Order Number 9418214

The multiple group confirmatory factor analysis of the Multidimensional Aptitude Battery

Moon, TaiHyong, Ph.D.

Wayne State University, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

THE MULTIPLE GROUP CONFIRMATORY FACTOR ANALYSIS OF THE MULTIDIMENSIONAL APTITUDE BATTERY

by

TAIHYONG MOON

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

1993

MAJOR : EVALUATION AND RESEARCH

Approved by:

9/27/93

Adviser Date

DEDICATION

To my parents

CHUNGROCK MOON and CHUNOCK MOON(LEE)

ACKNOWLEDGEMENTS

I would like to express my appreciation to the members of my dissertation committee: Dr. D. Marcotte, Dr. B. Markman, Dr. M. Smith, Jr., and Dr. J. Hankin. I am especially indebted to my adviser, Dr. D. Marcotte, for his advice and guidance given during the entire study.

I am also indebted to Dr. D. N. Jackson, the author of the MAB, for his great help. I really appreciated him for providing me with all the information which I needed for the study. I also wish to express my sincere gratitude to others, too numerous to name individually, who assisted me in many ways to complete my study.

I am extremely grateful to my family members for their endless assistance and encouragement. A great love and gratitude is extended to my parents, Chungrock Moon and Chunock Moon(Lee), for their devotional contribution to me. A special gratitude goes to my sister and her husband, Soobok Moon Kim and Insung Kim who provided the sacrifice and warm understanding for me during the study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Statement and background of the Problem

Intelligence tests play a significant role in assessing both groups and individuals in educational settings. Intelligence tests are a psychometric device, that is, sets of standardized questions and tasks for assessing an individual's potential for purposeful and useful behavior. In practice, they are designed to measure major mental abilities (Wechsler, 1981, p. 7). Despite the controversy surrounding intelligence tests, their use has been and continues to be influential in shaping educational policy and practice (Caroll, 1978).

An important goal of contemporary testing is to contribute to self-understanding and personal development. The information provided by tests is being used increasingly to assist individuals in educational and career planning and in making decisions about their own lives (Anastasi, 1988, p. 363).

The history of testing can date back to the system of civil service examination prevailing in the Chinese empire for some three thousand years. At present, schools are among the largest test users. The following are among the many educational uses of the tests: the classification of children based on their ability to profit from different types of school instruction, the identification of the intellectually retarded and gifted, the diagnosis of academic failures, the educational and vocational counseling of high school and college students, and the selection of applicants for professional and other special schools (Anastasi, 1988, pp. 4-5).

Psychological tests are among the many practices that counteract the idea that people are the same. Tests are designed to measure differences in intelligence and aptitude between individuals or between the reactions of the same individual on different occasions. Test scores demonstrating differences between people may suggest to some that people are not created with the same abilities. The most irritating problem is that certain ethnic groups, on the average, score differently on some psychological tests. The most controversial case concerns intelligence tests (Kaplan, 1985, p. 465).

The uses of intelligence tests are many and quite varied, ranging from predicting future academic levels, to distinguishing organic from psychiatric syndromes, to evaluating personality. Intelligence tests are among the most frequently administered of all tests by clinical psychologists and school psychologists (Reynolds and Kaufman, 1985, p. 601). The investigation of intelligence is rapidly becoming central to psychology as a discipline, and to all disciplines involved in the scientific study of the mind (Sternberg, 1982, p. xi). The results of intelligence measures have at least three major areas of educational application: (1) selection and placement, (2) screening, diagnosis, and remedial planning, and (3) accountability, research, and evaluation (Bohem, 1985, p. 933). However, it is true that intelligence testing is one of the most controversial areas of assessment.

Intelligence is omnipresent in our daily activities and influences much of what we do and are able to accomplish, but intelligence is not omnipotent. Failure to realize this simple difference between omnipresent and omnipotent has resulted in many abuses of intelligence, abuses which can be avoided if an adequate understanding of validity and of the limitations of intelligence as construct is obtained (Reynolds and Kaufman, 1985, p. 613).

Conventional intelligence tests, even the entire concept of intelligence testing, are the focus of considerable controversy. Reynolds (1982) pointed out that intelligence testing in the education area has been the target of criticism and attack. A principal criticism of intelligence tests is their possible cultural biases. Therefore, critics of testing demand a moratorium on their use with children (Reynolds and Kaufman, 1985, p. 609). Those who insist on the inherent unacceptability of IQ tests as measurement devices with no real utility are at one extreme of the issues, and those who believe in its immense value are at the other extreme (Reynolds and Kaufman, 1985, p. 609).

However, it is true that because of the great influence of intelligence tests on one's educational and job opportunities, the examination of psychological characteristics of intelligence gives rise to a great concern to many persons. In this context, the investigation of the psychometric properties of the intelligence test seems very important in terms of its valid use and the correct interpretation of test results, which eventually help one decide on what should be done for the development and improvement of an individual's capacity.

When psychological aptitude tests, such as Wechsler's scales and a series of Stanford-Binet intelligence scales, were newly developed and published, many studies were conducted with respect to their psychometric properties including reliability, validity, and factor structures. According to Reynolds and Kaufman (1985, p. 611), knowledge and skill in psychometrics and measurement are requisite to intelligence testing. The clinical evaluation of test performance must be directed by careful analyses of statistical properties of the test scores, the internal psychometric characteristics of the test, and the data regarding its relationship to external factors.

Since the Multidimensional Aptitude Battery (MAB: Jackson, 1984) is a rather recently developed group intelligence test, it should be submitted to the same scrutiny as the Wechsler and Stanford-Binet tests. Some controversies with respect to its value as a psychological test already existed (Krieshok & Harrington, p. 1985). Therefore, this research will evaluate the appropriateness of the MAB as a tool for psychological measurement.

Purpose of the Research

This research focuses on the examination of the psychometric properties of the MAB (Jackson, 1984). The MAB is a test of intelligence and mental abilities patterned quite closely after the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981).

Two factor analyses, using two different samples, of the MAB were reported in the manual. Furthermore, other studies conducted for the construct validity of the MAB (Lee, et al, 1990; Wallbrown, Carmin, & Barnett, 1988 & 1989), using the hierarchical factor analysis of subtest intercorrelations, provided considerable support for Jackson's (1984, p. 48) assertion that the battery measures the same ability dimensions as the WAIS-R.

However, no confirmatory factor analytic studies of the MAB, especially using the structural equation models with subjects in a wide age range, were reported. Previous studies used samples in a restricted age range or subjects from special populations. According to Joreskog & Sorbom (1989, p. 96), it is highly desirable that a hypothesis which has been suggested by mainly exploratory procedures should subsequently be confirmed, or disproved, by obtaining new data and subjecting these to more rigorous statistical techniques.

Furthermore, no studies were reported about the invariance of factor structure of the MAB across different age groups. The multiple group analysis allows one to compare factor structures across groups and to test increasingly restrictive hypotheses concerning how well the factor structure is replicated across groups.

Therefore, the purpose of this study is to factor analyze the data obtained from the 1990 norming study using the confirmatory factor analysis in order to investigate the extent to which the proposed two factor model explains the variation in the MAB subtests, and to examine the factorial invariance across groups.

Research Questions

The major research questions of this research are as following:

(1) Research Question 1:

Does the two factor model explain the variation in the MAB subtest scores for all age groups simultaneously?

(2) Research Question 2:

Are there any significant differences across the age groups with respect to factorial invariance?

Hypotheses

For research question one, the following null hypotheses are posited:

Ho1.1 There is no significant difference between the null model which represents no relationships among the subtests and the model that represents the actual data for all groups.

Ho1.2 There is no significant difference between the two factor model and the model that represents the actual data for all groups.

Ho1.3 There is no significant difference in explanatory power between the two factor model and the null model.

For the research question two, the following null hypothesis is posited:

Ho2 There is no significant difference across the age groups with respect to factorial invariance.

Significance of the Research

The obtained results and findings of this research will provide evidence for the possible use of the MAB as an instrument for screening and diagnosing people in business and industry, clinics and mental health facilities, and in educational and career counseling settings. With the findings of this research, school psychologists will be in a better position to draw conclusions as to the appropriateness of the use of the MAB in assessing intellectual abilities of their clients. In addition, the results obtained from this study will be used as a guideline for other psychometric studies of the MAB using other samples and statistical techniques.

Definitions of Terms

1. Intelligence : Scores on the Full Scale of the Multidimensional Aptitude Battery (MAB).
2. Verbal IQ : Scores on the Verbal Scale of the MAB.
3. Performance IQ : Scores on the Performance Scale of the MAB.
4. Full Scaled IQ : Combined scores of Verbal IQ and Performance IQ of the MAB.

CHAPTER II

REVIEW OF RELATED LITERATURES

The literatures reviewed for this chapter are organized into following sections: (1) the theories of intelligence, (2) the factor analysis (exploratory factor analysis and confirmatory factor analysis), (3) LISREL program, and (4) factor analytic studies of the Multidimensional Aptitude Battery.

Theories of Intelligence

The concepts of intelligence and education are so often discussed and studied independently that they are conventionally assumed to be distinct. Still, one can entertain the notion that intelligence and education cannot really exist independently --that the referents of these terms are not separable strands interwound in human mental life but rather are fundamentally confounded. In other words, human intelligence is fundamentally a product of education and education is fundamentally a product of the exercise of human intelligence (Snow, 1982, p. 493).

Intelligence is both a scientific and a folk concept. This fact is often the source of confusion and can disrupt communication (Linn, 1989, p. 29). Some of the older, more common definitions of intelligence considered it as an ability or a capacity (Wechsler, 1958, p. 7) and carry the surplus meaning that intelligence is an inherited, unchangeable characteristic. Intelligence is defined as the entire repertoire of acquired skills, knowledge, learning sets, and generalization tendencies considered intellectual in nature that are available at any one period in time (Cleary et al., 1975, p. 19). Most evidence would suggest a hallmark of intelligence is the ability to generalize information from one situation to another (Campione & Brown, 1979, p. 279).

In defining intelligence, there are two major factor analytic camps: those who support a general (g) factor theory of intelligence and those who favor a multiple factor theory. For example, Galton insisted on a general unitary function of intellectual abilities. On the other hand, researchers like Thurstone and Guilford proposed that the intellect consisted of many independent faculties. However, despite some differences in defining the nature of intelligence between two camps, the general intelligence theory is generally accepted with the belief of the multidimensional intelligent behavior.

Spearman (1927) suggested a two-factor theory of intelligence, a general factor (g) and one or more specified factors (s). According to him, the 'g' factor is a general mental energy, and the more complicated the mental activities, the greater amount of 'g' they have. Jensen (1979) pointed out that the 'g' factor is an index of general mental ability or intelligence and represents the inventive as contrasted with the reproductive aspect of mental ability.

Thurstone (1938) refusing the concept of a unitary trait of intelligence assumed a certain systematic organization of human intelligence. When he found eight primary mental abilities of Verbal, Perceptual Speed, Inductive Reasoning, Number, Rote Memory, Deductive Reasoning, Word Fluency, and Space or Visualization, he thought that intelligence could be divided into these multiple factors, each of which has equal weight.

Guilford (1967) proposed a three dimensional Structure of Intelligence (SI) model as a means for organizing intellectual factors into a system. The SI model consists of :

(1) operations involved in processing information (the ways people think),

(2) contents (what people think about), and

(3) products (the result of operations and contents).

According to the SI model, intellectual activities are interpreted in terms of the kind of mental operation performed, the type of content on which the mental operation is performed, and the resulting product. Each of these dimensions is divided into more detailed sub-areas.

The operations component consists of five areas: cognition, memory, divergent production, convergent production, and evaluation. The contents component is partitioned into four areas: figural, symbolic, semantic, and behavioral. The products component is subdivided into six areas: units, classes, relations, systems, transformations, and implications. Thus, in the SI model, a total of 120 possible factors from 5 x 4 x 6 matrix are postulated and a factor is produced by combining one element from each of the three dimensions.

Vernon (1950) proposed a hierarchical theory of intelligence. In his theory, general ability, 'g', is located at the highest level; at the next level, the two major group factors exist which represent skills in the verbal-numerical-educational *(v:ed)* field and in the practical-mechanical-spatial-physical *(k:m)* field. Each of these group factors has smaller subdivisions (or minor group factors) at the next lower level. The verbal-numerical-educational factor consists of creative abilities, verbal fluency, and numerical factors. The practical-mechanical-spatial-physical factor is composed of spatial, psychomotor, and mechanical information factors. Finally, specific factors peculiar to certain tests are located at the lowest level.

Cattell and Horn (Cattell, 1963; Horn, 1967, 1968, 1978a, 1978b, 1985; Horn & Cattell, 1967) postulated that there are two types of intelligence: Fluid and Crystallized. Fluid intelligence refers to nonverbal tasks (comparable to Vernon's practical-mechanical-spatial-physical factor), relatively culture-free mental efficiency, involving adaptive and new learning capabilities, and being related to mental operations and process. This intelligence increases till adolescence, then declines because of physiological degeneration.

Crystallized intelligence refers to acquired skills and knowledge that are strongly dependent on exposure to culture (comparable to Vernon's verbal-numerical-educational factor). This crystallized intelligence, reflecting cultural assimilation, is highly influenced by formal and informal educational factors through life, and increases continuously through middle adulthood. Crystallized intelligence involves well established cognitive functions and is related to mental products and achievement. Tasks measuring fluid intelligence (e.g., the WAIS-R Block Design subtest) may require more concentration and problem solving than do crystallized tasks (e.g., the WAIS-R Vocabulary and Information subtests), which tap retrieval and application of general knowledge abilities (Sattler, 1988).

Factor Analysis

Factor analysis explains the variation and covariation in a set of observed variables in terms of a set of unobserved factors (Long, 1983, p. 22). It attempts to simplify complex and diverse relationships that exist among a set of observed variables by uncovering common dimensions or factors that link together the seemingly unrelated variables, and consequently provides insight into the underlying structure of data (Dillon & Goldstein, 1984, p. 53). Factor analysis is based on the assumption that intercorrelations can be explained by some underlying set of (unobservable) factors that are fewer than the tests (or variables) themselves (SPSS/PC+, pp. B-41 - B-42).

A factor is defined as a cluster of common elements for a set of interrelated tests. In other words, a factor is a construct, a hypothetical entity, a latent variable that is assumed to underlie tests, scales, items, and, indeed, measures of almost any kind (Kerlinger, p. 569).

Information provided by factor analysis includes (1) an indication of the degree to which varying numbers of factors explain the correlations among tests, (2) communality, or the total amount of variability in test scores explained by common factors, and (3) factor loadings which are simply correlations between factors and tests, that is, the relative weight of each factor in determining the performance on each test.

The basic factor model assumes that a score on a variable can be expressed as a linear combination or as a weighted sum of scores on factors underlying performance in that variable. That is, each observed variable is defined as a linear function of one or more factors.

Mathematical expression of the relationship between the observed variables and the factor is

$$x = \Lambda\xi + \delta$$

where x is a (q x 1) vector of observed variables;

ξ is a (s x 1) vector of common factors;

Λ is a (q x s) matrix of factor loadings relating the observed x's to the latent ξ's; and

δ is a (q x 1) vector of the residual or unique factors.

The assumptions of factor analysis are as following:

1. Variables (both the observed and unobserved ones) are

measured from their means. Thus the expected value of each vector is a vector containing zeros: $E(x)=0$; $E(\xi)=0$; and $E(\delta)=0$.

2. The number of observed variables is greater than the number of common factors.

3. Common factors and unique factors are uncorrelated: $E(\xi\delta')=0$ or $E(\delta\xi')=0$.

Statistically, the task is to explain the interrelationships among the observed variables, as indicated by the covariances among the observed variables, in terms of relationships among the observed and latent variables.

In factor analysis, the primary concern is with the determination of the coefficients, or loadings. A factor loading is the coefficient used to express a standardized variable in terms of the factors. It indicates how much weight is assigned to each factor. The factor loadings are the standardized regression coefficients in the multiple regression equation with the original variable as the dependent variable and the factors as the independent variables. Factors with large coefficients for a variable are closely related to the variable. The factor loadings are multiplied by the factor and summed to equal the measured variable. This means that an observed variable is assumed to be a linear combination of the factors. If the factors are uncorrelated, the values of the coefficients are not dependent on each other. They represent the unique contribution of each factor; they are the correlations between the factors and the variable.

The variance in a factor analysis, associated with a test or subtest, can be partitioned into three categories: (1) communality (2) specificity and (3) error variance. Communality or common factor variance is the proportion of variance explained by the common factors (those that appear in more than one test) and is defined as the squared multiple correlation of the common factors with the measured variable. In an orthogonal factor model, it is equivalent to the sum of the squared factor loadings for each variable. Communality can range from 0 to 1, with 0 indicating that the common factors explain none of the variance, and 1 indicating that all the variance is explained by the common factors.

Specificity or specific factor variance refers to that part of the total variance that is due to factors that are specific to the particular test, and not to measurement error or common factors. Error variance refers to that part of the total variance that remains when reliability (the communality component plus the specific component) of the test is subtracted from the total variance. The combination of specificity and error variance is called the uniqueness of the variable, which is residual in the sense that it corresponds to that portion of the observed variable that is not explained by one or more common factors.

Factor analysis usually proceeds in four steps. The first step in factor analysis is collecting the relevant data for analysis and preparing a correlation matrix for all variables--the data used directly in the factor analysis. Variables that do not appear to be related to other variables can be identified from the correlation matrix and associated statistics. The appropriateness of the factor model can also be evaluated with such statistics as Kaiser-Meyer-Olkin measure of sampling adequacy (KMO MSA) and Bartlett's test of sphericity (SPSS/PC+, 1988, pp. B-43 - B-45).

The second step is the extraction of initial factors. In this step, the minimum number of factors that can adequately explain the observed correlations (or covariance) among the observed variables (SPSS/PC+, 1988, p. B-43, p. B-53) and the communalities of each variable are determined (Kim and Mueller, 1978b, p. 29).

