WECHSLER INTELLIGENCE SCALE FOR CHILDREN- FOURTH EDITION (WISC-IV): ADAPTATION, TRANSLATION, AND STANDARDIZATION IN PAKISTAN

BY

SAIMA AMBREEN

Dr. Muhammad Ajmal
NATIONAL INSTITUTE OF PSYCHOLOGY
Center of Excellence
Quaid-i-Azam University, Islamabad

2014

WECHSLER INTELLIGENCE SCALE FOR CHILDREN- FOURTH EDITION (WISC-IV): ADAPTATION, TRANSLATION, AND STANDARDIZATION IN PAKISTAN

By

SAIMA AMBREEN

A dissertation submitted to

Dr. Muhammad Ajmal
NATIONAL INSTITUTE OF PSYCHOLOGY
Centre of Excellence
Quaid-i-Azam University, Islamabad.

In partial fulfillment of the requirement for the degree of

DOCTOR OF PHILOSOPHY

IN

PSYCHOLOGY

2014

WECHSLER INTELLIGENCE SCALE FOR CHILDREN-FOURTH EDITION (WISC-IV): ADAPTATION, TRANSLATION, AND STANDARDIZATION IN PAKISTAN

By

SAIMA AMBREEN

Approved by

(Prof. Dr. Anila Kamal)
Supervisor

External Examiners

(Prof. Dr. Mehr Bano)
External Examiner

(Dr. Tanvir Akhtar)
External Examiner

(Prof. Dr. Anila Kamal)
Director, NIP

LIST OF FOREIGN EVALUATORS

The Ph.D. dissertation on "Wechsler Intelligence Scale for Children- Fourth Edition (WISC-IV): Adaptation, Translation, and Standardization in Pakistan" prepared by Miss Saima Ambreen has been evaluated by the following foreign examiners / evaluators:

1. Prof. Dr. Ralph W. Hood, Department of Psychology, College of Arts and Science, The University of Tennessee at Chattanooga, USA.
2. Prof. Thomas Holtgraves, Department of Psychological Sciences, College of Science and Humanities, Ball State University, USA.
3. Prof. Dr. Jennifer E. Lansford, Center for Child and Family Policy, the Duke University, Durham, USA.

CERTIFICATE

It is certified that Ph.D. dissertation on "Wechsler Intelligence Scale For Children- Fourth Edition (WISC-IV): Adaptation, Translation, and Standardization in Pakistan" prepared by Miss Saima Ambreen has been approved for submission.

(Prof. Dr. Anila Kamal)
Supervisor

List of Contents

Acknowledgements

All praise to Allah Almighty, who bestowed upon me the rigor and courage to complete this research successfully. The present work can not be an individual effort. Many hands and many minds are behind it. Even if I miss the names of these individuals in the following lines, I owe them much for their help and concern.

I am deeply grateful to my supervisor, Prof. Dr. Anila Kamal as her knowledge, technical expertise, and guidance enabled me to complete this challenging task. She remained a source of inspiration throughout the research and due to her brilliance and empirical approach; she will remain my role model for the rest of my life.

I would also like to pay my gratitude to Pearson Assessment, USA for allowing me to work on Wechsler Intelligence Scale for Children (WISC-IV). An agreement between NSC Pearson Private Limited and NIP has made this research possible. My special thanks to Angelina Mao and Dr. Joseph Michelle from NCS Pearson, India for their help, support, and guidance.

I am indebted to Dr. Mah Nazir Riaz, Dean and Chairperson, Department of Psychology, Frontier Women University, Peshawar for her expert review of adaptations in verbal subtests. I am also much grateful to employees of National Language Authority, especially to Dr. Anwar Ahmed, Ex. Director General and Syed Sardar Ahmed Pirzada; their expert opinion helped me a lot in finalizing translations of various subtests.

I would also like to thank Mr. Atif Mehmood Kiyani, Ex-Director General; Mr. Zafar Iqbal, Assistant Director Training, and other staff of Training section of Federal

Directorate of Education, Islamabad for allowing me to have test administrations in their schools and colleges. Similarly, I am thankful to Maj. Gen. Zahid Parvez, Ex-Director Welfare and Education, and Col. Muhammad Zubair, Senior Manager Walfare- Colleges and Campuses, Fauji Foundation for permitting me to take data from their prestigious institutes in the selected 10 districts of Pakistan. Special thanks to Principals and heads of all schools or colleges from which I selected sample for test administration.

I would like to thank all faculty members of NIP specially Dr. Rubina, Dr. Jamil, Dr. Moazama, and Dr. Humaira Jami for being part of my research work as members of my Expert Committee. I could not precede even step further without their precious time and thoughtful suggestions during our long committee approaches. Many thanks to the NIP faculty members who were member of my translation team including Ms. Rabia, Ms. Shaista, Ms. Nelopher, Ms. Raiha, and Ms. Irum Naqvi.

Much appreciations for the supporting staff of NIP including the computer laboratory staff, library staff, and the administration staff who always provided me help and required resources. My special thanks to Mr. Tahir Elahi for his continuous assistance in editing content of the manual and for his help in maintenance of my research data. Mr. Qayyum's help in finding litrary material; and Mr. Khalid's and Mr. Mohsin's help in maintaining Ph.D. related paper work secured me from many due stresses.

My team of test administrators Sufia, Amara, Mahpara, and Tabasum deserve much more than a thanks for their help in test administrations. A very special thanks and appreciation for great support of my students cum team members Sara, Maria, Aneeqa, and Misbah throughout the research period.

I have no words to express my gratitude for my family specially parents. Without their love, encouragement, and support I would not be able to achieve anything in my life. They have shown great tolerance for my long working hours and long absence from home for various research works. The concern, support, and encouragement of my brother, sister, and sister in law remained the source of motivation throughout my research. The love and attachment of my niece and nephews Linta, Rebal, and the youngest Shayaan boosted my spirits at the time of extreme stress.

I am thankful to my sincere friends, who supported me at every step of my work. Mr. Waseem Fayyaz with his scholarly discussions and confrontations always gave me guidance and motivation to achieve more. My hostel roommate and dearest friend Sitwat Aman with her immense tolerance for my studying habits was a big support for me throughout this period. My friends Mussarat, Nadia, and Fatima were always very helpful and supportive. I would also like to thank Miss Saba Farooq for her continuous help, encouragement, and enjoyable company.

I also owe deep appreciation for all the students who participated in the research for their time, feedback, and cooperation. Last but not the least; I wanted to thank all the participant teachers and members of my advisory panels, as without their insight and knowledge I would not be able to do this research.

Saima Ambreen

Abstract

The present research was primarily aimed to adapt, translate, and standardize WISC-IV for Pakistan. The research was conducted in collaboration with NCS Pearson Private Limited, India as part of a larger project aiming at standardization of WISC-IV for South Asia. Research process was completed through three studies. Study-I was concerned with the adaptation and translation of WISC-IV and it was completed through two phases. Phase I involved pre-testing of the original WISC-IV South Asia subtests (n=12). Its findings not only identified few items needing adaptation but also suggested translation of instructions and/or item content of all subtests (including performance subtests). Phase II was concerned with the steps involved in proper adaptation and translation of the WISC-IV subtests. It involved a priori procedures (judgmental procedures) for adaptation and translation of the subtests. These procedures included multiple-forward translation, committee approaches, and expert reviews. This process resulted in development of Urdu adaptation of WISC-IV in which child directed instructions for all subtests and item content of all verbal subtests have been translated into Urdu along with various adaptive changes.

Study II was aimed to evaluate the psychometric properties of the WISC-IV ^{PAK} *through conduction of tryout I (n=33), tryout II (n=88), and tryout III (n=110). Initial tryout assessed functioning and comprehensibility of items through response frequency and reliability analysis only. But other two tryouts involved Classical Test Theory (CTT) and Item Response Theory (IRT) based psychometric techniques to improve, item reliability, item difficulty, item discrimination, and item fit. This detailed item analysis*

involved item re-ordering of nine subtests and changes in item content and/or sampled responses of few other subtests and resulted in finalization of WISC-IV ^{PAK}.

The study III involved the standardization of the WISC-IV ^{PAK} and was completed through four phases. All the analyses conducted in this study are based on a normative sample of 800 children (50% girls) selected through following a stratified random sampling design. The normative sample was stratified into 11 age groups, two gender groups, five geographical regions, and three parental education levels. It was selected from govt., semi-govt., and private schools and colleges situated in 10 districts of Pakistan. In the first phase of study III, subtests temporal stability and internal consistency evidence was established. In the phase II a multi-model multi-trait matrix method was utilized to establish convergent and discriminant validity of WISC-IV ^{PAK}. The cross validation of WISC-IV factorial structure in Pakistan was also done during this phase. Pakistani norms for WISC-IV ^{PAK} were developed for 11 age groups of one year ranging from 6 to 16 years and 11 months in the phase III of the standardization study. Two types of norms including standard score norms (scaled scores and composite scores) and test-age equivalent norms were derived. Comparison of WISC-IV ^{PAK} composite scores by using Pakistani and UK norms was also conducted in this phase. In the last phase of study III influence of variables like age, gender, geographical region, and parental education level on children's intelligence level was explored. This concluded the efforts to provide a reliable and well-standardized tool to measure intelligence of Pakistani children.

List of Tables

List of Appendices

List of Figures

INTRODUCTION

Chapter I

Introduction

Intelligence and Intelligence Quotient (IQ) are among the most important concepts of psychology, yet these are also the most complex and sometimes disputatious issues. Researchers focus more on psychometric intelligence because it is considered as general academic aptitude and have strong educational implications (Storfer, 1990). Intelligence testing with its strong history and ever growing scope can be regarded as mother of psychometrics. From the last nearly one hundred years, intelligence testing has gained much attention in psychological research, clinical practice, and psycho-educational assessment throughout the world. Even in developing countries like Pakistan, growing progress in research, education, and mental health services is creating increased need for intelligence testing.

In Pakistan, Armed forces, Federal Public Service Commission, and Provincial Public Service Commissions are using many intelligence tests. But most of these tests are used and constructed for the specific purpose of personnel selection, so they may have psychometric limitations when used for other purposes. In clinical and educational settings few popular standardized intelligence tests such as Raven Progressive Matrices (Raven, 1936) and Otis-Lennon School Ability Test (Otis & Lennon, 1979) are also in use (as cited in Hashmi, 2000). These well-constructed standardized tests are of limited use in Pakistan due to differences from culture in which they were developed. For example, being less familiar task a test having only matrices can not effectively gauge intelligence in Pakistan. Items on an intelligence test mirror the culture of the society in which the test is developed and used. So great care is required while using these otherwise psychometrically strong tests in cultures that differ from the culture in which they were developed. This problem of cultural relevance can be resolved by proper adaptation and translation of such tests for the specific culture in which it is to be used. So present study is conducted to adapt, translate, and standardize the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV; Wechsler, 2004) for Pakistani population.

