	.50	.56	-.20
II	.66	1.00	.59	1.00	.49	.50	1.00	.35	-.18
III			.51	.49	1.00	.56	.35	1.00	-.22
IV						-.20	-.18	-.22	1.00

[a] Unlike orthogonal rotations where the pattern matrix and structure matrix are the same, in oblique rotations the pattern matrix and structure matrices are different. The structure matrix gives the variable-factor correlations, and is the matrix that is interpreted when oblique rotations are used.

[b] Missing data has been replaced with regression estimates (n=66).

Appendix D.5

Pattern Matrix[a] For Two, Three, and Four-Component Oblimin Rotated Principal Components For The Stanford-Binet IV: Ages 13-0 to 16-11 Years[b]

Stanford-Binet IV Age 13-0 to 16-11	Oblimin Rotated Principal Components											
	Two Factors Rotated			Three Factors Rotated				Four Factors * Rotated				
Tests	I	II	h^2	I	II	III	h^2	I	II	III	IV	h^2
Vocabulary	.82	.09	.80	.76	.09	-.29	.86	.10	.16	-.88	.14	.92
Comprehension	.98	-.04	.89	.92	-.04	-.22	.92	.23	-.04	-.74	-.10	.93
Pattern Analysis	.86	.06	.81	.86	.10	.15	.85	.60	.12	-.20	-.14	.85
Matrices	.92	.01	.85	.90	.05	.06	.86	.52	.01	-.30	-.24	.87
Quantitative	.94	.01	.89	.90	.03	-.12	.90	.32	-.08	-.51	-.31	.91
Number Series	.99	-.05	.91	.97	-.02	-.02	.91	.45	-.05	-.44	-.22	.92
Bead Memory	.88	-.01	.76	.91	.06	.34	.90	.85	.18	.01	-.02	.91
Memory for Sentences	.44	.52	.81	.87	.50	-.37	.90	-.20	.23	-.66	-.35	.90
Memory for Digits	.34	.61	.80	.32	.63	-.04	.80	.11	.15	.03	-.85	.97
Memory for Objects	-.07	.99	.89	-.07	1.01	.05	.91	.11	.87	-.04	-.06	.97

* The four-component oblimin solution required 63 iterations to converge; this suggests that rotating four factors obliquely may result in over-factoring of the correlation matrix.

[a] Unlike orthogonal rotations where the pattern matrix and structure matrix are the same, in oblique rotations the pattern matrix and structure matrices are different. The factor pattern matrix gives the weights by which factors reproduce the variables.

[b] Missing data has been replaced with regression estimates (n=56).

Appendix D.6

Structure Matrix[a] and Intercorrelations Between Components For Two-, Three-, and Four-Component Oblimin Rotated Principal Components For The Stanford-Binet IV: Ages 13-0 to 16-11 Years[b]

Stanford-Binet IV Age 13-0 to 16-11 Tests	Two Factors Rotated I	II	h^2	Three Factors Rotated I	II	III	h^2	Four Factors* Rotated I	II	III	IV	h^2
Vocabulary	.89	.71	.80	.88	.69	-.44	.86	.70	.60	-.95	-.58	.92
Comprehension	.94	.70	.89	.94	.68	-.39	.92	.78	.52	-.94	-.69	.93
Pattern Analysis	.90	.71	.81	.91	.73	-.03	.85	.88	.58	-.77	-.68	.85
Matrices	.92	.71	.85	.93	.72	-.12	.86	.86	.54	-.81	-.73	.87
Quantitative	.94	.72	.89	.94	.71	-.29	.90	.81	.51	-.88	-.78	.91
Number Series	.96	.70	.91	.96	.71	-.20	.91	.86	.52	-.87	-.74	.92
Bead Memory	.87	.66	.76	.88	.69	.17	.90	.94	.58	-.68	-.59	.91
Memory for Sentences	.83	.85	.81	.81	.82	-.50	.90	.56	.68	-.87	-.79	.90
Memory for Digits	.80	.87	.80	.79	.87	-.18	.80	.64	.64	-.67	-.97	.97
Memory for Objects	.68	.94	.89	.67	.95	-.06	.91	.57	.97	-.62	-.62	.97

* The four-component oblimin solution required 63 iterations to converge; this suggests that rotating four factors obliquely may involve over-factoring of the correlation matrix.

Correlation Among Factors (Age 13-0 to 16-11 Above)

Factor	Two Factors Rotated I	II	Three Factors Rotated I	II	III	Four Factors Rotated I	II	III	IV
I	1.00	.76	1.00	.75	-.19	1.00	.46	-.69	-.57
II	.76	1.00	.75	1.00	-.12	.46	1.00	-.53	-.54
III			-.19	-.12	1.00	-.69	-.53	1.00	.65
IV						-.57	-.54	.65	1.00

[a] Unlike orthogonal rotations where the pattern matrix and structure matrix are the same, in oblique rotations the pattern matrix and structure matrices are different. The structure matrix gives the variable-factor correlations, and is the matrix that is interpreted when oblique rotations are used.

[b] Missing data has been replaced with regression estimates (n=56).