All initial solutions are based on the orthogonal solution. At this stage, the chief concern is whether a smaller number of factors can account for the covariation among a much larger number of variables. The most commonly used factor extraction methods are (1) principal component analysis, (2) principal axis factoring (or principal factors with iterated communalities), and (3) maximum likelihood method.

The overall objective of the maximum likelihood solution is to find the factor solution which would best fit the observed correlations. In this method, it is assumed that the observed data comprise a sample from a population where a k-common factor model exactly applies, and where the distribution of variables (including the factors) is multivariate normal. What is assumed unknown is the exact configuration of parameters, i.g., the exact loadings on each variable. The objective is then to find the underlying population parameters (under the given hypothesis) that would have the greatest likelihood of producing the observed correlation matrix (Kim and Mueller, 1978b, p. 23). The most important advantage of this model is its large sample significance test. However, in practice, depending on the significance test alone is not desirable so that more common factors than are necessary will be obtained if the sample size is large. That is, with a large size sample, a minor misfit between the model and the data can produce additional significant factors.

Furthermore, the researcher also ascertains how well the chosen model fits the data. When factors are extracted using generalized least-squares (GLS) or maximum-likelihood (ML) estimation and it is assumed that the sample is from a multivariate normal population, it is possible to obtain goodness-of-fit tests for the adequacy of a k-factor model.

The third step, rotation, focuses on transforming the factors to make them more interpretable (SPSS/PC+, 1988, p. B-43). The rotation phase attempts to transform the initial matrix into one that is easier to interpret. Although the factor matrix obtained in the extraction stage indicates the relationship between the factors and the individual variables, it is usually difficult to identify meaningful factors based on this matrix. Often the variables and factors do not appear correlated in any interpretable pattern (SPSS/PC+, 1988, p. B-53). Therefore, this step involves finding simpler and more easily interpretable factors through rotations, while keeping the number of factors and communalities of each variable fixed (Kim and Mueller, 1978b, p. 29).

Any rotated factor solution explains exactly as much covariation in the data as the initial solution (Kim and Mueller, 1978a, p. 50). Rotation is not used to improve the quality of the mathematical fit between the observed and reproduced correlation matrices (Tabachnick and Fidell, 1989, p. 628), and does not affect the goodness of fit of a factor solution. That is, although the factor matrix changes, the communalities and the percentage of total variance explained do not change. However, the percentage of variance explained by each factor changes. What is attempted through rotation is a possible 'simplication'. However, there exist different criteria of simplicity leading to different methods of rotation (Kim and Mueller, 1978a, p. 50). Different rotation methods may actually result in the identification of somewhat different factors (SPSS/PC+, 1988, p. B-54). The criterion for ideal rotation is: Rotation should be performed so that each variable loads on one and only one factor (Nunnally, 1978, p. 377).

Although many factorists today talk about rotating to simple structure, no one can say for sure what constitutes simple structure. Rather than talk about simple structure, it would be better to talk about simpler structures. Rotation to achieve simple structure is a fairly objective way to achieve variable simplicity or to reduce variable complexity (Kerlinger, 1986, p. 581). Thus a rotated factor matrix usually is simpler to interpret than the unrotated matrix, and some rotations are simpler to interpret than others. What one seeks is a rotation where there are some relatively pure variables for each factor (Nunnally, 1978, pp. 377-378).

Thurstone (1947, p. 335) laid down five principles or rules of simple structure as a guide to rotation. These are:

1. Each variable should have at least one zero loading.

2. Each factor should have a set of linearly independent variables whose factor loadings are zero.

3. For every pair of factors, there should be several variables whose loadings are zero for one factor but not for the other.

4. For every pair of factors, a large proportion of the variables should have zero loadings on both factors whenever more than about four factors are extracted.

5. For every pair of factors, there should be only a small number of variables with nonzero loadings on both.

The rules are applicable to both orthogonal and oblique rotations. In effect, these criteria call for as pure variables as possible; that is, each variable loaded on as few factors as possible, and as many zeros as possible in the rotated factor matrix. In this way the simplest possible interpretation of the factors can be achieved (Kerlinger, 1986, p. 581).

There are three different approaches to rotation problem (Kim and Mueller, 1978b, p. 30). The first one is to examine the pattern of variables graphically and then rotate the axis or define new axes in such a way that the new axes best satisfy one's criterion of simple and meaningful structure.

The second approach relies on some analytic rotation method that is free of subjective judgment, at least after a particular criterion of simplicity is chosen. There are two subtypes of rotations: orthogonal rotations and oblique rotations.

A variety of algorithms are used for orthogonal rotations to a simple structure. Orthogonal rotations maintain the independence of factors, that is, the angles between the axes are kept at 90 degrees, which means that the correlations between the factors are zero. The most commonly used method is the varimax method. It minimizes the number of variables that have high loadings on a factor and enhances the interpretability of the factors. The quartimax method stresses simple interpretation of variables, since the solution minimizes the number of factors needed to explain a variable. The equamax method is a combination of the varimax method and the quartimax method (SPSS/PC+, 1988, p. B-55).

Rotations in which factor axes are allowed to form acute or obtuse angles are called oblique rotations. Obliqueness means that factors are correlated. Oblique rotations are more general than orthogonal rotations in that they do not arbitrarily impose the restriction that factors should be uncorrelated.

The third approach is to define a target matrix or configuration before actual rotation in order to find the factor patterns that are closest to the given target matrix. Since the specification of a target matrix presumes certain knowledge or a hypothesis about the nature of the factor structure, this strategy approaches confirmatory factor analysis (Kim & Mueller, 1978b, p. 30).

At the final step, factor scores can be computed for each case. Since one of the goals of factor analysis is to reduce a large number of variables to a smaller number of factors, it is often desirable to estimate factor scores for each case. The factor scores can be used in subsequent analyses to represent the values of the factors. A factor can be estimated as a linear combination of the original variables (SPSS/PC+, 1988, p. B-61).

Factor analysis is generally classified into two types: exploratory factor analysis and confirmatory factor analysis. Exploratory factor analysis is used mainly as a means of exploring the underlying factor structure without prior specification of the number of factors and their loadings. Exploratory factor analysis may be used as an expedient way of ascertaining the minimum number of hypothetical factors that can account for the observed covariation, and as a means of exploring the data for possible data reduction.

However, because of such limitation as subjectivity, confirmatory factor analysis was used. Confirmatory factor analysis is the one in which specific expectations concerning the number of factors and their loadings are tested on sample data. The uses of factor analysis need not be confined to exploring the underlying dimensions of the data. Factor analysis can be used as a means of testing specific hypotheses. If it is used to test a certain expectation which researcher may anticipate or hypothesize, then it is used as a means of confirming a certain hypothesis. Thus, it is referred to as confirmatory factor analysis (Kim & Mueller, 1978b).

The assumptions of exploratory factor analysis are as following (Long, 1983, p. 12): (1) All common factors are correlated; (2) All observed variables are directly affected by all common factors; (3) Unique factors are uncorrelated with one another; (4) All observed variables are affected by a unique factor; and (5) All common factors are uncorrelated with all unique factors. In exploratory factor analysis, the researcher does not specify the structure of the relationships among the variables in the model beyond the specification of the numbers of common factors and observed variables to be analyzed. Therefore, the researcher using exploratory factor analysis cannot incorporate substantively meaningful constraints.

Through the confirmatory factor analysis, the limitations of the exploratory factor analysis can be largely overcome by imposing substantively motivated constraints which determine: (1) the relationships among common factors; (2) the relationships between observed variables and common factors; (3) the relationships between observed variables and unique factors; and (4) the relationship among unique factors. In confirmatory factor analysis, it can be assumed that some of the unique factors are correlated and an observed variable has no error (unique) factor associated with it. Furthermore, the observed variables are influenced by only some of the common factors (Long, 1983, p. 15).

Confirmatory factor analysis is powerful because it provides explicit hypothesis testing for factor structure. The hypothesized factor structure is based on a strong theory or past data that allow for specification of a unique factor resolution (Gorsuch, 1983, p. 133). In confirmatory factor analysis, statistical tests are performed to determine whether the sample data are consistent with imposed constraints. Using confirmatory factor analysis, the researcher estimates the extent to which the obtained data confirm a specified model, which is usually derived from substantive knowledge of the research under investigation and the conditions under which the data were collected.

There are several advantages of the confirmatory maximum likelihood factor analysis over the usual least-squares exploratory methods. It allows a test of the fit of the overall hypothetical factor model as well as the various parameter estimates within the model. It also allows the comparison of various competing hypothetical factor models. Furthermore, It is possible to assess the adequacy of the model(s) under investigation across populations. That is, with CFA the researcher can formulate a specific model and test the invariance of specific parameters in the factor solution. In addition, an incremental fit index can be computed to compare models and to decide on an acceptable model (Marsh and Hocevar, 1985, p. 565; O'grady, 1983, p. 827).

In sum, exploratory factor analysis should be used only for those areas that are truly exploratory, that is, areas where no previous analyses have been conducted. If previous analyses have been done, then either those results or a strong theory challenging those results should be used as a hypothesized structure in the new study (Gorsuch, 1983, p. 134).

LISREL Program and Covariance Structural Models

LISREL (LInear Structural RELations) developed by Joreskog and Sorbom (Joreskog, 1969, 1970, 1971; Joreskog & Sorbom, 1984) is a statistical tool for analyzing covariance matrices using systems of structural equations. The LISREL computer program assesses linear measurement and/or structural equation relationships among variables. The variables in the equation may be observed or latent variables.

The LISREL framework consists of two conceptually distinct models; a measurement model and a linear structural equation model. The measurement model specifies the relationships between the observed variables and the unobserved latent variables, and includes measurement errors. It describes the measurement properties of the observed variables such as reliabilities and validities. The structural equation model specifies the causal relationships among the latent variables and is useful in describing the amount of unexplained variance (Joreskog and Sorbom, 1989, p. 2). LISREL fits the observed covariance matrix to the structure implied by these two models and provides indices of how well the hypothesized model fits the actual data. The LISREL program estimates the unknown coefficients (parameters) in a set of linear structural equations and provides their standard errors and t-values.

The standard error is a measure of the precision of the parameter estimate. The t-value for a parameter is defined as the parameter estimate divided by its standard error. This can be used to examine whether the true parameter is zero. Parameters with t-values larger than two in magnitude are normally judged to be different from zero (Joreskog and Sorbom, 1989, p. 89).

The confirmatory factor analysis using LISREL is performed through four steps. The first step is the specification of a model involving formal definitions of the various components of the model and a statement of the assumptions. After the specification of the model, the identification of the model must be determined. Identification involves determining if there is a unique solution for the parameters of the model. Parameters of the model cannot be estimated in an unidentified model, leading to the respecification of the model. Once identification has been established, estimation can proceed. After estimation, an assessment of the fit of the model can be made which involves hypothesis testing and specification searching (Long, 1983, p. 17).

Estimation involves finding values of elements of (1) the matrix of coefficients relating endogenous variables to one another, (2) the matrix of coefficients relating exogenous variables to endogenous variables, (3) the factor loading matrix, (4) the covariance matrix of errors in the equation, (5) the covariance matrix of common factors, and (6) the covariance matrix of unique factors, which generate an estimated covariance matrix that is as close as possible to the sample covariance matrix. An exogenous variable is a variable whose variability is assumed to be determined by causes outside the causal model under consideration. No attempt is made to explain the variability of an exogenous variable or its relations with other exogenous variables. An endogenous variable is one whose variation is to be explained by exogenous and other endogenous variables in the causal model (Pedhazur, 1982, p.178). Endogenous variables are explained by specifying that they are causally dependent on other endogenous variables and exogenous variables.

The assessment of fit or the detection of lack of fit of a model is important in the application of LISREL. The estimated coefficients and the strength of associations require close examination. To aid the evaluation of a model, many statistical measures of fit have been proposed. The goodness of fit of the overall model can be evaluated by means of the following measures of overall fit; (1) Chi-square (X^2), (2) Goodness-of-fit index (GFI), (3) Adjusted goodness-of-fit index (AGFI), and (4) Root mean-square residual (RMR) (Joreskog and Sorbom, 1989, p. 43).

Virtually all measures of overall fit involve functions of the sample covariance matrix and the population covariance matrix. These fit indices gauge the closeness of the sample covariance matrix to the estimated sample covariance matrix. (Bollen, 1989, p. 256).

The first one is an overall X^2 measure and its associated degrees of freedom and probability level. If a model is correct and a sample size is sufficiently large, a X^2 measure is the likelihood ratio test statistic for testing the model. For large sample sizes, the goodness-of-fit statistic tends to be distributed as a X^2 variate. The degrees of freedom for X^2 is:

$$df = (1/2)p(p + 1) - t,$$

where p is the number of observed variables analyzed and t is the total number of independent parameters estimated.

GFI is a measure of the relative amount of variances and covariances jointly accounted for by a model. The AGFI adjusts for the degree of freedom of a model relative to variables. Both of these measures are between zero and one, and reach their maximum of one when the sample covariance matrix and the estimated sample covariance matrix are equal (Joreskog and Sorbom, 1989, p. 44; Bollen, 1989, p. 276). Unlike X^2, GFI is not dependent on the sample size and is relatively robust against departures from normality. Unfortunately its statistical distribution is unknown, so there is no standard to compare it with.

The root mean square residual (RMR) is a measure of the average of the residual variances and covariances, and can only be interpreted in relation to the sizes of the observed variances and covariances in the sample covariance matrix. A GFI of over .90 and a RMR of less than .05 are considered to represent an adequate fit (Hoffman, Mathieu and Jacobs, 1990, p. 948).

X^2 , GFI, AGFI, and RMR are measures of the overall fit of a model to data and do not express the quality of the model. If any of the overall measures indicates that the model does not fit the data well, it does not tell what is wrong with the model or which part of the model is wrong (Joreskog and Sorbom, 1989, pp. 44-45). A more detailed assessment of fit can be obtained by an inspection of standardized residuals and/or the modification indices.

A standardized residual is a fitted residual, the difference between the sample covariance matrix and the estimated sample covariance matrix, divided by its asymptotic standard error. In principle, each standardized residual is interpreted as a standard normal deviate and considered large when it is equal or greater than 2.58 in absolute value.

The modification indices (MI) are measures which are associated with the fixed and constrained parameters of a model. For each fixed and constrained parameter, the modification index is a measure of predicted decrease in X^2 if a single constraint is relaxed and the model is reestimated. The MI may be judged by means of a X^2 distribution with 1 degree of freedom. The fixed parameter with the largest MI is the one which will improve fit maximally when relaxed. The improvement in fit is measured by a reduction in X^2 which is expected to be close to the modification index (Joreskog and Sorbom, 1989, p. 45).

In LISREL output, an overall X^2 test of goodness-of-fit of a proposed model to data is provided. However, this X^2 test of goodness-of-fit is dependent on the sample size. A X^2 value obtained for a given model, being a joint function of sample size and incongruence between the model and the data, will generally lead to the rejection of the null hypothesis that the population covariance matrix is equal to the reproduced population covariance matrix on a statistical basis when the sample size is large regardless of whether or not the difference between two matrices is trivial. That is, the statistical power of the X^2 goodness-of-fit test associated with a large sample may easily lead to the rejection of a theoretically useful model that is closely aligned with the sample covariance structure (Hoelter, 1983, p. 328).

Therefore, researchers using X^2 test statistic should be always aware of the fact that a X^2 measure is sensitive to sample size and very sensitive to departures from multivariate normality of observed variables. Large sample sizes and departures from normality tend to increase X^2 over and above what can be expected due to specification error in a model (Joreskog and Sorbom, 1989, p. 43).

As a result, alternative indices of fit have been developed to examine the triviality of a difference between the hypothesized model and the actual data. These indices include Tucker and Lewis' (1973) incremental index, Bentler and Bonett's (1980) Normed Fit Index (NFI) and Non-Normed Fit Index (NNFI), Wheaton et al.'s (1977) relative X^2, and Carmines and McIver's (1981) ratio of X^2 to degrees of freedom.

Especially, Bentler and Bonett (1980) developed a psychometric index of incremental fit representing the relative increase in precision of estimation of a hypothesized model over a quite restricted model or the null model. NFI can be used along with the goodness-of-fit test to aid both in the process of model confirmation and model comparison. The value of NFI lies between zero and one (O'Grady, 1983, p. 828). Wheaton et al. (1977) suggested a ratio of approximately 5 or less as beginning to be reasonable fit. Carmines and McIver's (1981) ratio of X^2 to degrees of freedom associated with the model should be less than 3 for nonsignificance. Through the application of these kinds of procedures, researchers test the goodness-of-fit of models, diagnose problems with models, and fix or constrain parameters.

In sum, the closeness of the sample covariance matrix to the estimated sample covariance matrix can be measured in many ways. The one most amenable to tests of statistical significance is the X^2 estimator where the null hypothesis is that there is no significant difference in the population covariance matrix and the estimated population covariance matrix. The other overall fit measures provide largely descriptive measures of fit since their distributions are unknown (Bollen, 1989, p. 281).

Factor Analytic Studies of the Multidimensional Aptitude Battery

Since the Multidimensional Aptitude Battery was developed rather recently, only a few factor studies have been reported. Two factor analyses, using two different samples, of the MAB were reported in the manual (Jackson, 1984, pp. 43-46). Of the two factor analyses, the first one, using a principal components factor analysis based on the intercorrelations between subtest raw scores of 3,121 male and female high school students between the ages of 16 and 19, yielded two clearly identifiable factors after the first general factor. All subtests had moderate to high first-factor loadings indicating the MAB had a strong general intelligence (or g) factor underlying it. The two rotated factors were identified as the Verbal and the Performance factors. For the second factor analysis, based on 516 high school students, two distinct Verbal and Performance factors were also yielded.