Intelligence

Intelligence has been a topic of concern and curiosity among philosophers, employers, educators, psychologists, anthropologists, and lay people that extends back hundreds of years (Cianciolo & Sternberg, 2004). But attempts to develop a precise but comprehensive definition of intelligence have been fraught with difficulty and debate due to its abstract and metaphorical nature. It is a general label for a group of internal processes that are inferred from more observable behaviors and responses and has been considered as a multifaceted faculty that displays itself in various ways throughout the life.

Numerous attempts have been made to define intelligence, but the first people to ponder the nature of intelligence were not psychologists or educators, but were philosophers. The ancient Greek philosopher Plato has resembled intelligence to wax blocks of different size, hardness, moistness, and purity. To him a person who is suffering from intellectual deficits would have an overly hard or soft and muddy wax block. Thomas Aquinas believed that intelligent people have more complete and universal comprehension skills than less intelligent people. While, eighteenth century philosopher Immanuel Kant was of the view that intelligence has many kinds or facets, and individuals clearly differ in the degree to which they possess them (as cited in Cianciolo & Sternberg, 2004). Galton (1869, 1883) saw intelligence as the underlying mental strength (energy) and believed it to have a physical or sensory-motor attribute (as cited in Mahmood, 1991), while Binet believed that intelligence includes abilities like reasoning, judgment, memory, and abstract thinking but emphasized judgment as the fundamental faculty (Binet & Simmon, 1916). Similarly, Wechsler in 1958 conceptualized intelligence as global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment (as cited in Marnat, 1990).

Cohen and Swerdlik (2005) has discussed four major approaches to study intelligence including the developmental or Piagetian approach, the neurological-biological approach, the cognitive or information processing approach, and the psychometric or factor-analytical approach. These approaches along with their relevant theories can be best understood on the basis of seven metaphors of mind that have guided scholarly exploration into the nature of intelligence.

Geographic Metaphor. Theories of intelligence that correspond to a geographic metaphor represent an attempt to develop a map for the human mind. These mental maps describe structure of mind and underlying intellectual abilities. Modern geographic theories of intelligence are devoted to identify the basic intellectual abilities, called ability factors, that supposedly underlie the range of intelligent acts and they also assume that people differ in level to which they possess these abilities/set of abilities or factors (as cited in Cianciolo & Sternberg, 2004).

Factor-analytical Theories of Intelligence. Statistically based factor-analytical research has resulted in emergence of various theories of intelligence since long. These geographic theories mainly differ in number of factors and in the particular kind of identified factors.

Among many presented theories of intelligence that were based on factor analysis is the Two-Factor theory of Intelligence formulated by Spearman in 1927. He proposed the existence of a general intellectual ability factor '*g*' and specific ability factor '*s*' in intelligence. He was primarily interested in what is common among various intellectual abilities rather than their differences. Criticisms of the two-factor theory have not been lacking, and many alternative theories have been proposed. Thorndike, for example, formulated a theory that viewed intelligence as a composite of many different abilities interconnecting in the brain. He proposed three kinds of intelligence labeled as social, concrete, and abstract. Similarly, Multiple-Factor theory was given by Thurston in 1938. He conceived intelligence as consisting of seven 'primary abilities' including Verbal Meaning, Number Facility, Inductive Reasoning, Perceptual Speed, Spatial Relations, Memory, and Verbal Fluency. The theory of Structure-of-intellect was proposed by Guilford (1967). He explained mental activities by deemphasizing the '*g*' factor. In his model there are not less than 120 distinctive abilities based on five kinds of operations (cognition, memory, divergent thinking, convergent thinking, and evaluation), four types of content (figural, symbolic, semantic, and behavioral), and six products (units, classes, relations, systems, transformations and implications). Later revisions of Guilford's model featured 150 and even up to 180 distinct abilities (as cited in Cohen & Swerdlik, 2005).

Factor-analytical Theories impelling Intelligence Testing. From the end of 20^{th} century, an intelligence theory that was proposed by Cattell (1971) has gained growing attention from test constructors and test users. The theory postulated the existence of two major types of cognitive abilities: crystallized intelligence, *Gc* (includes acquired skills and knowledge that is cultural-specific and its application such as vocabulary) and fluid intelligence, *Gf* (includes non-verbal, relatively cultural-free knowledge such as memory for digits). Modifying Cattell's theory, Horn (1991) gave the idea add several factors to Cattell's two factors like visual processing (*Gv*), auditory processing (*Ga*), quantitative processing (*Gq*), speed of processing (*Gs*), facility with reading and writing (*Grw*), short term memory (*Gsm*), long-term storage and retrieval (*Glr*). He agreed with the existence of crystallized and fluid intelligence but differed with Cattell in suggesting that both these factors are learned and are also based on heredity. He says that *Gf* is based on casual learning, while *Gc* is based on cultural and school type learning.

Carroll (1997) presented his Three-Stratum Theory of cognitive abilities based on the factor-analytical studies. On top stratum of Carroll's model is '*g*' or general intelligence; whereas, the second stratum consists of eight abilities or processes: Fluid Intelligence (*Gf*), Crystallized Intelligence (*Gc*), General Memory and Learning (*Y*), Broad Visual Perception (*V*), Broad Auditory Perception (*U*), Broad Retrieval Capacity (*R*), Broad Cognitive Speediness (*S*), and Processing/ Decision Speed (*T*). While, the third stratum is composed of several 'level factors' and 'speed factors', each varies on the bases of the second level stratum to which they are associated. So, Carroll's model is hierarchical in nature (as cited in Cohen & Swerdlik, 2005).

The Cattell-Horn and Carroll models have various similarities with each other. Beside the differences including the presence of a general '*g*' factor, an integration of these two models was proposed by McGrew (1997). The result was CHC model (Cattell-Horn-Carroll model). This model has greatly influenced the recent test development including the development and interpretation of fourth edition of WISC (Wechsler, 2003) and fifth edition of Stanford Binet test (SB5; Roid, 2003).

The Cattell-Horn-Carroll (CHC) model was subsequently modified on the basis of further factor-analytical work by McGrew and Flanagan in 1998. In its recent form the CHC model propose to have ten "broad-stratum" abilities and more than seventy "narrow-stratum" abilities, each of the broad-stratum ability has linking with two or more narrow-stratum abilities. The ten broad-stratum abilities are: Fluid Intelligence (*Gf*), Crystallized Intelligence (*Gc*), Quantitative Knowledge (*Gq*), Reading/ Writing Ability (*Grw*), Short-Term Memory (*Gsm*), Visual Processing (*Gv*), Auditory Processing (*Ga*), Long-Term Storage and Retrieval (*Glr*), Processing Speed (*Gs*), and Decision/Reaction Time or Speed (*Gt*). This model gives no apparent place to the general intellectual ability factor '*g*' but this does not mean that the integrated model does not subscribe to a separate general human ability or that '*g*' does not exist. Rather, it was omitted as it has little practical relevance in psycho-educational assessment and the model was proposed in an effort to improve the practice of psycho-educational assessment (McGrew & Flanagan as cited in Cohen & Swerdlik, 2005).

Geographic metaphor based theories explain the abilities underlying intelligence well that led to their measurement but they do not clarify how intelligence or intellectual abilities work, or how mental processes involved in these abilities work.

Computational Metaphor. These theories of intelligence use terms such as 'information processing' to explain the manner in which information is received, stored, and retrieved, and the ways these processes eventually result in a response. A representative information processing theory is the work of Campione and Brown (1978). They have divided process of cognition into two kinds. The meta-cognitive processes are used to control one's information processing, while other cognitive processes are used to implement task strategies. Another information processing theory is Sternberg's (1985) Cognitive Component or Triarchic Theory, who distinguishes three different kinds of information processing components: Meta components (higher order control processes), performance component (lower-order processes) and the knowledge component (processes involved in learning and storing new information). Sternberg emphasized that intelligence must be purposeful, goal oriented, relevant, and must also involve the development of effective information processing (as cited in Marnat, 1990). The computational theories though give satisfactory information regarding how intellectual abilities work, but the link between cognitive mechanisms and actual neurological functions is unclear in these theories (as cited in Cianciolo & Sternberg, 2004).

Biological Metaphor. This metaphor is based on the fact that all thoughts originate in the brain, so eventually intelligent behavior can be traced back to its biological source. All approaches to understand intelligence assume that there is an underlying neurological substrate on which intelligence is ultimately dependent. Halstead (1961) theorized that a number of brain functions relating to intelligence are biologically based and are relatively independent of cultural considerations. Two other contributors to biological approaches are Cattell (1963) and Hebb (1972). Both of them emphasized the existence of certain areas of intelligence that are innate and biological. Cattell refers to this as the fluid intelligence, while Hebb calls this as intelligence A. Biological approaches to intelligence generally have serious methodological and theoretical difficulties. So far, no specific neurological substrates have been found that clearly relate to intelligence. It is also extremely difficult to separate the effects of learning and culture from a hypothesized underlying biological structure (as cited in Marnat, 1990).

Epistemological Metaphor. Any epistemological theory of intelligence must have knowledge acquisition as its central focus. The basis of epistemological theories of intelligence rests largely on Piaget's theory of cognitive development (as cited in Cianciolo & Sternberg, 2004).

Piaget (1971, 1972) viewed intelligence as a special form of biological adaptation between a person and his or her environment. Piaget hypothesized that adaptation and the learning occurs with the help of two central mental operations: assimilation (actively organizing new information so that it fits in with what already is perceived and thought) and accommodation (changing what is already perceived or thought, to fit in with new information). He also emphasized the importance of physical activities and social peer interaction in producing the disequilibrium that leads the process through which mental structures change. Piaget stressed that both assimilation and accommodation occur simultaneously, independent of age, but also within all age groups. However, within these general processes there are specific age-related differences and Piaget has described four stages of cognitive development (as cited in Cohen & Swerdlik, 2005).

Sociological Metaphor. The influence of society on intellectual development is the focal point of sociological theories of intelligence (see for example Vygotsky, 1978 & Feuerstein, 1979). These theories highlighted the view that each of us is a collaborator in the improvement of people's intelligence, particularly of children. Vygotsky (1978) viewed culture as central to intellectual development. He argued that people use 'psychological tools' like language, imagery, thinking styles and other artifacts in the human culture to enhance the thinking of other people. He presented the concept of 'the zone of proximal development' to describe the situation in which psychological tools are shared and learned. He defined this zone as the difference between what a person is capable of doing without any help and what he is able to do with other's assistance or guidance. The greater is this difference between what people can do with help verses without help, the greater the zone of proximal development (as cited in Cianciolo & Sternberg, 2004). Similarly, Feuerstein (1979, 1980) stated that much of the intellectual development results from the mediation of the environment by mother or any other adult (as cited in Sternberg, Lautrey, & Lubart, 2003).