Other studies used a hierarchical factor solution with a sample of convicted felons. Two studies of the construct validity of the MAB for convicted felons provided considerable support for Jackson's (1984, p. 48) assertion that the battery measured the same ability dimensions as the WAIS-R.

Wallbrown, Carmin, and Barnett (1988) performed the construct validity study of the MAB using a hierarchical factor solution with 300 male felons referred to a reception and diagnostic center for psychological assessment. Intercorrelations among the 10 subtests of the MAB were computed and subjected to the Wherry (1959) and the Wherry and Wherry (1969) hierarchical factor analyses. Based on previous studies of Wechsler scales (Blaha and Wallbrown, 1982), the extraction of two factors at the primary level and one higher-order factor was specified. The results provided a factor structure consisting of a strong general ('g') factor and two primaries defined respectively by the Verbal and Performance subtests. Wallbron, Carmin, and Barnett (1989) obtained similar results using the same method with a second sample of 300 male offenders from the same state agency.

Lee, Wallbrown, and Blaha (1990) conducted a hierarchical factor analysis of subtest intercorrelations of the MAB to examine the construct validity. Data were obtained from the test manual reported by Jackson (1984, p. 45). A hierarchical factor solution was used on 10 subtest intercorrelations. Based on the results of the Wallbrown, Carmin, and Barnett's (1988, 1989) studies, factorization was controlled by specifying the extraction of two factors at the primary level. Higher order factorization was controlled by specifying the extraction of one factor. The results showed a strong general ('g') factor defined by all 10 subtests as well as the Verbal factor defined by the five verbal subtests and the Performance factor defined by the five performance subtests. Complete bifurcation between the verbal and performance subtests was found when the g-variation was removed. A hierarchical arrangement of abilities consisting of a strong 'g' factor and two primaries corresponding to the verbal-education (v:ed) and spatial-perceptual-mechanical (k:m) parameters from Vernon's (1950) structural model was obtained. This ability arrangement provided the construct validities for the Full Scale IQ, Verbal IQ, and Performance IQ of the MAB.

Stockwell (1984) undertook principal components factor analyses on the MAB and the WAIS-R separately to compare the underlying structure of abilities represented by the two tests. He extracted the Verbal and Performance factors from intercorrelations among the WAIS-R subtests reported in the WAIS-R manual. After two factors were extracted and rotated by varimax method, coefficients of factor congruence between the WAIS-R factors and the MAB factors were calculated. These congruence coefficients between the two tests were .97 and .96 for the Verbal factor and the Performance factor, respectively. From these analyses it was concluded that the patterns of abilities measured by the MAB and the WAIS-R were very similar in spite of such differences between the two tests as their content, and administration and response formats (Jackson, 1984, p. 48).

However, the construct validity study conducted by Kranzler (1991) reported different results. Results of his hierarchical factor analysis with 101 university students did not support the construct validity of the MAB on the first-order factor found in the Wallbrown et al.'s study (1988, 1989). Re-analyzing the Wallbrown et al's (1988, 1989) results, he underlined the importance of ascertaining the reading level prior to the administration of the MAB and insisted that marginal reading proficiency, especially on a timed paper-pencil test such as the MAB, would confound results. He reported that although the Full Scale IQ of the MAB seemed to be a valid measure of general mental ability, the validity of the Verbal and Performance IQs was not supported and concluded that the MAB did not appear to measure the identical pattern of abilities measured by the WAIS-R.

CHAPTER III

METHODOLOGY

This study was designed to examine the psychometric properties of the Multidimensional Aptitude Battery (MAB). This was done in an effort to factor analyze the MAB through a confirmatory factor analytic technique. Through this chapter, instrumentation used and actual statistical procedures to address the research questions and corresponding null hypotheses are described.

Instrumentation: The Multidimensional Aptitude Battery (MAB)

The Multidimensional Aptitude Battery (MAB) is a multiple choice test of intelligence and mental abilities patterned quite closely after Wechsler Adult Intelligence Scale-Revised (WAIS-R). According to Jackson (1984, p. 5), Wechsler's scales have been extraordinarily successful for a number of reasons: (1) their incorporation of a diversity of tasks including not only verbal and school learned content, but performance and practical skills as well; (2) content appropriate for adolescents and adults, as well as children; (3) their reflection of fresh conceptions of the nature of intelligence in which psychotic processes, neurological damage, or emotional disturbances might affect performance; (4) their extremely careful and thorough standardization; (5)

their high technical psychometric quality; and (6) their substantial validity. In fact, the Wechsler scales have become the standard against which other tests of intelligence have been appraised. However, a major drawback of the Wechsler scales is that they require individual administration and scoring by a specially-trained professional. Jackson (1984, p. 6) argued that for the majority of examinees, individual administration is costly and unnecessary.

Thus, the MAB was designed under its author's intention to evaluate the degree to which it is possible to incorporate some of the widely acknowledged positive features of individually-administered multi-scale tests, such as the WAIS-R, into a structured format which allows group administration, automated administration, and convenient hand or machine scoring (Jackson, 1984, p. 6). Because of much less expensive cost in administration and available results comparable to those of the more time-consuming individual tests, in educational situations, group aptitude tests are used far more extensively than individually administered intelligence tests. Although in many schools the actual test may be administered by a counselor or someone else with special advanced training, most group tests are designed so that any teacher with a minimum of training should be capable of the administrative task (Mehrens & Lehmann, 1984, p. 380).

It took over 10 years to publish the final version of the MAB, during which several revisions were done. The structure of the MAB is similar to that of the WAIS-R in terms of its scales and subtests. Like the WAIS-R, the MAB consists of two subscales, Verbal and Performance, each comprising five subtests, and is scored to provide the Verbal, Performance, and Full Scale IQ scores. It is designed to be used with adolescents and adults ranging in age from 16 to 74 years. The age groups are identical to those employed in the standardization of the WAIS-R. Therefore, there are nine different age groups (16-17, 18-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65-69, and 70-74). Five Verbal subtests are contained in one booklet, and five Performance subtests in another. Separate answer sheets that may be either hand or machine scored are provided for each booklet. Nine of the subtests have the same name as in the WAIS-R (Digit Span is omitted and Spatial replaces Block Design), but all items are original and a multiple-choice format is employed throughout. Examiner's instructions are standardized, so that the test can be administered by cassette. Seven minutes are allowed for each subtest and each half of the battery requires about fifty minutes for test administration to be completed.

The ten subtests which have the same name as the corresponding WAIS-R subtests with one exception are as following (Vernon, 1985).

Verbal Scale

1. Information: a 40-item test of general knowledge and information.

2. Comprehension: a 28-item test measuring a person's understanding of various social situations and conventions, and why things are done the way they are.

3. Arithmetic: a 26-item test, ranging in difficulty from simple addition and subtraction to complex problems involving fractions, percentages, and numerical reasoning.

4. Similarities: a 34-item test in which subjects must decide in what way pairs of words are alike. The items range in complexity from concrete to relatively abstract associations.

5. Vocabulary: a 46-item test in which subjects select words that have the same meaning as the test items.

Performance Scale

1. Digit Symbol: This 35-item test measures the examinee's ability to match symbols to numbers. At the top of each page, a coding chart is printed in which the digits 1 to 9 are matched (or coded) with different symbols. In each item of the test, a combination of the coded symbols appears (from 1 to 9 symbols in length) followed by five strings of digits. Only one of the digit strings completely matches the order of the symbols and it is the subjects' task to select the correct digit string.

2. Picture Completion: a 35-item test in which pictures of common objects are presented, each missing one important part or component. Below each picture are the first letters

of five possible missing parts and the subjects' task is to select the correct letter.

3. Spatial: a 50-item test whose purpose is to find out how well the examinee can see differences in figures. In this subtest subjects must perform spatial rotations of figures and select one of five possible rotations presented as their answer.

4. Picture Arrangement: a 21-item test in which subjects must mentally rearrange cartoon panels so that they tell a sensible story. The panels are numbered and the subjects' task is to select the correct sequence of numbers that conforms to the order in which they think the panels should be arranged.

5. Object Assembly: a 20-item test in which parts of well-known objects are placed in the wrong order. The examinee should first identify the objects and then mentally rearrange their parts into the correct order, from left to right, in which these parts should be placed to form the object. Examinees select a sequence of numbers that conforms to the order they have decided on.

Internal consistency reliabilities of the MAB were computed separately for 230 male and 285 female adolescents at each year level from 15 to 20. These reliabilities range in different age groups from .94 to .97 for the Verbal Scale, .95 to .98 for the Performance scale, and .96 to .98 for the Full Scale. For test-retest reliabilities, 52 young adult psychiatric patients were tested on two occasions, separated by an average of 45 days. Stability of the MAB is revealed by the test-retest correlations. Individual subtest Verbal Scale test-retest reliabilities ranged from .83 to .97 with a median of .90 while individual subtest Performance Scale reliabilities ranged from .87 to .94 with a median of .93. The test-retest reliabilities for the Verbal Scale, Performance Scale, and Full Scale scores were .95, .96, and .97, respectively (Jackson, 1984, pp. 39-42).

A major source of validity data for the MAB is sets of correlations with the WAIS-R. The MAB and the WAIS-R were administered individually or in small groups to the sample of 145 subjects ranging in age from 16 to 35 years. The median Verbal Scale subtest MAB/WAIS-R correlation is .82 and the median Performance Scale subtest correlation is .65 (ranging from .44 for Spatial/Block Design to .89 for Arithmetic and Vocabulary). The Verbal, Performance, and Full Scale scores for the MAB correlate .94, .79, and .91 with the corresponding IQ scores on the WAIS-R, respectively. Obviously, the MAB and the WAIS-R are highly correlated even though the MAB is different from the WAIS-R with the individually administered format and shares no items in common (Jackson, 1984, pp.47-48).

Data Collection

Data for this study were obtained from the 1990 norming study (Jackson, forthcoming) of the MAB.[1] They were collected between 1986 and 1989 by testing a number of groups covering a wide geographical area in the United States, with representation also from Canada.

Subjects' participation was voluntary and they were paid. The total sample size is 1600, approximately evenly divided by gender. The ratio of White to Non-white is approximately 6.7 to 1 and the subjects were selected from six arbitrary regions of the United States and Canada (North West, South West, Mid West, South East, North East, and Canada). Most of the subjects were American; Canadian subjects represented sixteen percent of the total. Average educational attainment (for sample over 25-yrs. old) was between high school graduate and the freshman year in college.

Subjects were divided into nine age groups (16-17, 18-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65-69, and 70-74 years old). Seven of the groups had 200 subjects, while two had 100 subjects. Only the groups which contained 200 subjects were selected for this research because maximum likelihood procedures have been found to be robust for this sample size (Fuller & Hemmerle, 1966). Therefore, a total of 1400 subjects were used for this research.

Model Specification

[1] The data for this research is not yet published by the author of the MAB. Please, contact the author of the MAB for more information on the data.

The specification of the proposed factor model is accomplished by freeing, fixing, or constraining elements in three matrices; (1) the factor loading matrix (lambda), (2) the factor variance-covariance matrix (phi), and (3) a diagonal matrix of error/uniqueness parameters for each subtest (theta delta).

For this study of confirmatory factor analysis using LISREL, a model that matched the actual organization of the MAB and the results of previous studies was specified. The common factor model which has been found to be appropriate representation of the MAB is the two factor model. Therefore, the two factor model was specified for this study. In the two factor model, the two common factors are the Verbal factor and the Performance factor.

All coefficients with the value of 0 or 1 were fixed according to *a priori* theoretical considerations. The factor loadings were restricted so that each subtest loaded only on the factor which was hypothesized to represent. For the proposed two factor model, the subtests of Information, Comprehension, Arithmetic, Similarities, and Vocabulary were specified as loading only on the Verbal factor; the subtests of Digit Symbol, Picture Completion, Spatial, Picture Arrangement, and Object Assembly were specified as loading only on the Performance factor. All other factor loadings were constrained to zero (see figure 1).

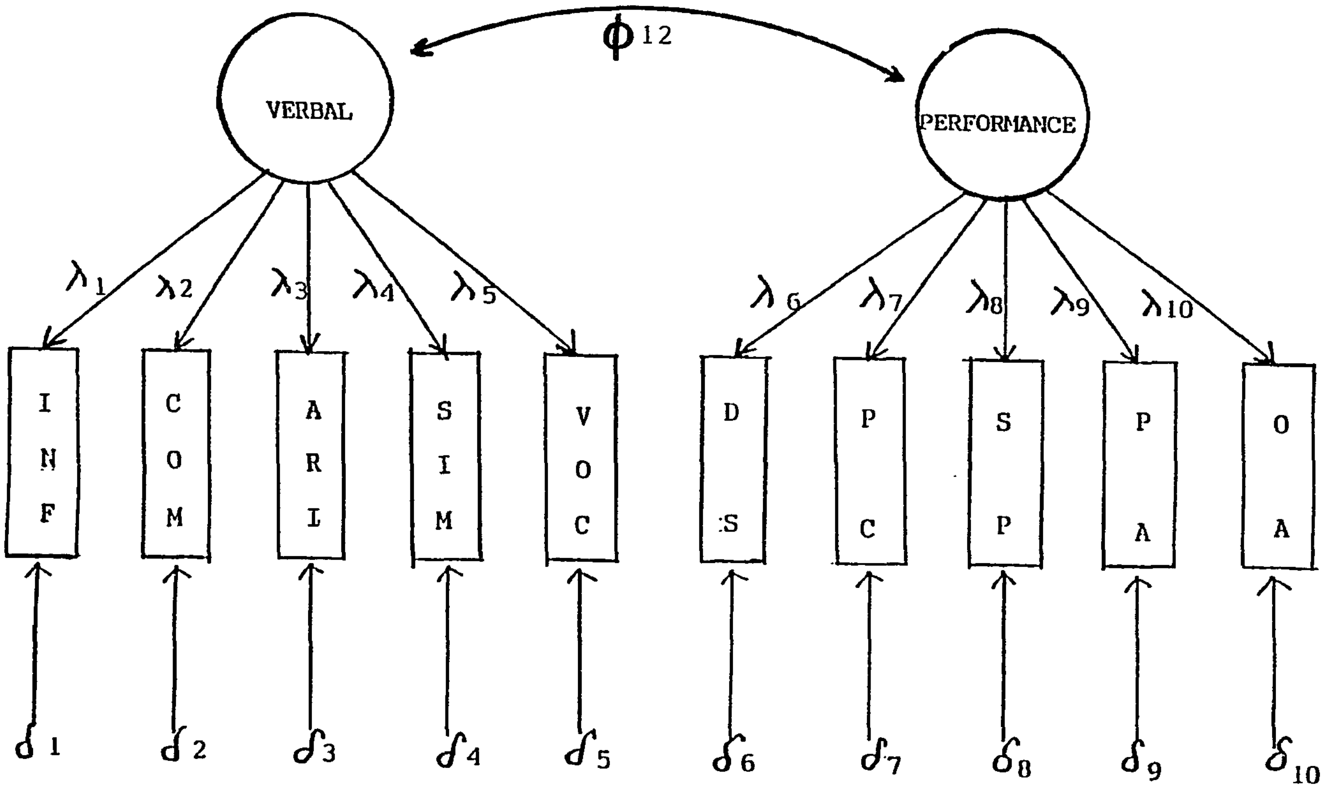
In addition, factors were specified to correlate with each other, because each factor was assumed to also reflect a 'g' or general intelligence component. The information about correlations among the factors provided information relevant to the higher-order factors (Jackson, 1984; Lee & Wallbrown, 1990; Wallbrown, Carmin & Barnett, 1988). Therefore, it was expected that there was a significant correlation between factors.

The error/uniqueness components of subtests, which were the diagonal values of the theta delta matrix, were set free. But the off-diagonal elements were set fixed to zero because of the assumption of no relationships among errors.

Other assumptions were also implied. First, the measurement errors were assumed to be independent from factors. Second, the observed indicators (subtests) were assumed to have multinormal distribution (Joreskog and Sorbom, 1989). Under these assumptions, the covariance matrices were used to evaluate the parameters in the model through the maximum likelihood procedure.

When analyzing multiple groups it is necessary to use covariance matrices and to scale each of the latent variables in a similar fashion (Joreskog, 1971; Joreskog and Sorbom, 1989). Therefore, for the present study the analyses comparing seven age groups used covariance matrices with the scaling of each factor by fixing one of factor loadings which loaded on the factor to 1.00. Furthermore, the same loading was fixed for each factor in seven age groups, thereby assigning the identical metric to the factors.

Thus, twenty one parameters for each age group--eight



INF : Information
COM : Comprehension
ARI : Arithmetic
SIM : Similarities
VOC : Vocabulary

D S : Digit Symbol
P C : Picture Completion
S P : Spatial
P A : Picture Arrangement
O A : Object Assembly

λ : Factor loading
Φ : Correlation between the latent variables
δ : Error associated with each observed variable

Figure 1 : The Specification of the Two Factor Model

factor loadings, three coefficients in the factor variance-covariance matrix, and ten parameters in error/uniqueness matrix--were estimated. Therefore, a total of 147 free parameters were estimated for the two factor model across seven age groups.

Statistical Analyses

The purpose of this study was to examine whether the two factor model explains the MAB factor structure for the American and Canadian samples of the 1990 norming study and to test factorial invariance across groups. For the statistical examination of the research questions, confirmatory maximum likelihood factor analyses were performed by entering the covariance matrices into the LISREL VII (Joreskog & Sorbom, 1989).

Confirmatory factor analysis is the proper technique to determine which factor analytical model provides the superior fit to a set of data (Long, 1983). With a confirmatory factor analysis researchers can test the tenability of a hypothesized factor structure of data. Based on the results of the statistical test, it is determined whether the hypothesized model will be rejected or not.

Maximum likelihood factor analysis provides statistical estimates of parameters and significance tests of goodness-of-fit of hypothesized models. Confirmatory maximum likelihood factor analysis identifies the population parameters with a maximum likelihood of generating the observed sample distribution (Gorsuch, 1983, pp. 127-141; Tabachnick and Fidell, 1989, pp. 626-627).

Major Research Questions

The major research questions were as following:

(1) Research Question 1:

Does the two factor model explain the variation in the MAB subtest scores for all groups simultaneously?

(2) Research Question 2:

Are there any significant differences across the age groups with respect to factorial invariance?

Hypotheses and Statistical Analyses

In order to examine the research questions, null hypotheses were posited. All hypotheses were tested by using the covariance matrices from ten subtests scores of the MAB for each age group.

Usually, the measurement model can be examined for goodness-of-fit using several different criteria. The most frequent approach is to use the X^2 statistic to determine whether the underlying structure of the actual data is explained by the proposed model, and which model better explains the relationships among the factors and subtests.

Because this measure has the shortcoming of being sample size dependent (Bentler and Bonett, 1980, p. 591), many other procedures have also been proposed to assess the degree of fit and significance of the model. Those strategies used for this research included: Wheaton et al.'s (1977) ratio of X^2 to degree of freedom; and Bentler and Bonett's (1980) Normed Fit Index (NFI) and Non-Normed Fit Index (NNFI). However, there is much disagreement in the literature regarding the best method to determine the fit between the measurement model and the original covariance matrix. Any criterion used in decisions regarding what constitutes a trivial difference has tentative nature (Hoelter, 1983, p. 331). Therefore, the safest recommendation is to report a X^2 estimate along with several of the other fit indices (Bollen, 1989, p. 281). Accordingly, the best method to evaluate the goodness-of-fit seems to combine several of these criteria.

Therefore, in this study, for the assessment of goodness-of-fit, the X^2 statistic was reported along with Wheaton et al's (1977) ratio of X^2 to degrees of freedom and Bentler and Bonett's (1980) NFI and NNFI. Especially, the cutoff point of 5 for X^2 ratio to df and that of .85 for NFI and NNFI were used. Bentler and Bonett's (1980) NFI and NNFI may be likened to the concept of practical significance, and it provides a way to compare one model directly with another by measuring improvement of fit. The overall X^2, on the other hand, gives an indication of statistical significance (Gridley and McIntosh, 1991, p. 240).

X^2 measures of overall fit and standard errors used to estimate the statistical significance of each parameter estimate were reported. The probability level, 'p', of a X^2 value was also given. The probability level of a X^2 is the probability of getting a X^2 value larger than the value actually obtained, given that the hypothesized pattern is true. Thus, small values of 'p' correspond to poor fit and large values to good fit (Joreskog, 1969, p. 195).

Prior to the actual analyses of research questions, a test was carried out to examine whether the sample covariance matrix resulted from random sampling errors. The null hypothesis for this test was that there was a single source of variation in each of the subtests, that is, the disturbance term, and that no common sources of variation exist for the subtests in the population. This analysis yields a X^2 value with corresponding degrees of freedom. If the obtained X^2 is significant, the null hypothesis of the sample covariance matrix due to sampling errors alone is rejected and the sample covariances reflect real covariation among the subtests in the population (Alwin and Jackson, 1980, p. 83).

Hypotheses for research question 1

For research question 1, the following null hypotheses were posited and simultaneous confirmatory factor analyses for all groups were conducted to investigate whether the two factor model explained the MAB subtest variation.

Ho1.1 There is no significant difference between the null model which represents no relationships among the subtests and the models that represents the actual data in all groups.

Ho1.2 There is no significant difference between the two factor model and the model that represents the actual data.

Ho1.3 There is no significant difference in explanatory power between the two factor model and the null model.

Each of these analyses had an associated X^2 value. Each X^2 was used to assess the fit of the model to the observed data and to determine whether there were differences between models, which was related to a X^2 difference test. Because the difference between two X^2 values is distributed as a X^2 variate with degrees of freedom corresponding to the difference between the degrees of freedom of the two models under examination, the various models, as proper subset of each other, can be tested to determine if a given model fits the data significantly better than some other model. If a significant difference exists between any pair of X^2 values, then one model is considered better able to explain the data and provides a better fit for the data (O'grady, 1983, p. 828). Like the usual X^2 test, a X^2 difference test has higher power to detect a false model in large samples than in small ones (Bollen, 1989, p. 292). Therefore, other supplementary tests were obtained such as Wheaton et. al's (1977) ratio of X^2 to degrees of freedom, and Bentler and Bonett's (1980) NFI and NNFI.

(1) Hypothesis 1.1

For Ho1.1, a multiple group confirmatory factor analysis was conducted by simultaneously entering the covariance matrices for the ten MAB subtests for each age group. In this step, the null model was tested. The null model is the model which has zero common factors and is regarded as the worst possible model with no paths or relations among subtests. Since many studies on the MAB factor structure (Jackson, 1984; Lee, Wallbrown and Blaha, 1990; Stockwell, 1984; Wallbrown, Carmin and Barnett, 1988 & 1989) provided that there were underlying relationships among the subtests under examination, this null model was expected to be rejected. Wheaton et al.'s (1977) ratio of X^2 to degree of freedom was also obtained to assess the degree of fit and significance of the model.

(2) Hypothesis 1.2

Ho1.2 was examined to determine whether the two factor model explained the relationships in the actual data. The unknown parameters of the two factor model were estimated for all groups simultaneously using LISREL procedures.

This multiple group confirmatory factor analysis provided an associated X^2 value which was used to assess the fit of the two factor model to the actual data. Wheaton et al.'s (1977) ratio of X^2 to degrees of freedom was also provided.

(3) Hypothesis 1.3

For the test of Ho1.3, the two factor model was tested against the null model. A statistical test of the difference between the X^2 value for the null model obtained from the test of Ho1.1 and the X^2 value for the two factor model obtained from the test of Ho1.2 was conducted. If this X^2 difference value is significant, it indicates that the two models are statistically different and one of the two models provides a better fit for the data.

In addition, Wheaton et al.'s (1977) ratio of X^2 to degrees of freedom as well as Bentler and Bonett's (1980) NFI and NNFI were also obtained. By using Bentler and Bonett's (1980) NFI, an examination of the additional amount of covariance which a proposed model can explain is provided. In the confirmatory factor analysis, the null model (i.e., a model with zero common factors) serves as the model against which the various hypothesized models are tested (Joreskog and Sorbom, 1980).

Hypotheses for research question 2

The purpose of research question two was to assess the potential differences in the factorial structure of the MAB among age groups. Statistically, the problem of making meaningful comparisons across groups involves the issue of factorial invariance. The issue of factorial invariance involves the degree of equivalence in the corresponding parameters of the models across populations (Alwin and Jackson, 1980, p. 100).

There are two subsets of issues in factorial invariance.

The first is related to the invariance of the factor pattern and the invariance of the measurement error variance, that is, measurement invariance. The second involves the invariance of the covariance structure of the underlying factors and the means of these factors, that is, factor space invariance (Alwin and Jackson, 1981, p. 253; Sorbom, 1974, p. 230).

The following null hypothesis was for research question two.

Ho2 There is no significant difference across the age groups with respect to factorial invariance.

For this research question, the two factor model was tested for invariance across the age groups. Generally any degree of invariance can be tested, from one extreme where all parameters are assumed to be invariant across groups to the other extreme where there are no constraints between groups (Joreskog, 1971, p. 410).

The factorial invariance was analyzed by testing a series of nested hypotheses involving the equivalence constraints of one or more parameter matrices for the age groups. Provided that models can be framed so as to be hierarchical or nested, that is, with one model able to be considered as a specialization of another model, both generalized least squares estimation and maximum likelihood estimation provide for a X^2 difference test that evaluates the statistical significance of the parameters that differentiate between two competing models (Bentler and Bonett, 1980, p. 592). Since the covariance matrix is a function of factor loadings matrix, the covariance matrix of factors, and the covariance matrix of measurement errors, the factorial invariance of the two factor model across the age groups was examined in terms of the equivalence of these parameter matrices.

Parameters were compared across the age groups by examining the degree of fit of models with different equivalence constraints (Joreskog, 1971, pp. 418-420). All five different models were evaluated across the age groups. Model 1 was a preliminary conclusive test for the invariance of three parameter matrices of the model involving the hypothesis of equality of covariance structure of the subtests across the age groups. Model 2 tested the invariance of the basic model form across all age groups. This model included no between-group equivalence constraints. That is, it was related to the test of the hypothesis of equality in the number of common factors (no constraints). Model 3 involved testing the invariance of the basic model and factor loadings across the age groups (equivalence constraints on factor loadings), and model 4 examined the invariance of the basic model, factor loadings, and measurement errors across the age groups. Finally, model 5 tested the invariance of the basic model, factor loadings, measurement errors, and factor covariances across the age groups. Models 2 through 5 represented different types of invariance and constituted a hierarchy. Implied in this hierarchy is that invariance in one parameter matrix (e.g., factor pattern matrix) carries a higher premium than that in other ones (e.g., the covariance matrix of measurement errors) (Liang et al., 1986, p. 28).

For the present study, the hierarchy suggested by Joreskog (1971) was used. The simultaneous factor analysis was undertaken by using the LISREL VII program (Joreskog and Sorbom, 1989). Several measures were used to assess the goodness-of-fit of these models. These were X^2 test which assessed the fit of the model to the data, the significant test of X^2 differences among models with the various equivalence constraints, Wheaton et al's ratio of X^2 to degrees of freedom, and Bentler and Bonett's Normed Fit Index (NFI).

Examining the issue of factorial invariance involved testing a series of sub-hypotheses. One began by testing the sub-hypothesis of equality of covariance structures of the subtests across groups as a general preliminary test for factorial invariance. This analysis produces a X^2 value.

When the hypothesis is not rejected, the issue is essentially settled. In other words, failure to reject the hypothesis of equality of covariance structures means that the parameters are invariant across the age groups. However, if the hypothesis is rejected, specific forms of invariance must be explored and one continues to test invariance hypotheses in order to specify the source responsible for the unequal observed covariance structures. For this purpose, a sequence of hypotheses, such that each hypothesis is a special case of the preceding, are considered (Joreskog, 1971, p. 419).

The second sub-hypothesis tested was the hypothesis of equality in the number of common factors. Namely, no between group constraints model (the basic model) was tested (Model 2). This analysis yields a X^2 value equal to the sum of the X^2 values for each group with degrees of freedom equal to the sum of the degrees of freedom for each group. If the hypothesis of a common number of factors is found tenable, the test of various between group constraints is continued.

The third sub-hypothesis was the hypothesis of the invariant factor pattern across the age groups. For this hypothesis, it was assumed that the number of factors in each group was known *a priori* and that these factors were held in common across all groups.

In order to examine this hypothesis, the estimation of a model constraining a common factor pattern across all age groups was required (Model 3). Then this model was compared with the model having no between-group constraints on the factor pattern across the age groups (Model 2). From these two models, two separate X^2 values were obtained. With these two X^2s, one from the no constraints model and the other from the invariant factor pattern model, a X^2 difference test was conducted to examine whether the sub-hypothesis of equality of factor pattern across the age groups was tenable. If a X^2 value from the X^2 difference test is significant, the hypothesis of invariant factor pattern may be rejected. If it is not significant, the invariant factor pattern is suggested and the test of next hypothesis of the invariant covariance structure of measurement errors across the age groups is conducted.

The sub-hypothesis of equality of the covariance structure of measurement errors was tested on the failure to reject the hypothesis of invariant factor pattern. This sub-hypothesis was related to Model 4, the invariance of the basic model, factor loadings, and measurement errors across the age groups.

The test of model 4 yields a X^2 value. Then, the sub-hypothesis of invariance in measurement errors is examined using a X^2 difference test between the X^2 value obtained from the test of model 4 and the test of model 3. If this X^2 value is significant, the sub-hypothesis of invariant covariance structure of measurement errors may be rejected. However, if it is not, invariance in the covariance structure of measurement errors across the age groups is suggested.

Finally, if the sub-hypothesis of invariant measurement errors was found tenable one continues to test the sub-hypothesis of invariance in factor covariance (Model 5). Joreskog (1971, p. 420) notes that this hypothesis is included in the hypothesis of equality of covariance structures of the subtests across the age groups (Model 1), but it is a stronger hypothesis since it specifies a particular form of invariance.

The same procedures for the previous hypotheses testings were used for this hypothesis testing. The test of model 5 yielded a X^2 value. Then, the sub-hypothesis of invariance in factor covariance was examined using a X^2 difference test between the X^2 values obtained from the test of model 5 and the test of model 4. If this X^2 value is significant, the sub-hypothesis of invariant factor covariance may be rejected. However, if it is not, invariance in factor covariance across the age groups is suggested.

Furthermore, Wheaton et al.'s (1977) ratio of X^2 to degrees of freedom and Bentler and Bonett's (1980) NFI were also calculated to estimate the fit of each model to the data and incremental fit under the consideration of the effect of a large sample size on X^2 tests and X^2 difference tests (Bollen, 1989, p. 292).

CHAPTER IV

RESULTS

Results of the Study

This chapter presents the results of the data analyses for the study. The purpose of this study was to examine the psychometric properties of the Multidimensional Aptitude Battery (MAB), especially, to factor analyze the data from the 1990 norming study through confirmatory factor analytic technique in order to investigate the extent to which the two factor model explained the variation in the MAB subtests and to examine the developmental invariance of the factor structure of the MAB across age groups.

Whereas previous exploratory factor analyses generated possible models, the present confirmatory factor analysis evaluated the relative fit of *a priori* specified models for independent samples. This allowed a statistical comparison of models.

In order to perform this study, two research questions were addressed:

1. Does the two factor (Verbal and Performance factors) model explain the variation in the MAB subtest scores for all groups simultaneously?

2. Are there any significant differences across the age groups with respect to factorial invariance?

The sample consists of a total of 1400 males and females approximately evenly divided by gender. Only the groups which contained 200 subjects were selected from a total of nine groups for this research because maximum likelihood procedures have been found to be robust for this sample size.

The analyses related to each research question will be presented below, followed by additional analyses that are appropriate in addressing questions.

For the assessment of the goodness-of-fit of the two factor model, the following criteria were used: (1) the X^2 statistic generated for the model by the LISREL program (Joreskog and Sorbom, 1989); (2) the ratio of X^2 to degrees of freedom suggested by Wheaton et al. (1977); and (3) the incremental fit indices [Normed Fit Index (NFI) and Non-Normed Fit Index (NNFI)] using the null model as a base line model (Bentler and Bonett 1980). In using these goodness-of-fit indices, if all criteria did not converge, more emphasis was placed on the NFI and NNFI. All competing models were evaluated using the X^2 difference tests with a non-significant X^2 indicative of congruence between the models (Bollen, 1989).

Prior to the analyses of research questions, a test was carried out to examine whether the sample covariance matrix resulted from random sampling errors for each group. This analysis yielded X^2 values for seven groups. They varied between 828.30 for the age group 1 and 1531.34 for the age group 5 with 45 degrees of freedom. These X^2 values all represented significant differences from zero leading to the rejection of the null hypothesis of sample covariance matrix due to sampling errors alone, and suggested that the sample covariances for each group reflected real covariations among the subtests in the population. Table 1 reported the results.

Research Question 1

Ho1.1 The result of the test of Ho1.1 was obtained by employing multiple group confirmatory factor analysis. The covariance matrices of seven age groups were simultaneously entered into the LISREL VII (Joreskog and Sorbom, 1989). The goodness-of-fit of the null model was evaluated by using several measures. The test of the null hypothesis of zero common factors yielded a statistically significant X^2 value of 9471.87 with 315 degrees of freedom. This result indicated that there was a statistically significant difference between the null model of no underlying relations among subtests and the model representing the actual data. Therefore, the null model did not explain the actual data implying that a model with specified relationships would better explain the actual data.

Considering the possible effects of a large sample size on the X^2 test, other index for the assessment of overall fit of the null model to the data was also provided. Wheaton et al's (1977) ratio of X^2 to degrees of freedom was 30.07. According to Wheaton et al's (1977), the ratio of X^2 to degrees of freedom in the range of 5 to 1 or less is

Table 1 : Results of Sampling Adequacy of Each Age Group

Group Age	X^2	df	X^2/df	GFI**	RMR***
G1:16-17	828.30*	45	18.41	.42	9.24
G2:18-19	1464.34*	45	32.54	.25	19.05
G3:20-24	1462.81*	45	32.51	.25	18.58
G4:25-34	1465.38*	45	32.56	.24	19.77
G5:35-44	1531.54*	45	34.03	.23	21.82
G6:45-54	1377.07*	45	30.60	.26	18.99
G7:55-64	1342.43*	45	29.83	.27	18.06

(N=200 for each group)

* $p < .05$

** GFI : Goodness-of-Fit Index

*** RMR : Root Mean-square Residuals

indicative of an acceptable fit between the hypothetical model and the sample data. Therefore, the obtained ratio indicated that a statistically significant difference existed between the null model and the model explaining the actual data. Table 2 reported the result.

Ho1.2 The goodness-of-fit of the two factor model was evaluated by using several measures. The analysis of the two factor model yielded the X^2 value of 601.16 with 238 degrees of freedom and the probability of the two factor model was near zero. This result indicated that the two factor model did not fit the data and there was a statistically significant difference between the two factor model and the model representing the actual data. However, considering the possible influence of a large sample size on the X^2 test, Wheaton et al's (1977) ratio of X^2 to degrees of freedom was also provided. A ratio of 2.53 indicated that the two factor model was able to explain the actual data. That is, there was no significant difference between the two factor model and the model representing the actual data for all age groups.

Tables 3, 4, and 5 present the actual lambda, phi, and theta-delta values of unstandardized and standardized maximum likelihood estimates for the two factor model for each age group and total groups. The lambda value represents the correlation between observed variables (subtests) and latent variables (factors). It is a factor loading. The standardized lambda is interpreted as the expected change in the observed variable (subtest) associated with a change of

Table 2 : Measures of Fit of Models for Total Groups

Model	X^2	df	X^2/df	X^2 difference test X^2	df	X^2/df	NFI**	NNFI***
Null Model	9471.87*	315	30.07					
				8870.71*	77	115.20	.94	.95
Two Factor Model	601.16*	238	2.53					

(N=1400)

* p <.05

** Normed Fit Index

*** Non-Normed Fit Index

one standard deviation in the latent variable (factor) (Bollen, 1989, pp. 199-200).