Anthropological Metaphor. This conceptualization of intellectual development considers culture as fundamental to describe what it means to be intelligent. Many studies concluded that people in different cultures may develop somewhat different kind of intellectual abilities, depending on what types of intellectual capabilities are appreciated by their particular culture (Heath; Okagaki & Sternberg as cited in Cianciolo & Sternberg, 2004). These theories highlighted the ethnocentric influences on the intelligence, but they still do not provide any explanation that why people in the same culture possess different intellectual abilities (as cited in Cianciolo & Sternberg, 2004).

System Metaphor. System theories view intelligence as set of multiple mutually dependent parts, or even multiple intelligences. These theories attempt to integrate multiple perspectives of intelligence. Gardner's Theory of Multiple Intelligences (1983, 1994) is one such example as it integrates methodological approaches and findings from the geographic metaphor, biological metaphor, and anthropological metaphor. Gardner conceived seven types of interdependent intelligences including logical-mathematical, bodily-kinesthetic, linguistic, musical, spatial, interpersonal, and intrapersonal. Sternberg's triarchic theory (1997) of successful intelligence is another system theory as it integrates geographic, computational, and anthropological metaphors. Ceci's biological model of intelligence (1996) involves all of the metaphors of intelligence. He argued existence of multiple 'cognitive potentials' that are biological predispositions. Cognitive potentials, knowledge, and environmental context interact to establish individual differences in the development of behaviors or in displaying intelligent behavior (as cited in Cianciolo & Sternberg, 2004).

Intelligence Testing

Intelligence is considered as one of the most important factor of an individual's personality and intelligence testing has gained much attention in psychological research, clinical practice, and psycho-educational assessment. As intellectual assessment has become a usual component of vocational, psycho-educational, and neuropsychological testing; so many tests have been constructed and are in use to assess the intellectual level of different individuals. Some of these tests are individually administered while others are administered in group settings.

Group tests are mostly considered useful screening tools when large numbers of examinees have to be evaluated either at the same time or within a limited time period as large number of test takers can be tested at one time, resulting in efficient use of time and resources. Many group intelligence tests are in use in armed forces, educational, and school setting, for example, the Armed Services Vocational Aptitude Battery (ASVAB, 1968; CAT-ASVAB, 1996, 1997), Otis-Lennon School Ability Test- OLSAT (Otis & Lennon, 1979; OLSAT- 7th Edition, 1995), the California Test of Mental Maturity (CTMM; Sullivan, Clark, & Tiegs, 1936, 1959), and the Cognitive Abilities Test (CogAT- Form 6; Lohman & Hagen, 2001) (as cited in Kiff, 2010).

Most of the intelligence tests are individually administered tests. Some of them have theoretical basis like Primary Mental Abilities Test (1938) was based on Thurston's theory of Primary Mental Abilities (PMAs). But through history it seemed that intelligence tests have been developed more out of necessity than anything else. For instance, Binet's efforts in 1904 for developing a test to screen developmentally disabled children resulted in first formal test of intelligence, the Binet-Simon scale in 1905 (as cited in Cohen & Swerdlik, 2005). From that time the intelligence testing spread all over the world. There are many individual intelligence tests but most popular and widely used of them are the Stanford-Binet Intelligence Scales and the Wechsler Intelligence Test series.

Binet Intelligence Scales. The first of original Binet Intelligence scales came in 1905 as Binet-Simmon Scale. It was revised in 1908 and then in 1911, while a downward extension of the test came in 1912 that brought down the age range to 3 months (Kuhlmann as cited in Cohen & Swerdlik, 2005). Binet scale underwent many adaptations and consequently, Stanford-Binet Intelligence Scale was published in 1916. It was the first intelligence test that employed the concept of Intelligence Quotient (IQ). It went through many revisions including the 1960 revision that came with the concept of "deviation IQ" to replace the concept of "ratio IQ", with the mean of 100 and standard deviation of 16. The most recent and the 5th edition of the Stanford-Binet (SB5; Roid, 2003 as cited in Cohen & Swerdlik, 2005) was intended for administration to assesses as young as 2 and as old as 85 years (or older). The test gives a Full Scale IQ resulting from the administration of 10 subtests and five Factor Index scores that corresponds to the five factors which the test is supposed to measure including Fluid Reasoning (FR), Knowledge (KN), Quantitative Reasoning (QR), Visual-Spatial Processing (VS), and Working Memory (WM). So the SB5 is based on the Cattell-Horn-Carroll (CHC) model of intellectual abilities (Cohen & Swerdlik, 2005).

The Wechsler Tests. Wechsler constructed a series of individually administered intelligence tests to appraise the intellectual abilities of people through preschool to adulthood. The series includes: Wechsler Adult Intelligence Scale (WAIS) and its revisions, Wechsler Intelligence Scale for Children (WISC) and its revisions, and Wechsler Preschool and Primary Scale of Intelligence (WPPSI) for much younger children. All Wechsler tests are scaled to yield a deviation IQ with a mean of 100 (interpreted as average) and standard deviation of 15.

Wechsler Adult Intelligence Scale (WAIS). Initially the Wechsler test was named as Wechsler-Bellevue Scale W-B (1939) and was criticized because of its poor standardization sample. Later in 1942, it was revised as W-B II. The test was revised again as Wechsler Adult Intelligence Scale (WAIS) in 1955. The WAIS scoring resulted into a Verbal IQ, a Performance IQ, and a Full Scale IQ. WAIS-R was published in 1981 while the third edition of the test came in 1997 as WAIS-III. The WAIS-III had the updated and colored materials. It assesses the IQ of individuals between the age ranges of 16 to 89 years. Three subtests were added in the WAIS-III including Symbol Search, Letter-Number Sequencing, and Matrix Reasoning (Cohen & Swerdlik, 2005). WAIS- IV (Wechsler, 2008) is the most recent revision and is comprised of four indices along with two generated broad scores which can be used to summarize general intellectual abilities: Full Scale IQ - FSIQ (based on the total combined performance on the VCI, PRI, WMI, and PSI), and General Ability Index - GAI (based only on the six subtests that comprise the VCI and PRI). The two new tests included are Visual Puzzles and Figure Weights, they are measures of non-verbal reasoning and analogical reasoning.

Wechsler Preschool and Primary Scale of Intelligence (WPPSI). The publication of WPPSI in 1967 resulted in downward extension of the age range of Wechsler tests to age 4. Its revision WPPSI-R was introduced in 1989. It was intended to assess the intelligence of children from ages 3 years through 7 years and 3 months. In 2002 publication of WPPSI-III extended the age range of the children who could be tested with this instrument downward to 2 years 6 months. This test yields three composite scores as Verbal IQ, Performance IQ, and a Full Scale IQ. Many previous subtests were dropped from WPPSI-III including: Arithmetic, Animal Pegs, Geometric Design, Mazes, and Sentences, while seven new subtests were added including: Matrix Reasoning, Word Reasoning, Picture Concept, Coding, Symbol Search, Receptive Vocabulary, and Picture Naming (Cohen & Swerdlik, 2005).

Wechsler Abbreviated Scale of Intelligence (WASI). It was introduced in 1999 and has an age range of 6 to 89 years. Its average completion time is 15 to 35 minutes. It is available in two or four subtests forms. Two subtest forms include the subtests of Vocabulary and Word Reasoning and takes about 15 minutes in completion. While the four subtest form include subtests of Vocabulary, Block Design, Similarities, and Matrix Reasoning, and takes about 35 minutes in its completion. It yields measure of Verbal, Performance, and Full Scale IQ (Wechsler, 1999).

Wechsler Intelligence Scale for Children (WISC). It is an individually administered clinical instrument for assessing the cognitive ability of children aged 6 years through 16 years 11 months. As a downward extension of the Wechsler Bellevue II, WISC was introduced in 1949. Its first revision came in 1974 as WISC-R. The Psychological Corporation revised it again in 1991 as WISC- III. The latest revision came in 2003 as WISC- IV. Researches on WISC-R suggested the probable existence of a third factor labeled as Freedom from Distractibility along with the two central factors, the Verbal IQ and the Performance IQ. Whereas, the WISC-III revisions attempted to strengthen this factor, and developed a fourth one too. Thus, the WISC-III offered a FSIQ (Full Scale IQ) as a measure of 'g', a Verbal and a Performance IQ, and four new Indices (an index is created where two or more sub-tests are related to a basic underlying skill). These include the Verbal Comprehension Index, Perceptual Organization Index, Freedom from Distractibility Index, and Processing Speed Index. WISC-III consisted of 13 subtests, three of which were supplementary (Cohen & Swerdlik, 2005).

Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV). The WISC-IV is an updated version of the WISC-III, and have 10 core subtests, and 5 additional/ supplemental subtests. These subtests yield four indices and one Full Scale IQ. The FSIQ can range from 40 at the lowest to 160 at the highest. Three subtests can be administered in advance forms to have additional assessment of processing abilities. Children assessed with the WISC-III were when reassessed with the WISC-IV showed about a 5 point decrease in FSIQ. This is because of the new aspects of the test and the novelty of some of the new items and subtests. It can be administered to the children of 6 years to 16 years and 11 months of age.

Revision Goals for the WISC-IV. Revision goals were drawn from ten years of research with the WISC-III; advice from experts in the field of neuropsychology, clinical psychology, and school psychology; and an extensive review of literature in the areas of intelligence theory, intellectual assessment, cognitive development and cognitive neuroscience (WISC-IV Technical and Interpretive Manual; Wechsler, 2003). The five primary revision goals are to update the instrument's theoretical foundation, to enhance its clinical utility, to increase developmental appropriateness, to improve its psychometric properties, and to increase user friendliness.

Structure of the Scale. The structure of the WISC-IV has been updated to reflect current theory and practice of cognitive assessment in children, including increased attention to working memory and processing speed. A total of five composite scores can be derived with the WISC-IV, Full Scale IQ (FSIQ) composite scores and four indices composite scores (WISC-IV Technical and Interpretive Manual; Wechsler, 2003). The four composite indices are:

1. Verbal Comprehension Index (VCI) having 5 subtests
2. Perceptual Reasoning Index (PRI) having 4 subtests
3. Working Memory Index (WMI) having 3 subtests

4. Processing Speed Index (PSI) having 3 subtests

Verbal Comprehension Index (VCI): It requires verbal conceptualization, stored knowledge access and oral expression. The VCI measures verbal concept formation. It assesses children's capacity to listen to a question, draw upon learned information from both formal and informal education, use their reasoning, and express their thoughts verbally. It can identify preferences for verbal information, a difficulty with novel and unexpected situations, or a desire for more time to process information rather than to decide immediately. It is considered as a good predictor of readiness for school and achievement orientation, but can be influenced by background, education, and cultural opportunities. The subtests included in this index are: Similarities, Vocabulary, Comprehension, Information, and Word Reasoning.

Perceptual Reasoning Index (PRI): It requires visual perception, organization and reasoning with visually presented, nonverbal material to solve the kinds of problems that are not school taught. The PRI basically measures non-verbal and fluid reasoning. It tests children's ability to examine a problem, draw upon visual-motor and visual-spatial skills, organize their thoughts, create solutions, and then test them. It can also identify preferences for visual information, comfort with novel and unexpected situations, or a preference to learn by doing. It includes sub-tests: Block Design, Picture Concepts, Matrix Reasoning, and Picture Completion.