All values of standardized lambda were greater than .45 and were significant at .05 level. The largest was .92 for Similarities subtest of group 5 and the smallest was .45 for Digit Symbol subtest of group 1. The lambda values indicated that the first five subtests of Information, Comprehension, Arithmetic, Similarities, and Vocabulary loaded significantly on the Verbal factor, and the remaining five subtests of Digit Symbol, Picture Completion, Spatial, Picture Arrangement, and Object Assembly loaded significantly on the Performance factor as expected. The factor loadings for the Verbal factor ranged from a low of .54 to a high of .92 with a median of .85, while those for Performance factor varied from a low of .45 to a high of .85 with a median of .76.

The phi values indicative of the correlation between factors were also reported. The correlations between factors were evident in all age groups ranging from .58 for the age group 1 to .86 for the age group 2 with a median of .82.

Unique variance (theta-delta) in the model is the sum of two uncorrelated components, a systematic component specific to a subtest and one representing pure random measurement error. Unique variances ranged from a low of .160 for Similarities subtest for the age group 5 to a high of .796 for Digit Symbol subtest for the age group 1. This pattern was also recognized through the investigation of table 6 which reported the squared multiple correlations (SMCs) for each

Table 3 : Unstandardized Parameter Estimates of the Two Factor Model

Prm*	G1	G2	G3	G4	G5	G6	G7
L1	1.00(.00)**	1.00(.00)	1.00(.00)	1.00(.00)	1.00(.00)	1.00(.00)	1.00(.00)
L2	.91(.07)	.78(.04)	.87(.05)	.85(.05)	.82(.05)	.86(.05)	.92(.06)
L3	.32(.04)	.40(.03)	.41(.03)	.41(.03)	.40(.03)	.38(.03)	.35(.04)
L4	1.02(.08)	.83(.05)	1.01(.06)	.87(.05)	1.03(.06)	1.08(.06)	1.15(.08)
L5	1.02(.08)	.95(.06)	1.33(.09)	1.17(.08)	1.46(.09)	1.71(.11)	1.72(.12)
L6	1.00(.00)	1.00(.00)	1.00(.00)	1.00(.00)	1.00(.00)	1.00(.00)	1.00(.00)
L7	1.46(.28)	1.67(.19)	1.02(.10)	1.20(.11)	1.33(.12)	1.16(.10)	1.08(.09)
L8	2.84(.57)	2.41(.32)	1.63(.18)	1.93(.20)	1.99(.21)	1.36(.17)	1.42(.15)
L9	.95(.18)	.70(.09)	.61(.06)	.71(.07)	.71(.07)	.66(.06)	.54(.05)
L10	1.10(.23)	1.39(.17)	.96(.09)	1.07(.10)	1.08(.10)	.73(.07)	.71(.07)
P12	5.24(1.21)	15.30(2.38)	17.80(2.52)	17.81(2.47)	17.31(2.47)	15.96(2.21)	16.20(2.27)
D1	8.84(1.20)	9.11(1.25)	12.15(1.44)	15.19(1.81)	11.81(1.43)	11.30(1.36)	11.44(1.36)
D2	7.41(1.00)	5.40(.75)	5.63(.76)	5.61(.82)	7.07(.88)	5.76(.77)	4.89(.71)
D3	4.14(.44)	5.69(.61)	4.43(.48)	4.85(.53)	4.67(.50)	4.44(.47)	7.05(.73)
D4	11.42(1.45)	9.99(1.21)	6.28(.91)	8.11(1.06)	6.89(1.01)	8.46(1.15)	9.31(1.25)
D5	11.62(1.47)	15.71(1.83)	21.51(2.56)	23.02(2.70)	23.23(2.86)	27.46(3.46)	27.91(3.45)

Table 3 : Unstandardized Parameter Estimates of the Two Factor Model, Continued

Prm*	G1	G2	G3	G4	G5	G6	G7
D6	15.84(1.72)	13.89(1.48)	12.67(1.49)	8.82(1.05)	12.82(1.44)	8.35(1.05)	10.42(1.28)
D7	12.65(1.60)	8.72(1.21)	10.42(1.29)	11.97(1.44)	11.75(1.49)	10.24(1.31)	7.54(1.07)
D8	64.83(7.65)	52.33(5.80)	52.74(5.83)	47.39(5.34)	47.57(5.39)	47.75(5.11)	42.75(4.71)
D9	3.29(.49)	3.96(.44)	3.48(.44)	4.12(.50)	3.57(.45)	3.26(.42)	3.58(.43)
D10	10.82(1.26)	7.67(.99)	9.90(1.21)	8.34(1.04)	7.84(.99)	7.58(.86)	8.32(.94)
X^2	89.47	81.48	94.29	94.48	79.34	63.97	98.13
df	34	34	34	34	34	34	34
GFI	.914	.917	.911	.912	.923	.993	.908
AGFI	.861	.866	.855	.858	.875	.902	.852
RMR	1.476	1.589	2.372	1.884	2.039	1.273	2.090

* Prm : Parameters

L : Factor Loading (Lambda)
P : Factor Covariance (Phi)
D : Measurement Error (Theta Delta)

1 : Information	6 : Digit Symbol
2 : Comprehension	7 : Picture Completion
3 : Arithmetic	8 : Spatial
4: Similarities	9 : Picture Arrangement
5 : Vocabulary	10 : Object Assembly

** Standard errors in parenthesis.

Table 4 : Standardized Parameter Estimates of the Two Factor Model

Prm*	G1	G2	G3	G4	G5	G6	G7
L1	.833	.910	.849	.846	.862	.851	.825
L2	.831	.912	.898	.911	.875	.890	.898
L3	.577	.743	.740	.752	.736	.697	.542
L4	.803	.868	.914	.884	.916	.896	.880
L5	.802	.845	.849	.834	.870	.872	.849
L6	.452	.586	.739	.760	.709	.782	.782
L7	.638	.837	.776	.769	.812	.795	.846
L8	.580	.669	.658	.697	.720	.581	.661
L9	.724	.687	.788	.771	.804	.799	.756
L10	.560	.805	.767	.790	.811	.695	.704
P12	.582	.858	.816	.831	.824	.807	.811
D1	.307	.172	.280	.285	.258	.276	.320
D2	.310	.168	.193	.171	.234	.207	.193
D3	.667	.449	.452	.435	.459	.514	.706
D4	.355	.247	.165	.219	.160	.198	.225
D5	.356	.285	.279	.305	.243	.240	.279
D6	.796	.656	.454	.422	.497	.388	.388
D7	.593	.300	.397	.409	.340	.367	.284
D8	.664	.553	.568	.514	.481	.663	.563
D9	.475	.528	.379	.405	.353	.362	.428
D10	.687	.352	.412	.376	.343	.517	.505

* Prm : Parameters
L : Factor Loading (Lambda) P : Factor Covariance (Phi) D : Measurement Error (Theta Delta)
1 : Information 2 : Comprehension 3 : Arithmetic 4 : Similarities 5 : Vocabulary
6 : Digit Symbol 7 : Picture Completion 8 : Spatial 9 : Picture Arrangement
10 : Object Assembly

Table 5 : Mean Unstandardized and Standardized Parameter Estimates for Total Groups

	<Unstandardized estimates>				<Standardized estimates>			
		M.V	M.P	Total		M.V	M.P	Total
L1	1.000				.854			
L2	.859				.888			
L3	.381	.915			.684	.830		
L4	.999				.880			
L5	1.337			1.049	.845			.776
L6	1.000				.687			
L7	1.274				.784			
L8	1.940		1.183		.652		.723	
L9	.697				.761			
L10	1.006				.734			
P12	15.089				.790			
D1	11.406				.272			
D2	5.967				.210			
D3	5.039	10.059			.527	.304		
D4	8.637				.227			
D5	21.494			13.785	.285			.386
D6	11.830				.515			
D7	10.470				.384			
D8	50.766		17.061		.570		.469	
D9	3.609				.420			
D10	8.639				.457			

M.V : Mean Verbal
M.P : Mean Performance

subtest separately and total coefficients of determination (TCD) for all subtests jointly. Since SMC plus unique variance equals unity, a subtest with a large unique variance has a small SMC and a subtest with a small unique variance has a large SMC. Squared Multiple Correlation (SMC) for a subtest is the relative amount of variance in the subtest which is accounted for by all factors jointly (Joreskog and Sorbom, 1989, p. 89). Since a SMC is a lower bound of the reliability of the subtest, each reliability is at least as large as the SMC indicates (Joreskog and Sorbom, 1989, p. 89). SMC and TCD show how well the subtests serve, separately and jointly, as measurement instruments for the factors. The measures range between zero and one, large values being associated with good models (Joreskog and Sorbom, 1989, p. 42).

Among age groups, the smallest squared multiple correlation was .204 for Digit Symbol subtest for the age group 1 and the largest was .840 for Similarities subtest for the age group 5. The Verbal factor subtests had, on the average, larger square multiple correlations than the Performance subtests. Among all subtests, Arithmetic, Digit Symbol, and Spatial subtests had relatively lower square multiple correlations. Of the five measures of the Verbal factor, Comprehension subtest was the most valid, and of the five measures of the Performance factor, Picture Completion subtest was the most valid. The most valid subtest among all tests was Comprehension subtest. Total coefficients of determination for subtests were very high, ranging from .966

Table 6 : Squared Multiple Correlations and Total Coefficient of Determination for Subtests

	SMC*										TCD**
Group Age	INF	COM	ARI	SIM	VOC	DS	PC	SP	PA	OA	
G1:16-17	.693	.690	.333	.645	.644	.204	.407	.336	.525	.313	.966
G2:18-19	.828	.832	.551	.753	.715	.344	.700	.447	.472	.648	.981
G3:20-24	.720	.807	.548	.835	.721	.546	.603	.432	.621	.588	.983
G4:25-34	.715	.829	.565	.781	.695	.578	.591	.486	.595	.624	.981
G5:35-44	.742	.766	.541	.840	.757	.503	.660	.519	.647	.657	.984
G6:45-54	.724	.793	.486	.802	.760	.612	.633	.337	.638	.483	.982
G7:55-64	.680	.807	.294	.775	.721	.612	.716	.437	.572	.495	.981
Mean Subtest	.729	.789	.474	.776	.716	.486	.616	.428	.581	.544	
Mean Verbal			.700								
Mean Performance								.531			
Mean Total					.614						.980

* SMC : Squared Multiple Correlations for Subtests
INF : Information COM : Comprehension ARI : Arithmetic SIM : Similarities
VOC : Vocabulary DS : Digit Symbol PC : Picture Completion SP : Spatial
PA : Picture Arrangement OA : Object Assembly

** TCD : Total Coefficient of Determination for Subtests to .984 with a mean of .980. These values indicate that all subtests jointly served as a good measurement instrument for two factors. Therefore, the measurement model for the proposed two factor model was good.

<u>Ho1.3</u> The difference in explanatory power between the null model and the two factor model was examined. For this hypothesis testing a X^2 difference test was conducted. The X^2 value obtained from the test of the two factor model (601.16 with 238 df) was compared with that of the null model (9471.87 with 315 df). The X^2 difference between the two models yielded the X^2 value of 8870.71 with 77 degrees of freedom. This X^2 value was highly statistically significant. Therefore, the hypothesis of no difference in explanatory power between the two factor model and the null model was rejected, indicating that the two factor model better explained the data.

Considering the effect of a large sample size on the X^2 difference test, Wheaton et al.'s (1977) ratio of X^2 to degrees of freedom, and Bentler and Bonett's (1980) NFI and NNFI were calculated. They were 115.20, .94, and .95, respectively. Especially, NFI indicated that the two factor model could explain an additional 94% of the variation in the MAB beyond the null model. These results further substantiated the finding that there was a statistically significant difference in explanatory power between the two factor model and the null model. Therefore, the two factor model was thought to better fit the data than the null model.

The results of the analyses for research question one were reported in table 2.

In sum, the findings from the analyses of research question one demonstrated that the null model was not able to explain the subtest variation in the actual data. The two factor model better explained the actual data.

Research Question 2

The purpose of research question two was to explore the developmental invariance of the factor structure of the MAB. The reason for assessing the factorial invariance in the structure of a measure of intelligence is that cross-age group differences may exist in the structure of the measure. Such differences would have serious implications in the comparative analyses.

Invariance is divided into two dimensions: model form and similarities in parameter values. Two models have the same form if the model for each group has the same parameter matrices with the same dimensions and the same location of fixed, free, and constrained parameters.

One must decide which elements or matrices of parameters should be tested for equality across groups and in what order these tests should be made, which is somewhat arbitrary. Therefore, the order in which parameter equalities are tested can be altered in accordance with substantive interest (Bollen, 1989, pp. 356-365). For this research, the order suggested by Joreskog was used.

The two factor model which was confirmed in research question one to provide the explanation of the actual data was tested for factorial invariance across the age groups. The covariance structure of the MAB subtests was viewed as a function of three parameter matrices of factor loadings, variances and covariances of measurement errors, and factor covariances. Therefore, cross-age group factorial invariance was analyzed by testing a series of nested hypotheses involving one or more equivalence constraints on those parameter matrices in all age groups of interest.

The base upon which the order of testing sub-hypotheses for factorial invariance across age groups was made is as following. In most applications to date, researchers assume that the form of models is the same, and they concentrate on the similarity of parameter values within a given form (Bollen, 1989, p. 356). Therefore, the sub-hypothesis of equality in the number of factors (Model 2) across the age groups was first tested. Since the equality of factor loadings (Model 3) is generally of a higher priority than the equality of measurement error variances (Model 4) or the equality of the covariance matrices (Model 5) in the different groups (Bollen, 1989, p. 360), the sub-hypothesis of the equality of factor loadings across age groups preceded the sub-hypotheses of invariant measurement covariances and invariant factor covariances. Next, the sub-hypothesis of the equality of variances and covariances of measurement error was tested. The highest step in the hierarchy was the sub-hypothesis of full constraints where all three parameter matrices were simultaneously tested for equality.

The simultaneous confirmatory factor analyses were undertaken by using the LISREL VII program (Joreskog and Sorbom, 1989). First, the test of the sub-hypothesis of equality of covariance structures of the subtests across groups was conducted as a general preliminary test. This analysis yielded a X^2 value of 800.55 with 330 degrees of freedom. It was statistically significant leading to the rejection of the hypothesis. However, considering the effect of a large sample size on the X^2 test, Wheaton et al's (1977) ratio of X^2 to degrees of freedom was also calculated. This ratio was 2.43 suggesting each group had equal observed covariance structures. Actually, the problem of factorial invariance across age groups was settled with this result. But, for the purpose of the cross validation of invariant factor structure, the subsequent analyses were continued.

The second sub-hypothesis tested was the hypothesis of equality in the number of common factors. This test was related to the test of no between-group constraints model (the basic model, Model 2). This analysis produced a X^2 value of 601.16 with 238 degrees of freedom, rejecting the hypothesis. However, obtained Wheaton et al's (1977) ratio of X^2 to degrees of freedom and Bentler and Bonett's (1980) NFI were 2.53 and .94, respectively, suggesting each group had a common number of factors. These results did not coincide with the X^2

result. However, because of placing emphasis on practical significance criteria when the fit indices did not converge, the assessment of overall goodness-of-fit was based on X^2 ratio to df and NFI. Therefore, it appeared that each group had a common number of factors. Since this sub-hypothesis was found tenable, the procedures to test various between group constraints were followed.

The third sub-hypothesis was the hypothesis of an invariant factor pattern (factor loadings) across the age groups. To examine the sub-hypothesis of invariant factor pattern, it was assumed that the number of factors in each group was known *a priori* and that these factors were held in common across all groups (obtained in the test of hypothesis of Model 2). In addition, the estimation of a model constraining common factor patterns across all groups was required (Model 3). This analysis produced a X^2 value of 787.75 with 286 degrees of freedom. Then this model was compared with the model with no between-group constraints on the factor patterns across the age groups (Model 2). With these two models, a X^2 difference test was conducted to examine whether the sub-hypothesis of equality of factor patterns across the age groups was tenable. This X^2 difference test yielded a X^2 value of 186.59 (787.75 - 601.16) with 48 (286 - 238) degrees of freedom leading to the rejection of the hypothesis.

However, other things being equal, a large X^2 value is more likely for a large sample. This makes the rejection of the null hypothesis and supports for the proposed model less likely. To determine if the lack of good fit obtained for the proposed model (maintained model, or less restrictive model) is due only to the large sample, it is important to compare a more restrictive model (e.g., the model of equality in the number of common factors and factor loadings) to a less restrictive model (e.g., the model of equality in the number of common factors) using the null model as a baseline model (Bentler and Bonett, 1980, pp. 599-600). If the less restrictive model has essentially the same fit as the more restrictive model, then no difference in explanatory power for subtest variation between the two models exists, indicating the failure to reject the hypothesis related to the test of invariance of parameters in one or more parameter matrices (e.g., invariance of factor loadings).

Therefore, Bentler and Bonett's (1980) NFI was calculated. This incremental fit index measures the improvement in the fit for the equal common factor model relative to the fit of the invariant factor patterns model as a proportion of the fit of the null model as a baseline model. This index yielded .02, a non-significant increase in the equal common factor model over the invariant factor patterns model in explaining the subtest variation of the MAB. The ratio of X^2 to degrees of freedom obtained from the X^2 difference test was also provided. It was 3.89, also suggesting no difference between the two models. Therefore, the invariance of the factor patterns was suggested.

The sub-hypothesis of equality of covariance structure of measurement errors was tested based on the failure to reject the sub-hypothesis of invariant factor pattern. The test of model 4 of invariant measurement errors plus previous constraints in model 2 and model 3 yielded a X^2 value of 950.07 with 346 degrees of freedom. Then the sub-hypothesis of invariance in measurement errors was examined using a X^2 difference test between the X^2 value obtained from the test of model 4 and the X^2 value from the test of model 3. It was 162.32 (950.07 - 787.75) with 60 (346 - 286) degrees of freedom. Since it was significant, the invariance of the covariance structure of measurement errors was rejected. However, under the consideration of a large sample size effect on a X^2 difference test, and more emphasis on the practical significance criterion, Wheaton et al's (1977) ratio of X^2 to degrees of freedom, and Bentler and Bonett's (1980) NFI were also provided. The X^2 ratio of 2.71 and NFI value of .02 substantiated the failure to reject the sub-hypothesis. Non-significant increase existed in the invariant factor loadings model over the invariant measurement errors model in explaining the data. Therefore, the invariance of covariance structure of measurement errors across the age groups was suggested.

Finally, based on the finding of the tenable hypothesis of invariant measurement errors, a X^2 test was conducted for the hypothesis of invariance in factor covariance (Model 5). The test of model 5 of invariant factor covariance plus previous constraints in model 2, model 3, and model 4 yielded a X^2 value of 1000.05 with 364 degrees of freedom. The sub-hypothesis of invariance in factor covariance was then examined using a X^2 difference test between the X^2 value obtained from the test of model 5 and the X^2 value obtained from the test of model 4. The X^2 difference test yielded a X^2 of 49.98 (1000.05 - 950.07) with 18 (364 - 346) degrees of freedom. Since this X^2 value was significant, the sub-hypothesis of invariance in factor covariance across the age groups was rejected. However, considering the effect of a large sample size on a X^2 difference test and the emphasis on the practical significance, Wheaton et al's (1977) X^2 ratio to degrees of freedom, and Bentler and Bonett's (1980) NFI were also obtained. The X^2 ratio to df was 2.78 and NFI was .01. Both indices indicated that there was non-significant difference between the invariant measurement errors model and the invariant factor covariance model in explaining the data. Therefore, the hypothesis of equality of factor covariance was substantiated.

As Joreskog (1971, p. 420) pointed out, this hypothesis is a stronger hypothesis than the hypothesis of equality of covariance structures of the subtests across the age groups (model 1) since it specifies a particular form of invariance. With the result from the test of sub-hypothesis of invariant factor covariance structure (model 5), it was confirmed again that each group had the same factor structure. The results of factorial invariance involving the two factor model are presented in tables 7, 8, 10, 11, and 12.

Furthermore, it was also found that the two factor model with no between group constraints explained only a very small increase over the two factor model with the equal constraints on the unique variances in addition to the constraints on the factor loadings across the age groups. Likewise, constraining the factor covariance across the age groups in addition to the constraints on the factor loadings and unique variance yielded only a small degree of decrease in fit when compared to the no between-groups constraints model. Therefore, there were no practically significant differences in explaining the subtest variation of the MAB between the model with no between-group constraints and the models with various between group constraints.

There were also no differences in explaining the subtest variation among the models with various between-group constraints. Tables 7 and 8 reported these results.

The correlations between the Verbal factor and the Performance factor ranged from .58 to .86. Based on the fact that these high interfactor correlations might suggest a strong 'g' factor underlying all subtests, the single factor model ('g'-only model), which assumed that all of the MAB subtests measured only 'g' factor, was tested to examine if it provided good fit to the data for all age groups. The obtained X^2 value was 1268.66 with degrees of freedom of 245 leading to the rejection of the null hypothesis that the single factor model explained the subtest variation of the

Table 7 : Goodness of Fit Associated with Models with Various Equivalence Constraints

				X^2 difference test				
Model*	X^2 [a]	df	X^2/df	X^2 [b]	df	X^2/df	NFI**	
Model 1	800.55	330	2.43					
Model 2	601.16	238	2.53				.94	
				186.59	48	3.89		.02
Model 3	787.75	286	2.75				.92	
				162.32	60	2.71		.02
Model 4	950.07	346	2.75				.90	
				49.98	18	2.78		.01
Model 5	1000.05	364	2.75				.89	

a: $p < .05$ b: $p < .05$ (N=1400)

* Model 1: Model for a preliminary test (equal observed covariance matrices)
Model 2: Model with no equivalence constraints
Model 3: Model with equivalence constraints on factor loadings
Model 4: Model with equivalence constraints on factor loadings and measurement errors
Model 5: Model with equivalence constraints on factor loadings, measurement errors, and factor covariance

** NFI : Normed Fit Index of Bentler and Bonett (1980) (based on the null model for total groups)

Table 8 : Differences in Goodness-of-Fit between No Constraints Model and the Models with Various Constraints

Model*	X^2-difference [a]	df-difference	X^2_d / df_d **
Model 3 - Model 2	186.59	48	3.89
Model 4 - Model 2	348.91	108	3.23
Model 5 - Model 2	398.89	126	3.17
Model 5 - Model 3	212.30	78	2.72

a: $p < .05$ (N=1400)

* Model 2: Model with no equivalence constraints
Model 3: Model with equivalence constraints on factor loadings
Model 4: Model with equivalence constraints on factor loadings and measurement errors
Model 5: Model with equivalence constraints on factor loadings, measurement errors, and factor covariance

** X^2_d : X^2 difference df_d : df difference

MAB. However, the obtained NFI and NNFI were .87 and .86, indicating good fit to the data. Therefore, the single factor model was able to explain the actual data. Table 13 reports the result.

On the basis of the result of the analysis of 'g' factor model, it seems that one's total subtest scores on the MAB can provide an estimate of general mental ability. That is, the interpretation of the Full Scale IQ as a global measure of intelligence may be possible. These correlated factors provided enough evidence supporting an underlying general factor in the MAB. Therefore, it should be noted that the substantial interfactor correlations mean that those factors (the Verbal factor and the Performance factor) themselves do not represent dimensions which can be considered completely independent of each other at all.

The single factor model was compared with the two factor model to see which model provided a better fit to the data. NFI and NNFI were obtained using the null model as a base line model. Both models provided similar overall fit, and little improvement was found when the two factor model was compared to the single factor model.

An analysis of the single factor model ('g'-only model) reveals that the MAB subtests cluster into three groups of good, fair, and poor measures of 'g'. All verbal subtests except Arithmetic subtest belong to good measures of 'g'. Arithmetic, Digit Symbol, Picture Completion, Picture Arrangement, and Object Assembly subtests are fair measures of 'g'. Spatial subtest is the only one which belongs to poor measures of 'g'. Tables 14 and 15 report factor loadings of the single factor model and the proportion of variance attributed to 'g'. Each factor loading indicates the extent to which each subtest serves as a measure of 'g'. Therefore, four verbal subtests may serve as a good measure of general intelligence. Comprehension subtest best serves as a measure of 'g' and the worst one is Spatial subtest.

The consistencies across the age groups are noticeable for some subtests. All verbal subtests (except Arithmetic subtests for group 1 and group 7) serve as good measures of 'g' across the age groups. Among the performance subtests, only Picture Completion subtest serves as a good measure of 'g' from age to age (except group 1 and group 6). However, in each age group, subtests are mixtures of good, fair, and poor measures of 'g'.

By squaring the mean factor loading of each subtest, the proportion of each subtest's variance attributed to general intelligence was obtained. In this study, for the good measures of 'g', this amounts to around 67%-77%; for the fair measures, the range is about 37%-49%; and for the poor measure, it is about 29%.

There are some implications of the MAB coming from the results of this study. First, the problem of separate use of the Verbal and Performance IQ scores of the MAB as a diagnostic tool in neuropsychological assessment is further understood as the result of the considerable overlap between the Verbal and Performance subscales. The significant loadings of verbal subtests on the Verbal factor and those of performance subtests on the Performance factor are probably sufficient to justify using the Verbal IQ and the Performance IQ as crude indicators of verbal ability and performance ability, respectively. However, because of the large interfactor correlations in the age groups, when separate Verbal and Performance IQs are maintained, other information such as subjects' reading levels should be taken into consideration for valid interpretation of test performance. Since two factors have a common part in explaining the subtest variation, some portion of the variance explained by one factor is also explained by the other factor. Therefore, each IQ would be interpreted in terms of the ability assumed to be interpreted mainly by one IQ, along with the ability assumed to be interpreted by the other IQ. For example, the ability in the Performance IQ would be understood through not only one's performance ability but also verbal ability in some degree.

Same cautions are required in clinical applications when interpreting mental ability in terms of a single total score, a Full Scale IQ, which is obtained from the combination of the Verbal and Performance IQs. Since the two factors are different constructs which define intelligence differently according to one's uniqueness in the structure of mental ability, the use of a combined unidimensional numerical index obtained by summing different multidimensional qualities may require some other related information for the valid clinical and job-related judgment or for comparative purpose.

The consistencies of the high loadings by the MAB subtests on the Verbal and the Performance factors across groups were remarkable. Especially, the consistency of factor loadings of the verbal subtests on the Verbal factor across age groups was notable. The loadings of each verbal subtest on the Verbal factor were stable and oscillated within a narrow range across age groups as indicated from the result of invariant factor loadings. The consistency of factor loadings of the performance subtests on the Performance factor from age to age was also noteworthy. Although the loadings of performance subtests on the Performance factor fluctuated rather larger than that of verbal subtests did on the Verbal factor, the mean loading of each performance subtest across age groups approached closely the mean of total loadings of the performance subtests on the Performance factor, which might contribute to invariant factor loadings.

However, a substantial amount of the variance of some subtests still remained unexplained, implying the possible emergence of other factors. Therefore, from the fact that some of the relationships in the model were not well determined as revealed by the low squared multiple correlations (SMCs) of some subtests, it is implied that although the overall fit of the model is good, the model still needs to be improved to increase the explanation of subtest variation (see tables 3, 4, 5, and 6).

Table 9 reported the variance partition for total groups using the two factor model. The subtest specificity indicates the proportion of reliable variance that is unique to that subtest in the battery and provides information about the extent to which one might be justified in making inferences about scores on individual subtests. The subtest specificity is obtained by subtracting the SMC (common variance) for each subtest from the reliability estimate. If the test specificity is relatively low, then a test cannot be said to be measuring the specific trait which the test is assumed to measure. Even though a test has considerable proportion of specific variance, if the test also has the error variance larger than the specific variance, meaningful interpretation of scores on the test is precluded. According to Kaufman (1975, p. 144), for a test to warrant specific interpretation, the specific variance should be greater than the error variance and should account for at least one quarter of the total variance. Then, one can draw inferences of specific ability or disability, but only when there is a substantial departure of the subtest score from the mean of the other subtests (Cohen, 1959, p. 293).

For this study, the specific variances of all subtests except Similarities exceeded their error variances. Test-retest reliability in the MAB manual (Jackson, 1984, p. 41) was used for the reliability estimates. Overall, while average error variance of the MAB was .09, the average proportion of variance due to specificity was .30. The

Table 9 : Variance Partition of Subtests for Total Groups

Subtests*	Factors		Variance Partition				
	V	P	C**	R	U	= (S*** +	E)
Verbal Subtests							
Information	.85		.73	.97	.27	.24	.03
Comprehension	.89		.79	.95	.21	.16	.05
Arithmetic	.68		.46	.88	.53	.41	.12
Similarities	.88		.77	.83	.23	.06	.17
Vocabulary	.85		.72	.90	.29	.19	.10
Mean-Verbal	.83		.69	.91	.30	.21	.09
Performance Subtests							
Digit Symbol		.69	.48	.90	.52	.42	.10
Picture Completion		.78	.61	.94	.38	.32	.06
Spatial		.65	.42	.93	.57	.50	.07
Picture Arrangement		.76	.58	.87	.42	.29	.13
Object Assembly		.73	.53	.93	.46	.39	.07
Mean Performance		.72	.52	.91	.47	.38	.09
Mean Total	.78		.61	.91	.39	.30	.09

* V : Verbal factor, P : Performance factor, R : Reliability estimates, C : Common variance (SMC), U : Unique variance, S : Specificity of a subtest, and E : Measurement error.

** 1 = C + U, R = C + S, and U = S + E. Therefore, 1 = C + S + E. The discrepancies between these formulas and the numbers from actual calculations are due to rounding errors.

*** The specificity is obtained by subtracting the communality or common variance from the reliability estimate for each subtest. For this research, the test-retest reliability reported in the MAB manual (Jackson, 1984, p. 41) is used for reliability estimates.

specificities for the Verbal subtests were, on the average, lower than those of the Performance subtests. If Kaufman's (1975) guidelines are applied in this study, then the Arithmetic subtest of Verbal scale and all Performance subtests warrant specific interpretations. Especially, Arithmetic (41%), Digit Symbol (42%), and Spatial (50%) subtests had specificities larger than .40. Specificities of this magnitude are probably large enough to justify making inferences about scores on these subtests scores with a reasonable degree of confidence when they are noticeably discrepant from the rest of the MAB test profile. Around one-third of the total variance in the following subtests consisted of specificities: Picture Arrangement (29%), Picture Completion (32%) and Object Assembly (39%). These specificities probably permit valid interpretation of scores on these subtests if they are noticeably discrepant from the rest of the test profile. However, the specificities for the remaining subtests were moderately or considerably smaller. Therefore, considerable restraints are required in the interpretation of scores of these subtests.

Since an individual's profile of scaled scores on the subtests is used to assess one's specific strengths and weaknesses, test specificity is very important in terms of clinical interpretation (Kaufman, 1975, p. 144). The result from the analyses of subtest specificity indicates that some subtests of the MAB have sufficient specificities to warrant clinical interpretation of their supposedly unique contributions to the battery. Therefore, it might be possible for subtests with high subtest specificity to be used as tools for assessing strengths and weaknesses on the abilities which they are assumed to measure.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This chapter presents a summary of the research described in the previous chapters, and conclusions and recommendations that can be drawn from the results of the statistical analyses.

The present study was undertaken to examine the psychometric properties of the Multidimensional Aptitude Battery (MAB), which is a group test of intelligence and mental abilities patterned quite closely after the Wechsler Adult Intelligence Scale-Revised (WAIS-R). Especially, this study was designed to factor analyze the data from the 1990 norming study through confirmatory factor analysis in order to investigate the degree to which the proposed two factor model explained the subtest variation of the MAB, and to examine the factorial invariance across age groups.

Summary

The first research question asked was: Does the two factor (the Verbal and Performance factors) model explain the variation in the MAB subtest scores for all groups simultaneously? For this research question, the following three null hypotheses were tested.

Ho1.1 There is no significant difference between the null model which represents no relationships among the subtests and the model that represents the actual data for all groups.

Ho1.2 There is no significant difference between the two factor model and the model that represents the actual data.

Ho1.3 There is no significant difference in explanatory power between the two factor model and the null model.

The results of analyses of this research question, using multiple group confirmatory factor analyses, provided that there was a significant difference between the null model and the model explaining the actual data for all age groups. The significance was demonstrated in a X^2 value of 9471.87 with 315 degrees of freedom. Considering the effects of a large sample size on a X^2 test, Wheaton et al's (1977) ratio of X^2 to degrees of freedom was also obtained. It was 30.07, substantiating the difference between the null model and the model representing the data. Therefore, the rejection of this hypothesis indicated that the model with specified relationships among the subtests and factors would better explain the actual data than the model of no relationships among the subtests.

Thus, the two factor model based on previous studies was proposed and tested. A significant difference was obtained between the two factor model and the model representing the actual data for all age groups. The significance was demonstrated in a X^2 value of 601.16 with 238 degrees of freedom indicating that the two factor model could not explain the variations in the MAB subtests scores for all age groups. However, to correct a large sample size effect on a X^2 value, Wheaton et al's ratio of X^2 to degrees of freedom was calculated. It was 2.53 indicating that there was no significant difference between the two factor model and the model representing the actual data for all age groups.

Next, the two factor model was compared to the null model to examine if there existed a difference in explanatory power between them. Through a X^2 difference test between the two factor model and the null model (8870.71 with 77 degrees of freedom), it was evident that the two factor model provided a better fit to the actual data than the null model. Furthermore, Wheaton et al's X^2 ratio of 115.20, and Bentler and Bonett's NFI of .94 and NNFI of .95 substantiated a statistically significant difference in explanatory power between the two factor model and the null model. All these analyses led to the conclusion that the two factor model had definitely a different factor structure than that of the null model and was more effective in explaining the actual data for all age groups when analyzed simultaneously.

The second research question asked was: Are there any significant differences across the age groups with respect to factorial invariance? This research question assessed the adequacy of fit of the two factor model with various constraints across the age groups. The corresponding null hypothesis was:

Ho2 There is no significant difference across the age groups with respect to factorial invariance.

The multiple group analyses made it possible to compare the factor structures across the age groups and to test increasingly restrictive hypotheses about how well the factor structure was replicated across the age groups. This analysis consisted of testing five sub-hypotheses;

(1) Sub-hypothesis 1: there is equality of covariance structures of the observed variables across the age groups;

(2) Sub-hypothesis 2: there is equality in the number of common factors across the age groups;

(3) Sub-hypothesis 3: there are invariant factor loadings as well as the same number of common factors across the age groups;

(4) Sub-hypothesis 4: there are invariant measurement errors as well as the same number of common factors and invariant factor loadings across the age groups; and

(5) Sub-hypothesis 5: there are invariant factor covariance as well as the same number of common factors, invariant factor loadings, and invariant measurement errors across the age groups.

Each sub-hypothesis was tested based on the tenability of a previous sub-hypothesis. Using the two factor model with no between-groups constraints as a basic model, the sub-hypotheses of 2, 3, 4, and 5 produced four hierarchical models. X^2 values obtained from X^2 difference tests among the models indicated that there were significant differences among them.