Working Memory Index (WMI): It requires working memory processes applied to the manipulation of orally presented verbal sequences. It is the ability to temporarily retain information in memory for performing some operation or manipulation to produce a result. It involves attention, concentration, mental control, and reasoning. It is closely related to achievement and learning (Learning Disability students frequently affected). It has its importance in higher-order thinking, learning, and achievement. It can measure concentration, planning ability, cognitive flexibility, and sequencing skill, but it also seems sensitive to anxiety. It is an important constituent of learning and achievement, and ability to self-monitor. Its sub-tests are: Digit Span, Letter-Number Sequencing, and Arithmetic.

Processing Speed Index (PSI): It requires visual perception and organization, visual scanning, and the efficient production of multiple motor responses. These tasks require executive control of attention and sustained effort for a 2-minute period of time while working with simple visual material as quickly as possible. It measures processing speed and assesses children's abilities to focus attention and quickly scan, discriminate between, and sequentially order visually presented information. It requires persistence and planning ability, but is sensitive to motivation, difficulty in working under time pressure, and motor coordination too. Cultural factors also seem to have little effect on it. It is related to Working Memory in that increased processing speed can decrease the load placed on working memory, while decreased processing speed can impair the effectiveness of working memory. The sub-tests included are: Coding, Symbol Search, and Cancellation.

Subtest Description. There are in total 15 subtests including 10 core and 5 supplemental subtests.

1. Block Design (BD): The examinee is asked to reproduce a set of modeled or printed two-dimensional geometric patterns using red and- white blocks within a given time limit.

2. Similarities (SI): The examinee has to describe in what way two words that represent common objects or concepts are similar.

3. Digit Span (DS): On Digit Span Forward, the child has to repeat numbers verbally as read by the examiner. Whereas, Digit Span Backward asks examinee to reiterate numbers in the reverse order.

4. Picture Concept (PC): The examinee has to pick one picture, from among two or three rows of presented pictures, to form a group with a common characteristic.

5. Coding (CD): The examinee is required to insert symbols that are paired with either geometric shapes or digits using a key within a given time limit.

6. Vocabulary (VC): The examinee is asked to verbally recognize pictures or provide meanings for the presented words.

7. Letter-Number Sequencing (LN): The child is presented with a number and letter sequence and is then asked to recall numbers in ascending order and letters in alphabetical order.

8. Matrix Reasoning (MR): The assessee is asked to visually complete the missing portion of a picture matrix by choosing one of five response options.

9. Comprehension (CO): The examinee has to answer a series of questions based on his or her understanding of general principles and social situations.

10. Symbol Search (SS): The examinee has to scan a search group and indicate the presence or absence of a target symbol(s) within the allotted time.

11. Picture Completion (PCm): The examinee is asked to identify and name the essential missing part of the presented picture within the allotted time.

12. Cancellation (CA): The child has to examine both a random and a nonrandom arrangement of pictures and cancel out target pictures within a specified time.

13. Information (IN): The examinee is requested to give answers to a wide range of general-knowledge based questions.

14. Arithmetic (AR): The assessee has to mentally solve several orally presented arithmetic problems within the given time period.

15. Word Reasoning (WR): The examinee is asked to identify a common concept being described by a series of clues (Flanagan & Kaufman, 2004).

Process Scores. The WISC-IV also includes seven process scores to give more detailed information on the cognitive abilities that add to a child's subtest performance. Derivation of these scores does not require additional administration procedures and is based on the child's performance on the corresponding subtest. Process scores include Block Design (with no time bonus), Digit Span Forwards, Digit Span Backwards, Longest Digit Span Forward, Longest Digit Span Backwards, Cancellation Random, and Cancellation Structured (Wechsler, 2004).

Differences from WISC–III. Many changes have been made in WISC-IV from its previous edition and they have increased the utility of the test in clinical, research and educational fields. The subtests such as Picture arrangement, object assembly and mazes have all been removed, while the subtests information, arithmetic, and picture completion, that were considered core subtests before are now supplementary subtests. Several new subtests are added to reflect current clinical knowledge and practice including Word Reasoning (measures reasoning with verbal material); Matrix Reasoning (measures fluid reasoning); Picture Concepts (measures fluid reasoning, perceptual organization, and categorization); Letter-Number Sequencing (measures working memory-adapted from WAIS-III); and Cancellation (measures processing speed using random and structured animal target forms).

Administration and Scoring. WISC-IV is an individually administered test. For most children test administration time is between 65 and 80 minutes, but more time is required if supplemental subtests are given or if the client is more intelligent, and less time is required if client is mentally retarded or deprived. It can be given to children as young as 6 years and as old as 16 years and 11 months. It is preferred to administer the entire battery in one session but if more than one session is required, second session should occur as soon as possible after the first testing, preferably within one week. Usually the Verbal Comprehension and Perceptual Reasoning subtests are administered in alternating order, with Working Memory and Processing Speed subtests interspersed. Core subtests are administered first, followed by the supplemental subtests if required. The *WISC-IV Administration and Scoring Manual* (2004) contains all the important information regarding administration in detail including information about subtests substitution rules; start point, reverse, and discontinue rules; and queries, prompts, and item repetition.

Scoring of majority WISC-IV subtests is an objective process but more judgment is required for the similarities, vocabulary, comprehension, information and word reasoning subtests. For items that are scored 1 or 0 points, 1 point is awarded for any response that is equivalent or superior to the 1-point sample responses, and 0 points are awarded for any response equivalent or inferior to the 0-point sample response. A similar principle governs the assignment of scores for multi-point items (Wechsler, 2004).

It has been observed that due to insufficient adherence to the scoring guidelines provided in the manual, scorers perform many mistakes. The most common mistakes examined in earlier researches are recording errors (not writing down responses or completion times on the record form), administration errors (not querying unclear answers on verbal subtests or incorrect assignment of points on various items) and computation errors (incorrect conversion of raw scores to standard scores). Verbal scale subtests, such as Comprehension, Vocabulary, and Similarities due to their judgmental nature are indicated to be the most vulnerable for examiner errors when considering administration and computation errors only. When recording errors are also included, nonverbal subtests, such as Picture Completion, Digit Span, and Arithmetic, have been the most error prone (Belk, LoBello, Ray, & Zachar as cited in Leo, Kadlubek, & Marks, 2007). Hopwood and Richard (2005) highlighted the importance of referring to the manual before scoring item responses, particularly on vocabulary, comprehension, and similarities subtests.

Norms and Standardization. The WISC-IV normative data was established using a sample collected from August 2001 to October 2002. Its normative sample has 2,200 children divided into 11 age groups (each group is of one year and each composed of 200 children). Sample also has an the same number of males and females in each group, and an ethnic divide that approximates the March 2000 US Census data on the variables of age, sex, race/ethnicity, parental education, and geographic regions. Sample includes five parental education levels, and four geographical areas covering all the United States and Hawaii. For arithmetic the normative sample was of 1,100 only (Wechsler, 2003).

Reliability. Its reliability was examined by estimating internal consistency values (split half correlations) and test-retest coefficients. The average internal consistency coefficient is .97 for FSIQ, and for indices coefficients are .94, .92, .92, and .88 for VCI, PRI, WMI, and PSI, respectively. For individual subtests, internal consistency coefficients across all ages ranged from .72 for Coding (for ages 6 and 7) to .94 for Vocabulary (for age 15). The median internal consistency coefficients for the individual subtests ranged from .79 (Symbol Search, Cancellation) to .90 (LNS).

Test-retest reliability was also established across the 11 age groups. Test was administered twice with an average interval of 32 days (13 to 63 days). The WISC-IV was found to be a stable with average test-retest coefficient of .93, .89, .89, .86, and .93 for the VCI, PRI, WMI, PSI, and FSIQ, respectively (*WISC-IV Technical and Interpretive Manual*; Wechsler, 2003). Two subtests showed improvement in reliability as compared to the older version of the test, may be because of the updates in the test. Inter-scorer reliability by experts was in general .98, with Comprehension dropping to .95. However, carelessness can effect this number drastically (Wechsler, 2003).

Validity. Validity for the WISC IV was assessed through utilizing various techniques and procedures. Content validity was established through reviewers and experts, and also by creating content similar to other established tests to expand the evaluation base of the WISC IV. The response process was also examined with multiple-choice formats to detect common errors, having children explain their responses to highlight alternate acceptable answers, and altering stimuli as a result. Inter-correlation studies were also done; this methodology provided the evidence of convergent and discriminant validity. Factor-analysis was also done to have evidence of structural validity. Exploratory Factor Analysis was done for that purpose. The principal axis factoring method was used for factor extraction, and oblique rotation was applied to allow for correlations between the factors. Cross-Validation Analysis was also done.

Along with factor analysis and content validity research, the validity of the WISC-IV is supported by correlations with other established assessment tools. The correlations between the WISC-IV FSIQ and the WISC-III FSIQ (.89) and the FSIQs of other Wechsler scales (i.e., WPPSI-III, WAIS-III, and WASI) are fairly high. The WISC-IV also displays moderate to high convergent / discriminant validity evidence. The VCI has an average correlation of .83 with other measures of verbal ability compared to a mean of .61 with measures of perceptual abilities. Similarly, the PRI has an average correlation of .76 with other measures of visual-perceptual ability compared to a mean of .61 with measures of verbal abilities. The validity of the WISC-IV was explored further through assessing its relationship with other measures like Wechsler Abbreviated Scale of Intelligence (WASI), Wechsler Individual Achievement Test (WIAT-II), Children memory Scale (CMS), and BarOn EQ. In addition, the WISC-IV Technical and Interpretive Manual describe various special-group studies to inspect the diagnostic usefulness of the instrument (Wechsler, 2003).

Keith (as cited in Flanagan & Kaufmann, 2004) suggested that a CHC theory (1998) based factor structure provide a good fit to the WISC-IV standardization data. According to Keith (2004), the WISC-IV assesses Crystallized Ability *(Gc),* Visual Processing *(Gv),* Fluid Reasoning *(Gf),* Short-Term Memory *(Gsm),* and Processing Speed *(Gs).* These findings are consistent with the results of a content validity study of the WISC-IV, based on CHC theory that used an expert consensus format (Caltabiano & Flanagan, as cited in Flanagan & Kaufmann, 2004). Though the Psychological Corporation documented four factors to explain the constructs underlying the WISC-IV, but Keith; and Caltabiano and Flanagan researched it to be five. The results of these latter two studies were quite consistent, with the exception of the CHC abilities supposed to underlie the Arithmetic subtest. Keith identified this test as measures of *Gf* and *Gsm,* whereas, Caltabiano and Flanagan classified this test as Quantitative Knowledge *Gq* and *Gf.* After the publication of the WISC-IV (Wechsler, 2003), the Psychological Corporation also classified all of the WISC-IV subtests on the basis of CHC theory on its web page, which indicated that the Keith and the Psychological Corporation classifications are largely similar, with only a few exceptions. The Psychological Corporation categorized 'Similarities' and 'Word Reasoning' as mainly measures of *Gf* and 'Arithmetic' as primarily a measure of *Gq* and *Gs;* while Keith categorized 'Similarities' and 'Word Reasoning' as measures of *Gc* and arithmetic as measure of *Gf* (as cited in Flanagan & Kaufman, 2004).