However, under the consideration of the practical significance, the examination of NFI in constraining the factor loading across the age groups revealed that the two factor model with no between-group constraints had only a very small increase in the explanation of the subtest variation of the MAB over the two factor model with factor loadings held equal across age groups. There was also no significant increase in the model with invariant factor loadings over the model with another constraint of invariant measurement errors. The same was true between the model with invariant measurement errors and the model with full constraints. The obtained results indicated that all groups had equal observed covariance structures. Therefore, it was found that there existed invariant factor structures across the age groups.

The finding that the factor structure of the MAB closely agreed with the actual organization of the subtest seemed to give strong empirical support to the interpretation of the Verbal and the Performance IQs as separately functioning entities in the MAB. Furthermore, the substantial correlations between the two factors provided support for the argument that the MAB was a measure of general intelligence. In other words, the significant factor correlations supported that general intelligence, 'g', might exercise influence over individual differences in abilities assessed by the MAB.

The subtest specificity was tested. The subtest specificity is the proportion of reliable variance which is unique to that subtest in the battery and provides information about the extent to which one might be justified in making inferences about the scores on individual subtest. The subtests of Arithmetic, Digit Symbol, Spatial, Picture Completion, Object Assembly, and Picture Arrangement had sufficient specificities to warrant specific interpretation.

In sum, it was substantiated that the proposed two factor model fitted the actual data well and would be a good measurement model of the MAB for all age groups under examination. Furthermore, it was also supported that there existed factorial invariance across the age groups.

Conclusions

All findings from this study may provide an empirical ground for valid interpretation of the results from future studies. The major findings of research analyses of the present study are as following:

(1) The null model failed to provide a good fit;

(2) The two factor model was able to explain the subtest variation of the MAB for all groups;

(3) The two factor model was superior to the null model in explaining the actual data for all age groups;

(4) Invariant factor structure existed across the age

groups;

(5) The result from the analysis of the single factor model supported an underlying general factor in the MAB; and

(6) Some subtests had sufficient specificities to warrant the assessment of strengths and weaknesses on the abilities which they were assumed to measure.

The dimensional structure and invariance of factor structure of the MAB were assessed in this study. It was substantiated that the hypothetical model with two factor structure (the Verbal and Performance factors) fitted the actual data of the MAB.

Based on the results of the two factor model and the single factor model, the present study provided a sound empirical basis for the conclusion that the MAB is a measure of multidimensional mental ability, along with general intelligence. That is, this study not only provided support for the presence of a general factor but also furnished support for maintaining two separate dimensions for the MAB: the Verbal ability and the Performance ability.

Compared to the single factor solution, the two factor solution of the MAB seemed to assess more limited cognitive domains. However, the more valid interpretation on individual differences would be obtained if clinicians take the two factor solution in interpreting their clients' mental abilities measured with the MAB.

From the fact that both the two factor model and the single factor model provide better fits to the data than the null model, it is implied that the interpretation of the MAB data using either of these models would be more valuable than the interpretation of individual subtests (the two factor model and the single factor model were able to explain additional 94% and 87% covariances, respectively, beyond what was explained by the null model).

Furthermore, the fact that factor structure of the MAB remained fairly constant across the adult life span implies that it can be used as a tool for assessing and comparing one's intellectual faculties across the developmental span from adolescence through adulthood to old age. Therefore, the MAB can be used as a tool for longitudinal research of the developmental phases of individuals' mental ability.

The findings based on the statistical analyses confirmed the previous findings of the two factor structure of the MAB and its similarity in factor structure to that of the WAIS-R. Actually, the MAB was designed under its author's intention to evaluate the degree to which it is possible to incorporate some of the widely acknowledged positive features of individually-administered multi-scale tests, such as the WAIS-R, into a structured format which allows group administration, automated administration, and convenient hand or machine scoring. Therefore, the use of the MAB as an alternative for the WAIS-R when used for group testing might be possible.

This study, as the first confirmatory factor analysis of the MAB, using structural equation models, confirmed previous research findings. However, It seems that still many studies are required to properly appraise its psychometric properties and to determine its adequacy as a psychoeducational assessment tool.

Recommendations for Future Studies

The major limitation of the present research comes from the sample selection procedure in that a true random sample was not used. Many factor analytic studies for the MAB were conducted using different samples and various statistical methods. As Reynolds and Kaufman (1985, p. 612) pointed out, today, the major intelligence scales are developed from large nationally stratified random samples. Samples are typically stratified on the basis of age, sex, race, socioeconomic status (usually determined by parental occupation), geographic region of residence (North, South, Central, West), and the subjects' residence (in an urban or a rural setting). Such careful sampling procedures are required to ensure the stability and the generalizability of scores on the battery. Less careful standardization and norming should not be considered acceptable for tests that will possibly impact on individuals' lives. Accordingly, some restricted characteristics of the data of this study, especially with respect to sampling procedure, limit, in some degree, the generalization of the results to other situations. In this context, future studies using the same statistical methods but much stricter sampling procedures to obtain more representative samples are suggested for more valid interpretation of the results of the MAB and for stronger grounds for the generalization of the findings to whole populations.

Although the two factor model provided a good fit to the data, there may be other factor structures that provide a fit as good as or better than the model reported here. Therefore, for future studies, other models should be considered.

Furthermore, since this study was based on only seven age groups among a total of nine age groups which the MAB covers, the results of this study might take different phases if all nine age groups were used. Therefore, future studies using all nine age groups are recommended to detect possible differences in the results of this study and future ones.

Finally, it is recommended that the MAB be examined to see if it has stability in factor structures when used with samples from different populations such as learning disabled, hearing-impaired, anxiety disorder, and gifted ones as well as ethnic minorities with different cultural backgrounds. In addition, if there are translated versions of the MAB into many different languages other than English, then it is also recommended that cross-cultural comparisons of the MAB be examined to see if the same psychometric properties of the MAB occur across various cultures. All these kinds of efforts will eventually provide empirical bases for the correct use of the MAB and the valid interpretation of the results obtained from its clients within a culture and between cultures.

APPENDIX

Table 10 : Unstandardized and Standardized Parameter Estimates : Invariant Factor Loadings

Prm*	G1	G2	G3	G4	G5	G6	G7
L1	1.00(.00)** .82	1.00(.00) .88	1.00(.00) .86	1.00(.00) .83	1.00(.00) .87	1.00(.00) .87	1.00(.00) .85
L2	.85(.02) .78	.85(.02) .90	.85(.02) .90	.85(02) .90	.85(.02) .89	.85(.02) .90	.85 (.02) .90
L3	.39(.01) .63	.39(.01) .69	.39(.01) .73	.39(.01) .72	.39(.01) .74	.39(.01) .73	.39(.01) .63
L4	.98(.02) .76	.98(.02) .88	.98(.02) .91	.98(.02) .90	.98(.02) .91	.98(.02) .89	.98(.02) .86
L5	1.27(.03) .85	1.27(.03) .88	1.27(.03) .84	1.27(.03) .84	1.27(.03) .84	1.27(.03) .78	1.27(.03) .77
L6	1.00(.00) ..54	1.00(.00) .68	1.00(.00) .70	1.00(.00) .77	1.00(.00) .74	1.00(.00) .74	1.00(.00) .71
L7	1.23(.05) .66	1.23(.05) .81	1.23(.05) .81	1.23(.05) .80	1.23(.05) .81	1.23(.05) .78	1.23(.05) .83
L8	1.80(.08) .48	1.80(.08) .64	1.80(.08) .66	1.80(.08) .68	1.80(.08) .71	1.80(.08) .65	1.80(.08) .69
L9	.66(.03) .66	.66(.03) .75	.66(.03) .78	.66(.03) .76	.66(.03) .81	.66(.03) .76	.66(.03) .77
L10	.97(.04) .60	.97(.04) .75	.97(.04) .73	.97(.04) .76	.97(.04) .80	.97(.04) .76	.97(.04) .76
P12	6.19(1.05) .58	17.11(2.06) .85	16.88(2.08) .82	17.80(2.16) .83	19.40(2.36) .83	15.52(1.92) .81	15.20(1.88) .82
D1	8.72(1.13) .33	10.40(1.29) .23	11.99(1.43) .27	15.51(1.80) .31	1.58(1.42) .24	11.02(1.38) .25	11.20(1.39) .27
D2	8.48(1.02) .40	5.57(.75) .18	5.60(.74) .19	5.88(.80) .19	6.89(.89) .21	5.63(.78) .19	4.79(.71) .18
D3	4.13(.45) .61	5.88(.62) .53	4.47(.48) .47	4.97(.53) .49	4.64(.50) .46	4.34(.47) .46	6.99(.73) .61
D4	12.47(1.47) .42	9.59(1.20) .22	6.36(.90) .17	7.84(.1.07) .19	7.18(1.00) .17	8.98(1.17) .22	10.11(1.27) .26
D5	10.83(1.53) .27	16.71(2.07) .23	21.61(2.52) .29	22.91(2.71) .29	25.20(2.92) .30	34.45(3.84) .39	33.11(3.70) .41

Table 10 : Unstandardized and Standardized Parameter Estimates :
Invariant Factor Loadings, Continued

Prm*	G1	G2	G3	G4	G5	G6	G7
D6	15.59(1.74) .71	13.77(1.52) .54	13.17(1.48) .51	8.70(1.04) .40	12.62(1.44) .46	8.91(1.04) .45	11.49(1.29) .50
D7	12.38(1.58) .56	9.36(1.20) .35	9.95(1.28) .34	11.73(1.44) .37	11.79(1.47) .34	10.34(1.27) .39	7.72(1.03) .31
D8	68.47(7.40) .77	52.84(5.75) .59	52.98(5.82) .56	48.14(5.32) .54	47.96(5.34) .50	46.99(5.15) .58	41.75(4.64) .53
D9	3.60(.48) .59	3.90(.46) .44	3.52(.43) .39	4.17(.49) .42	3.56(.44) .35	3.50(.42) .43	3.54(.42) .41
D10	10.50(1.22) .64	8.29(.98) .44	10.26(1.19) .46	8.70(1.03) .42	8.06(.98) .37	7.34(.88) .42	8.08(.95) .43

X^2 of 787.75 with 286 df

* Prm : Parameters

L : Factor Loading (Lambda)
P : Factor Covariance (Phi)
D : Measurement Error (Theta Delta)

1 : Information	6 : Digit Symbol
2 : Comprehension	7 : Picture Completion
3 : Arithmetic	8 : Spatial
4 : Similarities	9 : Picture Arrangement
5 : Vocabulary	10 : Object Assembly

** First raw of each parameter represents unscaled maximum likelihood estimates (standard errors in parenthesis). Second raw represents standardized estimates.

Table 11 : Unstandardized and Standardized Parameter Estimates : Invariant Measurement Errors

Prm*	G1	G2	G3	G4	G5	G6	G7
L1	1.00(.00)** .79	1.00(.00) .87	1.00(.00) .86	1.00(.00) .87	1.00(.00) .87	1.00(.00) .87	1.00(.00) .85
L2	.85(.02) .84	.85(.02) .90	.85(.02) .89	.85(.02) .90	.85(.02) .90	.85(.02) .90	.85(.02) .89
L3	.39(.01) .60	.39(.01) .71	.39(.01) .70	.39(.01) .71	.39(.01) .72	.39(.01) .71	.39(.01) .69
L4	.99(.02) .82	.99(.02) .89	.99(.02) .88	.99(.02) .89	.99(.02) .89	.99(.02) .89	.99(.02) .88
L5	1.33(.03) .77	1.33(.03) .85	1.33(.03) .84	1.33(.03) .85	1.33(.03) .86	1.33(.03) .85	1.33(.03) .83
L6	1.00(.00) .60	1.00(.00) .69	1.00(.00) .72	1.00(.00) .72	1.00(.00) .74	1.00(.00) .68	1.00(.00) .69
L7	1.23(.05) .70	1.23(.05) .79	1.23(.05) .81	1.23(.05) .81	1.23(.05) .83	1.23(.05) .77	1.23(.05) .79
L8	1.83(.08) .55	1.83(.08) .65	1.83(.08) .68	1.83(.08) .68	1.83(.08) .70	1.83(.08) .64	1.83(.08) .65
L9	.67(.03) .67	.67(.03) .76	.67(.03) .78	.67(.03) .78	.67(.03) .80	.67(.03) .74	.67(.03) .76
L10	.99(.04) .65	.99(.04) .75	.99(.04) .77	.99(.04) .77	.99(.04) .79	.99(.04) .73	.99(.04) .75
P12	6.18((1.07) .55	16.73(2.02) .86	16.50((2.05) .81	17.44(2.14) .83	18.89(2.31) .83	15.26(1.91) .81	14.87(1.86) .81
D1	11.57(.54) .38	11.57(.54) .25	11.57(.54) .27	11.57(.54) .25	11.57(.54) .25	11.57(.54) .25	11.57(.54) .28
D2	5.97(.31) .30	5.97(.31) .19	5.97(.31) .20	5.97(.31) .19	5.97(.31) .19	5.97(.31) .19	5.97(.31) .22
D3	5.09(.21) .64	5.09(.21) .50	5.09(.21) .51	5.09(.21) .50	5.09(.21) .49	5.09(.21) .50	5.09(.21) .53
D4	8.75(.44) .32	8.75(.44) .21	8.75(.44) .22	8.75(.44) .21	8.75(.44) .20	8.75(.44) .21	8.75(.44) .23
D5	23.14(1.05) .41	23.14(1.05) .28	23.14(1.05) .29	23.14(1.05) .28	23.14(1.05) .27	23.14(1.05) .28	23.14(1.05) .30

Table 11 : Unstandardized and Standardized Parameter Estimates : Invariant Measurement Errors, Continued

Prm*	G1	G2	G3	G4	G5	G6	G7
D6	12.05(.52) .65	12.05(.52) .52	12.05(.52) .49	12.05(.52) .49	12.05(.52) .45	12.05(.52) .54	12.05(.52) .52
D7	10.52(.51) .51	10.52(.51) .38	10.52(.51) .35	10.52(.51) .35	10.52(.51) .32	10.52(.51) .40	10.52(.51) .38
D8	50.97(2.15) .70	50.97(2.15) .58	50.97(2.15) .54	50.97(2.15) .54	50.97(2.15) .51	50.97(2.15) .60	50.97(2.15) .58
D9	3.70(.17) .56	3.70(.17) .43	3.70(.17) .40	3.70(.17) .39	3.70(.17) .36	3.70(.17) .45	3.70(.17) .43
D10	8.64(.40) 57	8.64(.40) .44	8.64(.40) .41	8.64(.40) .41	8.64(.40) 38	8.64(.40) .46	8.64(.40) .44

X^2 of 950.07 with 346 df

* Prm : Parameters

L : Factor Loading (Lambda)
P : Factor Covariance (Phi)
D : Measurement Error (Theta Delta)

1 : Information	6 : Digit Symbol
2 : Comprehension	7 : Picture Completion
3 : Arithmetic	8 : Spatial
4 : Similarities	9 : Picture Arrangement
5 : Vocabulary	10 : Object Assembly

** First raw of each parameter represents unscaled maximum likelihood estimates (standard errors in parenthesis). Second raw represents standardized estimates.

Table 12 : Unstandardized and Standardized Parameter Estimates : Invariant Factor Covariance

Prm*	G1	G2	G3	G4	G5	G6	G7
L1	1.00(.00)** .86	1.00(.00) .86	1.00(.00) .86	1.00(.00) .86	1.00(.00) .86	1.00(.00) .86	1.00(.00) .86
L2	.85(.02) .89	.85(.02) .89	.85(.02) .89	.85(.02) .89	.85(.02) .89	.85(.02) .89	.85(.02) .89
L3	.39(.01) .69	.39(.01) .69	.39(.01) .69	.39(.01) .69	.39(.01) .69	.39(.01) .69	.39(.01) .69
L4	.99(.02) .88	.99(.02) .88	.99(.02) .88	.99(.02) .88	.99(.02) .88	.99(.02) .88	.99(.02) .88
L5	1.33(.03) .84	1.33(.03) .84	1.33(.03) .84	1.33(.03) .84	1.33(.03) .84	1.33(.03) .84	1.33(.03) .84
L6	1.00(.00) .70	1.00(.00) .70	1.00(.00) .70	1.00(.00) .70	1.00(.00) .70	1.00(.00) .70	1.00(.00) .70
L7	1.24(.05) .79	1.24(.05) .79	1.24(.05) .79	1.24(.05) .79	1.24(.05) .79	1.24(.05) .79	1.24(.05) .79
L8	1.84(.08) .65	1.84(.08) .65	1.84(.08) .65	1.84(.08) .65	1.84(.08) .65	1.84(.08) .65	1.84(.08) .65
L9	.67(.03) .76	.67(.03) .76	.67(.03) .76	.67(.03) .76	.67(.03) .76	.67(.03) .76	.67(.03) .76
L10	.99(.04) .75	.99(.04) .75	.99(.04) .75	.99(.04) .75	.99(.04) .75	.99(.04) .75	.99(.04) .75
P 1	15.11(.84) .80	15.11(.84) .80	15.11(.84) .80	15.11(.84) .80	15.11(.84) .80	15.11(.84) .80	15.11(.84) .80
D1	11.54(.54) .27	11.54(.54) .27	11.54(.54) .27	11.54(.54) .27	11.54(.54) .27	11.54(.54) .27	11.54(.54) .27
D2	5.98(.31) .21	5.98(.31) .21	5.98(.31) .21	5.98(.31) .21	5.98(.31) .21	5.98(.31) .21	5.98(.31) .21
D3	5.10(21) .52	5.10(.21) .52	5.10(.21) .52	5.10(21) .52	5.10(.21) .52	5.10(.21) .52	5.10(.21) .52
D4	8.73(.44) .22	8.73(.44) .22	8.73(.44) .22	8.73(.44) .22	8.73(.44) .22	8.73(.44) .22	8.73(.44) .22
D5	23.15(1.06) .30	23.15(1.06) .30	23.15(1.06) .30	23.15(1.06) .30	23.15(1.06) .30	23.15(1.06) .30	23.15(1.06) .30

Table 12 : Unstandardized and Standardized Parameter Estimates : Invariant Factor Covariance, Continued

Prm*	G1	G2	G3	G4	G5	G6	G7
D6	12.09(.53) .52	12.09(.53) .52	12.09(.53) .52	12.09(.53) .52	12.09(.53) .52	12.09(.53) .52	12.09(.53) .52
D7	10.52(.52) .38	10.52(.52) .38	10.52(.52) .38	10.52(.52) .38	10.52(.52) .38	10.52(.52) .38	10.52(.52) .38
D8	50.98(2.15) .57	50.98(2.15) .57	50.98(2.15) .57	50.98(2.15) .57	50.98(2.15) .57	50.98(2.15) .57	50.98(2.15) .57
D9	3.67(.17) .42	3.67(.17) .42	3.67(.17) .42	3.67(.17) .42	3.67(.17) .42	3.67(.17) .42	3.67(.17) .42
D10	8.69(.40) .44	8.69(.40) .44	8.69(.40) .44	8.69(.40) .44	8.69(.40) .44	8.69(.40) .44	8.69(.40) .44

X^2 of 1000.05 with 364 df

* Prm : Parameters

L : Factor Loading (Lambda)
P : Factor Covariance (Phi)
D : Measurement Error (Theta Delta)

1 : Information	6 : Digit Symbol
2 : Comprehension	7 : Picture Completion
3 : Arithmetic	8 : Spatial
4 : Similarities	9 : Picture Arrangement
5 : Vocabulary	10 : Object Assembly

** First raw of each parameter represents unscaled maximum likelihood estimates (standard errors in parenthesis). Second raw represents standardized estimates.