Research Studies on Wechsler Intelligence Scale for Children (WISC)

Wechsler Intelligence Scale for Children (WISC) with its various revisions is used throughout the world due to its strong psychometric strength and clinical utility.

Besides its clinical and psycho-educational use, it has also been considered as an important research tool and research object.

Research Studies on Normative Differences and Cross-cultural/group Comparisons of WISC. Various comparative studies based on cross-cultural adaptive or normative differences, and clinical cross-group differences have been conducted on different revisions of WISC.

Tempest (1998) has conducted a research to see the influence of local Navajo norms for the WISC-III. For that purpose he developed local norms for the test. Results indicated that urban Navajo students performed better than students in the country side on verbal IQ. Similarly, the students proficient in English performed better than students functional in English. But on performance subtests all students have shown equal performance.

Panicker, Hirisave, and Subbakrishna (2006) compared the scores of Indian primary school children on adapted WISC-III UK when Indian and UK norms were used. The sample included 300 children from the age group of 6 years to 10 years and 11 months studying in various primary schools. The children scores showed a decrease of 16, 21 and 20 points on Verbal, Performance and Full scales, respectively when UK norms were applied. The findings emphasized the need of developing national/regional norms for evaluation of Indian children's intelligence.

Khaleefa (2006) compared the adaptation process of the WISC-III in Sudan and Japan. He also compared the performance of Sudanese and Japanese children on WISC-III subtests. In Sudan, the WISC-III (6-16 years) was translated from English to Arabic and back translated from Arabic to English. The same approach for translation was used in Japan. The sample size for the two countries differed as the adapted test was applied to a group of 330 and 1125 children in Sudan and Japan, respectively. The results showed that the WISC-III has generally enjoyed adequate structural equivalence in Sudan as well as in Japan, and has high level of reliability and validity in the two countries. In Japan, the time limit for some subtests was shortened from 120 to 90 seconds. By contrast, in Sudan it was increased from 120 to 150 seconds. Additionally, the study showed that Japanese children performed better in visio-spatial tests while Sudanese performed better in verbal tests.

Grob et al. (2008) have explored the differences in Swiss and Germen children's intelligence as measured by the Germen language adaptation of the WISC-IV (HAWIK-IV; Petermann & Petermann, 2007). Normative sample for this adaptation is based on 1650 children from Germany, Austria, and Switzerland. Comparison was carried out at subtest, index, and FSIQ levels. Results indicated that for cross-sample mean differences two out of 15 tests, and for correlations 17 out of 55 tests showed significant cross national differences.

Allen, Thaler, Donohue, and Mayfield (2010) have gathered WISC-IV profiles of Children with Traumatic Brain Injury as it is often used to assess such injuries. Their purpose was to compare these profiles with profiles of WISC-III in terms of similarities and differences. WISC-IV profiles showed relative deficits on all subtests and index scores with greatest deficits on Processing Speed Index and Coding subtest.

Research Studies on Factor Analysis and Validation of WISC. Establishment or replication of factorial structure and construct validation of various versions of WISC seems another popular field of study among researchers.

Guilmette, Kennedy, and Queally (2001) compared Wechsler Intelligence Scale for Children-Third edition (WISC-III) and Otis-Lennon School Ability Test-Sixth edition (OLSAT-6) with students referred for learning disabilities. Both tests were found to be significantly correlated, although the WISC-III Full Scale IQs were significantly higher than the OLSAT-6 Total Scores.

Watkins, Wilson, Kotz, Carbone, and Babula (2006) applied factor analysis to WISC-IV scores of 432 Pennsylvania students referred for evaluation for special education services to determine and replicate the factor structure of the WISC-IV. A first order, four factor oblique solution that was found in the WISC-IV normative sample was supported. When transformed to an orthogonalized higher order model, the general factor accounted for the greatest amount of common (75.7%) and total (46.7%) variance. It was recommended that interpretation of WISC-IV should not mark down the strong general factor.

Bodin, Pardini, Burns, and Stevens (2009) conducted higher order factor structuring of WISC-IV in a clinical sample of 344 children having some neuropsychological problems. The study resulted in a four-factor solution and indicated that the first-order processing speed factor exhibited the most unique variance beyond the influence of *g*. The results suggested that clinicians should not ignore the contribution of *g* when interpreting the first-order factors.

Chen, Keith, Chen, and Chang (2009) had investigated the validity of WISC-IV four-factor structure and the CHC theory (1998) based model of WISC-IV. They conducted CFA on Taiwan's WISC-IV standardization sample. Both models, the four-factor structure model and CHC theory based model indicated good-fit to the data.

Montes, Puente, Allen, and Neblina (2010) have validated the WISC-IV Spanish for a clinically referred sample of Hispanic children. The sample comprised of 107 clinically referred brain dysfunctioning children. They have not done specific validation of WISC-IV for separate Learning Disabilities.

Watkins (2010) has analyzed the structure of WISC-IV (Wechsler, 2003) through confirmatory factor analysis among a national sample of 355 students referred for psycho-educational from 35 states. Analysis resulted in a general intelligence factor along with the four first order factors in a direct hierarchical model. The general factor was found to be pre-dominant source of variation among WISC-IV subtests, accounting for 48% of the total variance and 75% of the common variance. The researcher concluded that the recommendations for preferring interpretation of 1st-order factor scores over general intelligence scores appear to be misguided.

Watkins, Canivez, James, James, and Good (in press) have also researched the construct validity of WISC-IV UK core subtests on a large referred Irish sample. Though the oblique four-factor structure provided the best fit, but researchers could not found meaningful differences between four-factor model and its rival hierarchical models. Researchers have also discussed about the importance of focusing primary interpretation on the FSIQ.

Research Studies aiming enhanced Efficiency and Usage of WISC. Various revisions of WISC have also gone through vigorous research in order to enhance its efficiency and usage in clinical and educational settings. To increase efficiency its various shorter versions and indices comprising of lesser number of subtests have been researched. Similarly, ways of improving scoring practices have also been researched to facilitate its effective usage.

Beebe, McBurnett, and Pfiffner (2000) evaluated WISC-III Comprehension and Picture Completion subtests as measures of Social Intelligence. They compared these subtests scores with mother- and teacher-reported social functioning in 142 children with ADHD and 30 control children. Results showed that after the General Intelligence was partialled out, the comprehension subtest related to some aspects of social functioning, but the clinical significance was limited. Whereas, all other subtest proved totally unrelated to the social functioning.

Reiter (2004) has evaluated the application of a Dumont-Faro short form of WISC-III (Dumont & Faro, 1993) to screen gifted elementary school children in Canada. Results indicated high correlation with Full Scale IQ when it is administered separately. A reduction of 55% of total administration time was also observed. So researcher suggested this short form to be an effective screening tool for gifted children in Canada.

Scott, Austin, and Reid (2007) have compared WISC-III and WASI on 25 children of age 6 – 15 years from clinical child population. A correlational and within participant design was used. Results indicating significant correlations between both tests favored the use of WASI in place of full length test to have a quick and valid measure of intelligence under certain clinical circumstances.

Saklofske, Zhu, Coalson, Raiford, and Weiss (2010) have developed a Cognitive Proficiency Index- CPI for the Canadian edition of the WISC-IV. They have discussed the calculation and uses of CPI based on Canadian norms. The CPI comprises the working memory and processing speed subtests.

Considering the improvement in scoring practice, Leo et al. (2007) have conducted a study on administration and scoring errors on the WISC-IV among graduate student examiners. They indicated that students committed errors on 98% of the protocols examined. The most common errors were failure to query verbal responses, assigning too many points to an answer, and failure to record an examinee's response on the test protocol. Errors resulted in inaccurate test composite scores, with the Full Scale IQ and Verbal Comprehension Index most frequently affected. Error rates did not improve significantly over the course of three practice administrations.

Similarly, Erdodi, Richard, and Hopwood (2009) suggested the importance of relying on the manual scoring for controlling error variance in the WISC-IV Vocabulary Subtest. They revealed that graduate students were more prone to make scoring errors in the extremely low and superior ranges of the IQ distribution and the participants with more clinical experience made more errors. The study showed that reliance on the manual was the strongest protective factor against scoring errors in this sample.

Considering WISC's clinical utility, Mayes and Calhoun (2006) compared WISC's 3rd and 4th edition (WISC-III & WISC-IV) profiles for children with ADHD and normal intelligence. The index discrepancies were greater for the WISC-IV, suggesting that the WISC-IV might be better than the WISC-III in delineating the strengths and weaknesses of children with ADHD. All children in the WISC-IV sample scored lowest on WMI or PSI, whereas only 88% of the WISC-III children scored lowest on FDI or PSI. Thus, the WISC-IV may be more helpful in diagnosing ADHD than the WISC-III.

Zander and Dahlgren (2010) have gathered WISC-III Index score profiles of Swedish children with Pervasive Developmental Disorders. Significant differences were found between performance levels of children with Asperger's disorder, who performed in the average range according to Swedish norms, and children with either autistic disorder or pervasive developmental disorder not otherwise specified, who performed below the average range.

Beside these areas WISC has been researched or used in researches for various other purposes. For example, Canivez and Watkins (2004) evaluated the Temporal Stability of WISC-III subtests based cognitive strengths and weaknesses. This was evaluated with 579 students who were twice tested with the WISC-III. Based on 66 subtests composites, six or seven interpretable cognitive strength and weaknesses were found on each WISC-III administration. Results showed that subtests based cognitive strengths and weaknesses were unreliable.

Oakland and Harris (2009) stated that research on children's counterproductive test behavior supported a three-factor model for behaviors: inattentiveness, avoidance, and uncooperative mood. In their study, test behaviors measured by the Guide to the Assessment of Test Session Behaviors (GATSB) were rated. They used a sample of 110

Hispanic Spanish-speaking children included in the WISC-IV Spanish standardization to further understand the relationship of test behavior to test performance in Spanish-speaking children tested in their native language.

Brooks (2010) has studied the prevalence of low scores on WISC-IV by establishing the base rates of low scores for WISC-IV. Researcher observed that low score typically increased with lesser intelligence and fewer years of parental education.

Review of all this literature can easily lead to the conclusion that being a popular test WISC has been well researched. Some prominent trends of these researches can also be pointed out through the literature review. First, investigation or replication of factor structure of WISC with different samples or in different countries is the most frequent issue. Secondly, most of the researches have been done on western countries, especially researches on WISC in South Asian countries are quite rare. The same research trends prevail for other tests of Wechsler series.

Research Studies on other Wechsler Intelligence Scales

Due to their popularity and worldwide use other Wechsler Intelligence Scales also underwent many research studies throughout the world. Few such studies on different editions of Wechsler Intelligence Scales in different research areas are as below.

Factorial Validation of Wechsler Intelligence Scales. Literature again indicates that most researches were conducted for scale validations or to establish/replicate factorial structure of these scales. For example, Gorsuch, Hildebrand, and Saklofske (2000) conducted exploratory and confirmatory factor analysis of the subtests of WAIS-III on a stratified sample of 718 Canadian adults. Results of the factor analysis confirmed the presence of four WAIS-III factors: Verbal Comprehension, Perceptual Organization, Working Memory and Processing Speed.

Similarly Axelrod, Ryan, and Ward (2000) performed a confirmatory factor analysis with the standardization data of the WAIS-III to compare 6 models with 1 to 4 factors for 11- and 13-subtest versions of the test. Results showed that three factors usually fit the data better than 2 factors, but 2-factor models were more parsimonious.

Price and Tulsky (2003) have also presented a series of confirmatory factor analysis (CFA) to determine the joint Wechsler Adult Intelligence Scale, 3rd edition (WAIS-III) and Wechsler Memory Scale, 3rd edition (WMS-III) factor structure. Using a structural equation modeling approach, a six factor model that included verbal, perceptual, processing speed, working memory, auditory memory and visual memory constructs provided the best model fit for the data.

Whereas, McGrew, Taub, and Witta (2004) investigated the factorial invariance of the WAIS-III across the instrument's 13 age groups. The overall results from this study generally support both configural and factorial invariance of the WAIS-III when the 11 primary tests were administered.

Following the same trend, Egeland, Bosnes, and Johansen (2009) reported that CFA of the Norwegian version of WAIS-III showed partial support to the four-factor model. However, the Arithmetic subtest proved difficult to get allocated to one factor. Allowing Arithmetic to load on both the Verbal Comprehension and Working Memory factors provided a more parsimonious solution compared to considering the subtest only as a measure of Working Memory.

Canivez and Watkins (2010) have investigated the factor structure of WAIS-IV (Wechsler, 2008) standardization sample by using exploratory factor analysis, multiple factor extraction criteria, and higher order exploratory factor analysis. Results indicated that the second order 'g' factor accounted for large portions of total and common variance, whereas the four first-order factors accounted for small portion of total and common variance. So it was concluded that he WAIS-IV provides strong measurement of general intelligence, and clinical interpretation should be primarily based at that level.

Weiss, Keith, Zhu, and Chen (2013) have also clinically validated the four and five-factor interpretive approaches for WAIS-IV. They concluded that when all 15 subtests are used though both four and five-factor structure show adequate fit but five-factor approach showed better fit. This five-factor model has WAIS-IV PRI differentiated into two components POI (*Gv*) and FRI (*Gf*).

Considering the length of Wechsler scales, evaluation of clinical utility of short forms is also researched. For example, Axelrod and Donders (2002) evaluated the validity and reliability of various short forms of WAIS-III in a sample of 100 patients with traumatic brain injury and in a demographically matched subgroup from the standardization sample. The purpose was to evaluate the potential clinical utility of short forms of the WAIS-III that would maintain the four-factor structure, based on two subtests per factor index. Although the results were satisfactory for all possible combinations of VC subtests, none of the short forms for the PO index were sufficiently reliable for both samples. It was concluded that short-form estimates of WAIS-III are not appropriate for clinical use when the goal is to obtain factor indexes.

Different revisions of same intelligence scales have also been compared time to time in order to evaluate their strengths and weaknesses. In one such study Crawford, Allan, Besson, Cochrane, and Stewart (1990) had employed a matched sample design to compare WAIS and WAIS-R IQ in UK subjects. The WAIS yielded significantly higher mean Full scale, Verbal and Performance IQs. Mean WAIS IQ was 108.6, suggesting that the WAIS yields inflated IQ scores in the contemporary UK population. Mean WAIS-R Full scale IQ was 101.1, suggesting that it neither markedly underestimates nor over estimates IQ in the UK.

To enhance interpretive strength of the scale Saklofske, Tulsky, Weiss, and Wilkins (2001) have developed a General Ability Index (GAI) for the WAIS-III. It is based on the sum of scaled scores of subtests that make up the VCI and POI scores. The developers of these index scores have posited that the GAI is a better indicator of the general ability or g, than the FSIQ. Researchers also provided GAI normative tables for the WAIS-III standardization sample.

Similarly Longman and Saklofske (2007) presented tables for the WAIS-IIIIQ and index scores by education level for both the U.S. and Canadian normative samples. The purpose was to allow clinicians to provide more accurate identification of relative strengths or weaknesses, compared to expectations from an individual's background, rather than the general population. The similarities and differences between the two national samples were noted, with particular reference to the relatively weaker demographic effects found in the Canadian sample.

Intelligence Testing in Pakistan

In the developing countries, progress in education and mental health services has created a need for intelligence testing. For research and clinical purposes, IQ testing is gaining the same importance as it once did and to some extent still does in the west (Mahmood, 1991). In Pakistan many research studies have been conducted on development, adaptation and validation of ability, aptitude, achievement, personality tests, and other related issues. However, not much work has been done regarding intelligence testing. Though some attempts have been made to translate, adapt and validate well-known tests of intelligence like Binet, Raven Matrices, and Wechsler tests. Few studies have also been taken to develop indigenous verbal or non-verbal test of intelligence (see for example Gardezi, 1994).

In Pakistan, intelligence tests are being used for personnel selection since independence of the country. The major users of intelligence tests in the country are Armed Forces, Federal Public Service Commission and four Provincial Public Service Commissions. Intelligence tests have also been included in some Entry Tests being used for the selection of candidates for admission in Medical Colleges, Engineering Universities and other professional institutes throughout Pakistan. But these tests cannot be considered psychometrically strong as they are usually taken from the foreign sources and proper care has not been given towards their construction, adaptation or validation (Shah as cited in Gardezi, 1994, 2001).

Few efforts were made by the Board of Secondary Education Karachi, which in 1962 established a project to develop and standardize general ability and aptitude tests for educational guidance at the secondary school level. Senior teachers wrote items for the tests, which were standardized on a sample of 2000 children. However, the use of those tests was rather limited and was soon abandoned (Hassan as cited in Gardezi, 1994). So there have been no efforts at national level to develop tests and establish norms to ascertain the use of these tests across different cultural and regional groups throughout Pakistan. Though different institutions or universities, and organizations are using psychometric methods to develop or adapt tests according to their own requirements and available expertise. Some of the research studies conducted on the intelligence testing and related issues are briefly described below.

Jamal (1964) used Raven's Colour Progressive Matrices (CPM; Raven, 1963) with kindergarten children to explore the relationship between socio-economic level and intelligence. Researcher found non-significant differences in intellectual ability of children belonging to different socio-economic levels.

Ansari (1976) developed an abbreviated version of Wallach-Kogan Creativity test (1965), which was correlated with AH5 Verbal Part and Standard Progressive Matrices. It was found that while the intelligence is equally related to the achievement test of the lower and higher Cognitive Objectives, the creativity measures show significantly higher correlations with the achievement test of Higher Cognitive Objectives.

Riaz (1979) conducted a study to find out distinction between the constructs, intelligence and creativity, and their relationship with academic achievement. Multiple correlations of intelligence and creativity with academic achievement showed that addition of creativity tests to intelligence tests adds significantly to the prediction of achievement over and above obtained by intelligence test alone.

Hassan (1981) assessed the effects of bilingualism on the performance of Pakistani school girls on tests of verbal intelligence and reasoning. The results indicated that bilingualism was significantly related to poor performance in verbal intelligence and reasoning tests.

Sheikh (1982) developed *Zahanat Paimana* at the Department of Applied Psychology, University of the Punjab. It is an adaptation of Otis Quick Scoring Mental Ability test in Urdu. However, its use has been limited to the department concerned and has not been available to other researchers.

Zoofashan (1982) developed a computer model for statistical analysis of Progressive matrices. The objective of the project was to develop a computer model for psychologist with little or no knowledge of computer to check the reliability and validity of progressive matrices in a sample of Pakistani children of both sexes, belonging to urban and rural areas.

Ansari and Iftikhar (1984) conducted a study to determine the validity of Raven's Standard Progressive Matrices (RSPM) for urban and rural school children in Pakistan. It was found that RSPM is useful as a test of intellectual performance in the urban school children. For the rural school children the utility of the test was limited.

Ahmed and Khan (1984) developed a cultural adaptation of Columbia Mental Maturity Scale. It is meant for a lower age group (3 years 6 month to 9 years 11 months). It is a test of general reasoning containing geometrical and figural material, some of which are coloured. The major changes in the adaptation were replacing colours with line design for the ease of reproductions and changes in figural material to make them closer to the experiences of Pakistani children. Geometrical designs were retained as such.

Ain (1985) validated Cattell's Cultural-Fair Test (1949) on Pakistani children. The researcher administered the scale II of the test to 1129 students of 5^{th}, 6^{th}, and 7^{th} grades of English and Urdu medium schools in Peshawar. The validity study demonstrated that although test scores were not significantly correlated with age, they were strongly related to school grades, academic achievement and teacher's ratings. The author explained the low correlation of test scores with age in terms of non-availability of accurate record of the age of these students.

Israr (1985) tried out thirteen Piagetian tasks on a sample of 360 primary school children from all over Pakistan. The results showed that grade 1 children were at early concrete operational stage, grade 3 at mid concrete operational stage, and grade 5 were at late concrete operational stage.

Ismail and Mehmood (1986) administered Raven's SPM to 300 students to study the effect of sex and social class. A significant difference was found among the performances of three different classes. However, no significant difference was found between boys and girls.

Imam and Munaf (1988) also administered Raven's SPM on 66 students of 5^{th}grade having different birth order. A significant difference in intellectual performance was found in the first, second and third born children.

Israr (1988) studied psychological interpretation of mathematical learning problems among secondary school children from urban and rural background. The results revealed that there were more problems in learning mathematics at grades 6^{th} and 7th as compared to grade 8^{th}. The findings also revealed that the worded mathematical questions were more difficult to learn than the non-worded questions.

Mahmud (1990) developed and validated Educational Ability Test for Pakistani pre-school children. The test consisted of 56 items covering six areas: Visual matching, Reasoning, School Language, Quantitative concepts, Auditory Memory, and Rhyming. The test-retest reliability demonstrated temporal stability of the test over a period of two months. Similarly, reliability of the test computed by KR-20 formula (.90) showed the homogeneity of the test. Significant differences were found in the rural-urban samples, while the gender differences were indicated only in rural sample.