Table 13 : Measures of Fit of the Null model, the Single Factor Model, and the Two Factor Model

				X^2 difference test				
Model	X^2	df	X^2/df	X^2	df	X^2/df	NFI**	NNFI***
Null Model	9471.87*	315	30.07					
				8203.21*	70	117.19		
Single Factor Model	1268.66*	245	5.18				.87	.86
Two Factor Model	601.16*	238	2.53				.94	.95

(N=1400)

* $p < .05$

** Normed Fit Index

*** Non-Normed Fit Index

Table 14 : The MAB subtests as a measure of 'g' ;
Standardized Factor Loadings for the Single Factor Model

Prm*	G1	G2	G3	G4	G5	G6	G7	Total
L1	.835	.895	.839	.826	.859	.841	.812	.844
L2	.817	.909	.885	.904	.858	.875	.882	.876
L3	.585	.756	.753	.751	.738	.724	.580	.698
L4	.790	.861	.904	.880	.903	.882	.850	.867
L5	.791	.827	.826	.801	.843	.844	.810	.820
L6	.244	.532	.674	.696	.676	.725	.717	.609
L7	.472	.778	.713	.722	.719	.701	.798	.700
L8	.349	.596	.520	.621	.618	.510	.582	.542
L9	.514	.599	.711	.694	.742	.685	.675	.660
L10	.326	.731	.654	.679	.709	.596	.582	.611

* Prm : Parameters

L : Factor Loading (Lambda)
P : Factor Covariance (Phi)
D : Measurement Error (Theta Delta)

1 : Information
2 : Comprehension
3 : Arithmetic
4: Similarities
5 : Vocabulary
6 : Digit Symbol
7 : Picture Completion
8 : Spatial
9 : Picture Arrangement
10 : Object Assembly

Table 15 : The MAB Subtests as Measures of 'g' for Total Groups*

Good measures of 'g'		
subtest	Mean loading of 'g'	proportion of variance attributed to 'g' (%)
Information	.84	71
Comprehension	.88	77
Similarities	.87	76
Vocabulary	.82	67

Fair measures of 'g'		
subtest	Mean loading of 'g'	proportion of variance attributed to 'g' (%)
Arithmetic	.70	49
Picture Completion	.70	49
Digit Symbol	.61	37
Picture Arrangement	.66	44
Object Assembly	.61	37

Poor measure of 'g'		
subtest	Mean loading of 'g'	proportion of variance attributed to 'g' (%)
Spatial	.54	29

* Based on total groups

REFERENCES

REFERENCES

Alwin, D. F., & Jackson, D. J. (1980). Measurement models for response errors in surveys: issues and applications. In K. F. Schuessler (ed.), Sociological methodology 80. San Francisco: Jossey-Bass, pp.68-119.

Alwin, D. F., & Jackson, D. J. (1981). Applications of simultaneous factor analysis to issues of factorial invariance. In D. J. Jackson and E. F. Borgatta (Eds.), Factor analysis and measurement in sociological research: a multi-dimensional perspective. Beverly Hills: Sage Publications, Inc.

Anastasi, A. (1988). Psychological testing (6th ed.). New York: Macmillan.

Beck, N. C., Horwitz, E., Seidenberg, M., Parker, J., & Frank, R. (1985). WAIS-R factor structure in psychiatric and general medical patients. Journal of Consulting and Clinical Psychology, 53(3), 402-405.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88(3), 588-606.

Binet, A., & Simon, T. (1916). The development of intelligence in children. Vineland, N.J.: Training School Publication No. 11, 193.

Bohem, A. E. (1985). Educational applications of intelligence testing. In B. B. Wolman (Ed.), Handbook of intelligence; theories, measurement, and applications. N.Y.: Wiley and Sons.

Bollen, K. A. (1989). Structural equations with latent variables. New York: John Wiley & Sons.

Campione, J. C., & Brown, A. L. (1979). Toward a theory of intelligence: contributions from research with retarded children. In R. J. Sternberg, and D. K. Detterman (Eds.), Human Intelligence: Perspectives on its theory and measurement. Norwood, N.J.: Ablex Publishing Corporation.

Carmines, E. G., & McIver, J. P. (1981). Analyzing models with unobserved variables: analyses on covariance structures. In G. W. Bohrnsted and E. F. Borgatta (Eds.), Social Measurement: Current Issues. Beverly Hills: Sage Publisher.

Carroll, J. B. (1978). On the theory-practice interface in the measurement of intellectual abilities. In P. Suppes (Ed.), Impact of research on education: Some case studies. Washington, DC: National Academy of Education.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence : A critical experiment. Journal of Educational Psychology, 54, 1-22.

Cleary, T. A., Humphreys, L. G., Kendrick, A. S., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. American Psychologist, 30, 15-41.

Cohen, J. (1959). The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. Journal of Consulting Psychology, 23, 285-299.

Das, J. P., Kirby, J. R., & Jarman, R. F. (1975). Simultaneous and successive syntheses: An alternative model for cognitive abilities. Psychological Bulletin, 82, 87-103.

Dillon, W. R., & Goldstein, M. (1984). Multivariate analysis: methods and application. New York: John Wiley and Sons.

Fuller, Jr., E. L., & Hemmerle, W. J. (1966). Robustness of the maximum likelihood estimation procedure in factor analysis. Psychometrika, 31, 255-266.

Gleitman, H. (1986). Psychology (2nd. ed.). New York: W. W. Norton & Company, Inc.

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

Guilford, J. P. (1967). The Nature of Human Intelligence. New York: McGraw Hill.

Hoelter, J. W. (1983). The analysis of covariance structures: Goodness of fit indices. Sociological Methods and Research, 11(3), 325-344.

Hoffman, D. A., Mathieu, J. E., & Jacobs, R. (1990). A multiple group confirmatory factor analysis evaluations of teacher's work-related perceptions and reactions. Educational and Psychological Measurement, 50, 943-955.

Horn, J. L. (1967). Intelligence-why it grows, why it declines. Trans-action, 5, 23-31.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. Psychological Review, 75, 242-259.

Horn, J. L. (1978a). Human ability systems. In P. B. Baltes (Ed.)., Life-span development and behavior (Vol. 1, pp. 211-256). New York: Academic Press.

Horn, J. L. (1978b). The nature and development of intellectual abilities. In R. T. Osborne, C. E. Noble, & N. Wely (Eds.), Human variation: The biopsychology of age, race, and sex (pp. 107-136). New York: Academic Press.

Horn, J. L. (1985). Remolding old models of intelligence. In B. Wolman(Ed.), Handbook of intelligence (pp. 267-300). New York: Wiley.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. Acta Psychologica, 26, 107-129.

Jackson, D. N. (1984). Manual for the Multidimensional Aptitude Battery. Port Huron, MI: Research Psychologists Press.

Jackson, D. N. (forthcoming). 1990 norming study of MAB.

Jensen, A. R. (1979). g: Outmoded theory or unconquered frontier? Creative science and technology, 11, 16- 29.

Joreskog, K. G. (1969). A general approach to confirmatory factor analysis. Psychometrika, 34, 183-202.

Joreskog, K. G. (1970). A general method for analysis of covariance structures. Biometrika, 57, 239-251.

Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. Psychometrika, 36, 109-133.

Joreskog, K. G., & Sorbom, D. (1980). Structural equation models workshop. Cambridge, Mass.: Abt Associates, pp.26-29, June.

Joreskog, K. G., & Sorbom, D. (1984). LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares method. Mooresville, Ind: Scientific Software, Inc.

Joreskog, K. G., & Sorbom, D. (1989). LISREL 7: A guide to the program and applications (2nd ed.). Chicago: Joreskog & Sorbom/SPSS, Inc.

Kaplan, R. M. (1985). The controversy related to the use of psychological tests. In B. B. Wolman (Ed.). Handbook of intelligence: theories, measurements, and applications. New York: John Wiley & Sons.

Kaufman, A. S. (1975). Factor analysis of the WISC-R at 11 age levels between 6 and 16 years. Journal of Consulting and Clinical Psychology, 43, 135-147.

Kim, J. O., & Mueller, C. W. (1978a). Introduction to factor analysis: what it is and how to do it. Beverly Hills: Sage Publishers.

Kim, J. O., & Mueller, C. W. (1978b). Factor analysis: Statistical methods and practical issues. Beverly Hills: Sage Publishers.

Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rinehart and Winston.

Kranzler, J. H. (1991). The construct validity of the Multidimensional Aptitude Battery. Journal of Clinical Psychology, 47, 691-697.

Krieshok, T. S., & Harrington, R. G. (1985). A review of the Multidimensional Aptitude Battery. Journal of Counseling and Development, 64, 87-89.

Lee, M. S., Wallbrown, F. H., & Blaha, J. (1990). Note on the validity of the Multidimensional Aptitude Battery. Psychological Reports, 67, 1219-1222.

Levy, J. (1972). Lateral specialization of the human brain: Behavior manifestations and possible evolutionary basis. In J. A. Kieger (Ed.), Biology of behavior (pp. 159-180). Corvallis, OR: Oregon State University Press.

Liang, J., Lawrence, R. H., & Bollen, K. A. (1986). Age differences in the structure of the Philadelphia Geriatric Center for Morale Scale. Journal of Psychology and Aging, 1, 27-33.

Linn, R. L. (1989). Intelligence: measurement, theory, and public policy. Urbana & Chicago: University of Illinois Press.

Long, J. S. (1983). Confirmatory factor analysis: A preface to LISREL. Beverly Hills: Sage Publications.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: first- and higher order factor models and their invariance across groups. Psychological Bulletin, 97, 562-582.

Matarazzo, J. D. (1972). Wechsler's measurement and appraisal of adult intelligence (5th ed). New York: Oxford University Press.

Mehrens, W. A., & Lehmann, I. J. (1984). Measurement and evaluation in education and psychology (3rd ed.). New York: Holt, Rinehart, and Winston.

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: Mcgraw Hill

O'Grady, K. E. (1983). A confirmatory maximum likelihood factor analysis of the WAIS -R. Journal of Consulting and Clinical Psychology, 51, 826-831.

Pedhazur, E. J. (1982). Multiple regression in behavioral research: explanation and prediction. New York: Holt, Rinehart and Winston.

Reynolds, C. R. (1982). The problem of bias in psychological assessment. In C. R. Reynolds, & T. B. Gutkin (Eds.) The handbook of school psychology (pp. 178-208). New York: Wiley.

Reynolds, C. R., & Kaufman, A. S. (1985). Clinical assessment of children's intelligence with the Wechsler Scales. In B. B. Wolman (ed.) Handbook of intelligence: theories, measurements, and applications (pp. 601-661). New York: John Wiley and Sons.

Sattler, J. M. (1982). Assessment of children's intelligence and special abilities (2nd ed.). Boston: Allyn and Bacon.

Sattler, J. M. (1988) Assessment of children (3rd ed). San Diego: Jerome M. Sattler, Publisher.

Snow, R. E. in collaboration with E. Yalow (1982). Education and Intelligence. In R. J. Sternberg. (1982). Handbook of human intelligence. Cambridge: Cambridge University Press, 493- 585.

Sorbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. British Journal of Mathematical and Statistical Psychology, 27, 229-239.

Spearman, C. E. (1927). The abilities of man. New York: Macmillan.

SPSS/PC+. (1988). Statistical Packages for Social Sciences: Advanced Statistics V2.0. Chicago : SPSS Inc.

Sternberg, R. J. (Ed.) (1982). Handbook of human intelligence. Cambridge: Cambridge University Press.

Stockwell, R. G. (1984). Factor structure comparison between MAB and the WAIS-R. In L.J. Stricker(Chair), The

Multidimensional Aptitude Battery (MAB): A new group intelligence test. Symposium conducted at the meeting of the American Psychological Association, Toronto.

Stoddard, G. D. (1943). The meaning of intelligence. New York: Macmillan.

Tabachnick, B. G., & Fidell, L. S. (1989). Using multivariate statistics (2nd ed). New York: Harper Collins Publishers.

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

Thorndike, E. L. (1927). The measurement of intelligence. New York: Bureau of Publications, Teachers College, Columbia University.

Thurston, L. L. (1938). Primary mental abilities. Psychometric Monographs, No 1.

Thurston, L. L. (1947). Multiple factor analysis. Chicago: University of Chicago Press.

Tucker, L. R., & Lewis, C. A. (1973). A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

Vernon, P. A. (1985). Multidimensional Aptitude Battery. In D. J. Keyser and R. C. Sweetland (Eds.). Test critiques, vol. 2. Kansas City: Test Corporation of America.

Vernon, P. E. (1950). The structure of human abilities. New York: Wiley

Wallbrown, F. H., Carmin, C., & Barnett, R. (1988). Investigating the construct validity of the Multidimensional Aptitude Battery. Psychological Reports, 62, 871-878.

Wallbrown, F. H., Carmin, C., & Barnett, R. (1989). A further note on the construct validity of the Multidimensional Aptitude Battery. Journal of Clinical Psychology, 45, 429-433.

Wheaton, B., Muthen, B., Alwin, D. F., & Summers, G. F. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed.), Sociological Methodology 1977. San Francisco : Jossey-Bass, pp. 84-136.

Wechsler, D. (1958). The Measurement and Appraisal of Adult Intelligence. Baltimore: Williams & Wilkins.

Wechsler, D. (1981). Manual for the Wechsler Adult

Intelligence Scale-Revised. New York: Psychological Corp.

Wherry, R. J. (1959). Hierarchical factor solutions without rotation. Psychometrika, 24, 45-51.

Wherry, R. J., & Wherry, R. J. Jr. (1969). WHEWH program. In R. J. Wherry (Ed.), Psychological department computer programs. Columbus, OH: Department of Psychology, The Ohio State University.

ABSTRACT

THE MULTIPLE GROUP CONFIRMATORY FACTOR ANALYSIS
OF THE MULTIDIMENSIONAL APTITUDE BATTERY

by

TAIHYONG MOON
DECEMBER 1993

Adviser: Dr. D. MARCOTTE
Major: EVALUATION AND RESEARCH
Degree : DOCTOR OF PHILOSOPHY

This research focuses on the examination of the psychometric properties of the Multidimensional Aptitude Battery (the MAB). The purpose of this study is to factor analyze the data obtained from the 1990 norming study using the maximum likelihood confirmatory factor analysis in order to investigate the extent to which the proposed two factor model explains the variation in the MAB subtests and to examine the factorial invariance across age groups.

Data for this research were obtained from the 1990 norming study of the MAB. The total sample size is 1400, approximately evenly divided by gender. The subjects were selected from six arbitrary regions of the United States and Canada. Only the groups which contained 200 subjects were selected for this research.

For this study the two factor (Verbal factor and Performance factor) model was specified. All hypotheses were tested by using the covariance matrices from the 10 MAB subtests scores for each age group. LISREL VII was used for the data analyses. The X^2 tests of fits of the proposed two

factor model and the null model to the observed data was conducted. The X^2 test was used to determine whether the underlying structure of the observed data was explained by the proposed model. A X^2 difference test was used to assess which model better explained the data. Furthermore, Wheaton et al.'s ratio of X^2 to degrees of freedom, and Bentler and Bonett's Normed Fit Index (NFI) were also obtained.

The factorial invariance across age groups was examined. The factorial invariance was investigated in terms of three parameter matrices (the matrices of factor loadings, of variances and covariances for measurement errors, and of factor variances and covariances) as well as the common number of factors across age groups. Five hypotheses were tested with X^2 tests and X^2 difference tests. Sub-hypotheses of the equal number of factors, invariant factor loadings, invariant measurement errors, and invariant factor covariances constituted a hierarchy.

The proposed two factor model well explained the subtest variation of the MAB. The two factor model better explained the data than the null model with no common variances. Furthermore, the factorial invariance across age groups was also supported. From the results of this research, it is implied that if used for group testings, the MAB can be an alternative for the WAIS-R.

AUTOBIOGRAPHICAL STATEMENT

Name : Taihyong Moon

Birth of Date : September 21, 1957

Place of Birth : Pusan, Korea

Education :

B.A., Dong-A University, Pusan, Korea, 1985
Major: Education

M.A., Wayne State University, 1986
Major: Educational Sociology

Honors and Awards

Honor Student Scholarship,
Dong-A University, 1982-1984

Graduate Professional Scholarship,
Wayne State University, 1986-1990

Professional Memberships :

The Gerontological Society of America