Abbas and Israr (1990) developed a Test of Intellectual Development for pre-school children. The contents of the test were both verbal and non-verbal. It consists of eight subtests: Colour Naming, Reasoning, Seriation, Verbal Memory, Pictorial Memory, Perceptual motor Task, One to One Correspondence, and Conversation. The reliability of the test was determined by test-retest and KR-20 methods. The test was validated against the criterion of age differentiation. The reliability and the validity studies confirmed the utility of the test as a sound psychometric tool.

Ansari, Tariq, and Iftikhar (1990) developed and validated the Educational Ability Test level 5. The test purports to evaluate the current status of a student in terms of a broad range of cognitive educational objectives including his/her ability to recall, comprehend, reason, and analyze materials that a student comes across in the environment in the school and outside the school. The internal consistency and test-retest reliability were found satisfactory, ranging from .87 to .90 for various groups.

Hussain (1993) developed a Group Verbal Intelligence Test in Urdu for high school students. The test comprised two subtests: Vocabulary Test and the Numerical Ability test. It was validated against the criterion of school marks. The reliability and validity were found to be satisfactory. Significant differences were found in male and female samples.

Syed (1993) developed a Non Verbal Test of Intelligence for Pakistani urban Primary School children. The test comprised of two subtests: Block Design and Picture Completion. The test reliability was determined by KR-20 method while validity evidence was obtained by correlating the test scores with school marks of the subjects. The reliability (r ranging from .82 to .86) and validity (r=.85) indices were found significant.

Naheed (1993) has developed a Verbal Test of Intelligence for Pakistani Urban Primary School children. The test is comprised of two subtests: Vocabulary and Arithmetic. The school marks were used as a validity criterion for the test while KR-20 method was used to determine reliability. Both the reliability (r= .80 to .85) and validity (r=.89 to .90) indices were found to be highly significant.

Gardezi (1994) developed a Non-Verbal Intelligence test for students of grade 10 in the age range of 15 to 17 years. The test comprised of four subtests: Series, Analogies, classification and Matrices. Both the reliability (r=.77 to .82) and validity (r=.76 to .82) indices were found to be highly significant. Percentile norms were developed separately for boys and girls.

Aziz (1997) developed a Pakistani Version of Columbia Mental Maturity Scale (CMMS) for children aged 3 years 6 months to 10 years. The test consists of 92 pictorial and figural classification items arranged in a series of seven overlapping levels: two for pre-school children, and five for each grade 1 through 5. The reliability and validity indices characterize the test as a useful tool for children in Pakistan, especially in school setting.

Jamil (1997) conducted a study on the development and calibration of a 24-itemed intelligence test for children of 8^{th} grade only. He implied discrimination index for item analysis. Similarly, Khan conducted a semi-standardization study of an intelligence test for students of grade six and seven (as cited in Hussain, Jamil, Siraj, & Maroof, 2012).

Ahmed (2000) investigated the emic perspective of intelligence in Punjabi culture and for that purpose developed an indigenous test on 'Emic perspective of Intelligence in Punjabi culture' for the Punjabi rural population. Researcher used the emic research strategy and concluded that emic perspective of intelligence in Punjabi culture is pluralistic and can be manifested in local perspective.

Hashmi (2000) has worked on standardization of an intelligence test for the middle level students. Researcher has developed five tests each for 6^{th} graders and 7^{th} graders. He used facility Index, D-index, and power of discrimination for initial item analysis. Later he implied Rasch model for item calibration through manual procedures by using PROX (Wright & Stone, 1979 as cited in Hashmi, 2000). The standardization of test included reliability and validity establishment and was done on a sample gather from Multan, D. G. Khan, Bahawalpur, and Sargodha divisions.

Gardezi (2001) conducted a study for development and standardization of an indigenous Non-Verbal Test of Intelligence for 12 grade students. For standardization of the test a sample of 200 participants was taken from urban and rural population.

Hussain (2001) conducted a study for development, validation and standardization of a Group Verbal Intelligence Test in Urdu for adolescents. Sample for the standardization of the test comprised of 535 students who have completed the 12 grade of education and were in age range of 17 to 20 years.

Kausar (2007) explored the gender differences and effect of different factors on intelligence. For this purpose she developed indigenous norms for WISC subtests of Information and Arithmetic for public and private school children of 6th and 7th class. Results showed that there was no significant difference in performance of boys and girls on subtest of Information but boys performed significantly better than girls on the subtest of Arithmetic. It was also found that many factors like parental education level effects intelligence.

Khan (2008) has developed a self-report measure of Emotional Intelligence (SRMEI) for healthy individuals and heart patients due to non-availability of an indigenous instrument to measure emotional intelligence specifically in Pakistan's context and for cardiac patients. The instrument emerged as having three sub-scales including emotional self-regulation scale, emotional self-awareness scale, and interpersonal skills scale. The convergent (with Urdu version of BarOn's EQ-I; Akram, 2004 & Urdu version of Mini-Marker; Manzoor, 2000) and discriminent validity (with Urdu version of BDI; Khan, 1996) of the self-report measure was also established.

Ambreen (2008) adapted Verbal Comprehension Index subtests of WISC-IV UK and also developed its norms for Pakistani children. Considering the inappropriate difficulty level or cultural irrelevance, five items in the vocabulary tests were replaced by new words while order of administration of two items was also changed. Similarly in the information subtest, four items were replaced by new items. Scaled score, composite score, percentile, and age-equivalent norms were developed.

Noor and Najma (2009) have translated Wechsler Memory Scale (WMS- III; Wechsler, 1997) in Urdu for their comparative study of memory deficits in younger and older adults. After translation the subtests were pre-tested and then scores on original and translated versions were compared and assessed for consistency.

Shamama-tus-Sabah, Gilani, and Iftikhar (2012) has investigated the psychometric strength of Raven Progressive Matrices (RPM) and concluded it to be highly reliable for use as a non-verbal intelligence test. They also explored gender and social class differences on RPM in middle childhood. Girls performed better than boys but researchers found non-significant differences on social class.

Hussain, Jamil, Siraj, and Maroof (2012) has also developed an intelligence test for school going children of age 6- 11 years. They have analyzed the items on the basis of item difficulty and discrimination on a sample of 600 children of Dera Ismail Khan only.

All these studies show realization of Pakistani researchers about the importance of intelligence testing especially for children and adolescents. But a review of these researches also point out serious limitations in their methodology when compared to international standards (see for example Guidelines for Test adaptation and Development, International Test Commission, 2010, 2012). Most of these researches aimed for development or standardization of various intelligence tests but with much restrictive methodology. Considering test development, most tests are developed for a very specific age group and even for these tests mostly classical item analysis techniques were used for item finalization with small samples. Similarly, test standardization is mostly limited to establishment of reliability and validity evidence, whereas norms development is not given due attention. Furthermore, the standardization samples are non-representative and mostly selected on convenience forcing serious limitation to test use and interpretation. Overall, though intelligence testing is common in Pakistan, but adherence to the international test usage and development guidelines, and use of latest psychometric techniques seems deficient.

Cross-cultural Testing and Test Bias

Cross-cultural testing has pin pointed the ways in which culture and testing can interact. Usually the term culture or cultural differences is used to delineate people living in different countries or to refer minority/various groups within a country as they may differ in their beliefs, customs, morals, practices, and values (Domino & Domino, 2006).

The testing of persons with highly dissimilar cultural backgrounds has received increasing attention since 1950s (Anastasi & Urbina, 1997). The issue of test or measurement bias is a central one for all those who are concerned with developing and using tests, especially for those who are developing tests that are intended to be used cross-culturally, and for those who are using tests that have been constructed in a contrasting culture. For example, one can suspect for measurement bias when using a test developed in a westernized culture for a non-westernized culture. Most criticism regarding test bias has been directed towards cognitive-ability tests (Domino & Domino, 2006), especially for tests of intelligence and, to a lesser extent, to tests of aptitude, ability, and achievement. These biases are due to the variables along which cultures vary.

Controlling Bias in Cross-cultural Testing. In cross-cultural studies, bias is a broad expression used for all factors that intimidate the validity of intergroup comparisons (Van de Vijver & Hambleton as cited in Malda et al., 2008). Bias results from a test's cultural loading, which refers to the extent to which the test implicitly or overtly relates to the particular cultural conditions. The three major types of bias include: construct bias, method bias, and item bias. An test that displays 'construct bias' in a cross-cultural comparison do not measure the same psychological construct across cultures. Method bias means those sources of bias that come up from methodological aspects of a research, such as instrument bias and administration bias. Item bias (differential item functioning) refers to item-specific problems in cross-cultural comparisons, such as item ambiguity due to poor item translations or culture-specific elements (e.g., an item about a microwave oven is biased against cultures in which this appliance is not common). Two types of procedures were suggested to control such bias: a priori procedures (also called judgmental procedures) and a posteriori procedures (statistical procedures) (as cited in Malda et al., 2008).

Reynolds (1982) stated six possible forms of test bias in cross-cultural testing. These include: inappropriate test content; inappropriate standardization samples;

examiner and language bias; inequitable social consequences; cultural difference in meaning of construct; and test having differential predictive validity (as cited in Domino & Domino, 2006).

Cross-cultural tests have tried to rule out one or more of the factors leading towards bias. Considering language, if the cultural groups to be tested spoke different languages, tests were constructed that do not demand language on the part of either examiner or examinee. When educational backgrounds differed widely, or illiteracy was prevalent, 'reading' was minimized. Nowadays many non-verbal assessment tools are available to rule out such parameters, like the new Wechsler Non-verbal Scale of Ability WNV, developed by Naglieri in 2006 (as cited in Weiss, Prifitera, Saklofske, & Holdnack, 2006). Another parameter on which cultures/sub-cultures differ is that of 'speed'. Accordingly, cross-cultural tests have often tried to eliminate the influence of speed by allowing long time limits and giving no bonuses for faster performance. Still another factor along which cultures differ is 'test content'. To minimize the influences of various cultural factors, several noteworthy but largely unsuccessful attempts were made to develop "culture free" or "culture fair" tests, that may only include items related to experiences common to a wide range of cultures and eliminating certain parameters, such as reading and response speed, etc. (Anastasi & Urbina, 1997). This may reduce test bias but can not eliminate it fully.

Approaches to Cross-cultural Testing. Three different approaches have been followed in construction of cross-cultural tests (Anastasi, 1982). The first approach involves the selection of items common to many cultures and validation of the resulting test against local criteria in many different cultures. But it is unlikely that any single test could be designed that would fully meet these requirements across a wide range of cultures. Moreover, there periodic validation is required which is often neglected. The second approach is to develop a test within one culture and administer it to persons with different cultural backgrounds. Such an approach provides cultural distance between groups, as well as the individual's degree of acculturation and his/her readiness for educational and vocational activities that are culture specific. As a third approach, different tests (or substantial adaptation of the existing tests) may be developed within each culture, validated against the local criteria, and used only within that culture. Each test is applied only within the culture in which it was developed and no cross-cultural comparison is attempted.

With the continuous development in cross-cultural testing it was noticed that no single test could be universally applicable or fair to all cultures. Cultural differentials in test performance can be reduced, but a cross-cultural test can not completely eliminate them. The mere use of paper and pencil or the presentation of abstract tasks having no immediate practical significance will favor persons of some cultural and handicap others. Emotional and motivational factors also influence test performance such as the intrinsic interest of the test content, rapport with the examiner, drive to perform well on a test, desire to excel others, and past habits of solving problems individually and cooperatively (Anastasi & Urbina, 1997).

Furthermore, in developing countries many children face multiple risks in their environment and display suboptimal developmental capabilities (physical, cognitive, and socio-emotional). This developmental lacking results from poor nutrition, housing, and hygiene; low socio-economic status; and restricted learning opportunities. Cognitive tests having Western origin may be are not effective to assess these children as these tests cannot be considered suitable cross-culturally. Additionally, contrary to the test-wise children of West, children in non-westernized countries may be are unfamiliar with testing procedures and materials. For example, working with figures and matrices may be a completely new experience for children of developing countries, whereas Western children are familiar with these tasks from preschool level (as cited in Malda et al., 2008).

The problems in cross-cultural testing lead to the conclusion that the third approach of cross-cultural testing, that is adequate cultural adaptation of the test is the only right approach. Even after adequate adaptations, cross-cultural comparisons between different cultures can be miss-leading and so should be avoided.

Test Translation and Adaptation

Test adaptation may be defined as the literal translation of some stimuli and to change of others so as to minimize bias and maximize their cultural appropriateness in the target culture (Van de Vijver & Poortinga as cited in Hassan, 2006). *Adaptation* is considered as a broad term for any procedure involved in transformation of an instrument developed in one culture for usage in another culture. The term has taken the place of the conventional concept of 'translation', because of the growing understanding that transferring a test to a new cultural and linguistic context involves more than merely translating an instrument.

The term *adaptation* is also used in specific sense. Three terms have been discussed to explain the transformations that are required to transfer a test to another culture: 'adoption' (or application), 'adaptation', and 'assembly'. Adoption of an instrument means a close translation into the target language, and it can be used if the purpose of a study is to compare scores across cultures directly. Assembly means the development of an completely new instrument, and it is usually done when the translation of an existing instrument would yield an invalid measure in the target culture or when a new research topic has to be explored for which no suitable instrument is available yet. Adaptation has qualities of both adoption and assembly; it includes a combination of close translation of the parts of the instrument that seem appropriate in the target culture, such as test instructions and items, and a change of other parts when a close translation would be inappropriate for linguistic, cultural, or psychometric reasons.

The two different usages of the term *adaptation* (broad and specific) are closely associated if we do not take adoption, adaptation, and assembly as three totally different kinds of procedures, but as terms on a continuum that ranges from a close translation of all instrument features (adoption) to a complete change of these features (assembly). Adaptation can then be considered as a term for all transfers that do not exist on the extremes of the continuum. In this explanation, adaptation covers a wide range of changes to tests (which may explain the popularity of adaptation in the current literature) and is the main method of transfer in current qualitative evaluation of test appropriateness (Malda et al., 2008).

Adapting or translating achievement, ability, and personality tests constructed in one language and culture into other languages and cultures has a long history in educational and psychological testing. Some of the reasons found in the literature for adapting tests are as following (Hambleton & Patsula, 1999):

1. Adapting a test is usually significantly cheaper and quickly than developing a new test in a second language.

2. An adapted test is the most appropriate way to create an equivalent test in a second language if the purpose of testing is cross-cultural or cross-national assessment (such as with many credentialing exams),.

3. This can be helpful as there may be an unavailability of expertise for constructing a new test in a second language.

4. More assurance and reliability is connected with an adapted test instead of a newly developed test especially when the original test is well known.

5. Fairness towards examinees often results from the existence of multiple language versions of a test.

Besides these, another major reason for considering test adaptation or translation is that it facilitates comparative studies across cultures and language groups (Yopp & Brown, 1999).

Types of Cognitive Test Adaptations. Different types of adaptations are appropriate for different types of tests. Five types of cognitive test adaptation can be identified (Malda et al., 2008).

Construct-driven adaptations. These are related to differences in definitions/explanations of psychological concepts across cultures. For example, when the goal is to measure "intelligence", the test should be adapted according to the target culture's explanations of intelligence.

Language-driven adaptations. These are applied when there is unavailability of semantically equivalent words across languages (e.g., there is no Dutch equivalent for the English word "distress") or there are structural differences between languages (e.g., words or grammatical structures automatically refer to gender in some languages, which makes it difficult to avoid gender-specific references. For example, the English word "friend" can indicate both a male and a female person, whereas the German word "Freund" refers to a male friend and "Freundin" to a female friend).

Culture-driven adaptations. These adaptations result from different cultural norms, values, communication styles, customs, or practices. For example, an item about celebrating marriages should consider that cultures differ noticeably in practices and cultural relevance of marriage.

Theory-driven adaptations. They involve changes that are required because of theoretical reasons. For example, digit span items should ideally have digit names that are all of similar length. Similarity in digit length may be lost when the items are translated into another language.

Familiarity/Recognizability-driven adaptations. These are based on differential familiarity with task or item characteristics (e.g., a classical drawing of a house in one culture is not necessarily recognized as such in another culture) or stimulus materials (e.g., in some cultures children might not be used to work with matrices).

Considering the language and content of WISC-IV the culture-driven and familiarity/recognizability-driven adaptations seem more probable in Pakistan. Language-driven adaptation could also be required during the process of test translation.

Misconceptions about Test Adaptation. There are a number of misconceptions linked with adapting tests that appear in psychometric practice and should be corrected as quickly as possible (Hambleton & Patsula, 1999):

1. It is mistakenly assumed that preferable strategy is always to adapt an existing test rather than develop a new test for a second language group. There are many good reasons for adapting a test, but there are reasons for not proceeding with a test adaptation as well. Especially when cross-cultural comparisons are not of interest, it may be more appropriate to construct a new test for a second language group. This avoids any complications with copyright, insures that the format will be suitable, and any desired modifications in the definition of the construct of interest can be made at the outset of the test development process. So keeping in view the purpose of the study, a researcher should weigh and then decide about going for an adaptation or not. This decision should be taken before delineating the specific objectives of the study.

2. Quite prevalent is the view that anyone who knows the two languages can produce an acceptable translation of the test. This may result in unqualified persons adapting tests. There is considerable evidence suggesting that test translators need to be familiar with both source and target languages and the cultures, and they need to be generally familiar with the construct being assessed, and the principles of good test development practices. So, learned and trained translators should be involved in the translation process.

3. It is also a prevailing view that a well-translated test guarantees that the test scores will be valid in a second language or culture for cross-language comparative

purposes. Van de Vijver and Poortinga (as cited in Hassan, 2006) stated that not only should the meaning of a test be consistent across persons within a language group and culture, but that meaning, must be consistent across language groups and cultures as well. Even if the test is equally valid in each language group and culture, it may still not be suitable for making cross-cultural comparisons. Similarly, the unequal familiarity of students of different cultures with certain item formats, like, the multiple-choice format, places examinees from the second cultural group at a serious disadvantage. The translation could be excellent, but the scores from the two language versions are still not equally valid. So instead of relying on mere translation, appropriate adaptation requirements (in construct meaning, theory, or task familiarity) should be taken in consideration.

4. It is also often misconceptulized that the constructs are universal and therefore all tests can be translated into other languages and cultures. For example, in case of intelligence tests, the Western concept of intelligence places considerable emphasis on speed of response. In some cultures, speed of response is of minor importance, so members of these cultural groups often score lower on Westernized intelligence tests because of not performing quickly. But if intelligence is defined in a way that devalues speed of response and emphasizes other human attributes of intelligence, then the results would be opposite (Hambleton & Patsula, 1999). So care should be taken in deciding whether or not the construct or the theoretical concept is equally applicable in both the original and new cultures (Yopp & Brown, 1999).

5. It is also wrongly practiced that translators are capable of finding flaws in a test adaptation and field testing is not usually necessary. The fact is that translators are

not able to anticipate all of the problems encountered by examinees taking a test in a second language. Field testing assess item functioning of the translated versions in the actual testing situation. Apparently well translated items can lead to invalid results if not field tested properly.

In summary, all of these misconceptions can seriously compromise the validity of a test in a second language or culture, or negatively influence the validity of adapted tests for use in cross-language comparison studies. Fortunately, each misconception is quite simple to address in practice (Hambleton & Patsula, 1999). By following the appropriate steps and test adaptation guidelines we can practically avoid these misconceptions.

Steps for Adapting Tests. Although interest in the field of test adaptation has existed for quite some time, the methods and guidelines for how to adapt a test have not been well established or understood (Yopp & Brown, 1999). In an effort to improve this situation the International Test Commission (ITC), an international committee of cross-cultural and educational psychologists, has developed a comprehensive set of practical guidelines. These guidelines are organized into four sections that include context, test development and adaptation, administration, and documentation/score interpretation.

According to ITC Test Adaptation Guidelines, the amount of overlap in the construct to be measured by the test in the intended populations should be assessed and effects of cultural differences that are irrelevant or unimportant to the main objective of the study should be kept as minimum as possible. Test developers/publishers should make sure that the adaptation process has fully considered the differences in language and culture among the intended populations. They should also provide evidence that the language used in the test; testing techniques and item format; and the item content and stimulus material is familiar and appropriate for all intended populations. Moreover, test developers/publishers should utilize systematic linguistic and psychological judgmental evidence along with statistical evidence of item/test equivalence in order to have valid adaptation versions. Considering administration, test administration should be aided with clear instructions in both source and target languages, and test manual should clearly identify all aspects of administration that require care in a new cultural context. Furthermore, when a test is adapted for another population, documentation of the changes should be provided, along with evidence of the equivalence. The test developer should specify and inform about the ways in which the socio-cultural contexts of the populations might influence test performance, and should suggest procedures to account for these influences in the interpretation of results (International Test Commission, 2010, 2012).

To establish proper test adaptation from one culture or language to another few considerations should be kept in mind. Firstly, in order to have meaningful cross-cultural comparisons it is important to assess whether construct equivalence exists between the cultures of interest and if it does not, then either opt for project discontinuation or consider revising the definition (decentering) of the construct to be equally equivalent in each language and cultural group (Hambleton & Patsula, 1999).

It is also important to consider the purpose of the adapted test, and the advantages and disadvantages of opting for adapting an existing test rather than developing a new test. With tests intended for cross-cultural comparisons, test adaptation (possibly with some decentering) may be the only option. One method of increasing the likelihood of a valid test adaptation is to take up one of the two (or both) standard designs: forward- and back-translation. Forward translation designs are more technically sound as their focus of the review is on both the source and target language versions of the test. Moreover, reviewing the adapted version of the test by judges and making necessary revisions is also important. In addition, it is also important to focus on the quality of the translators. Selection of translators is often one of the major limitations of a test adaptation project. Hambleton and Kanjee (1995) stated that translators should be fully proficient in both languages of interest, be familiar with the cultures associated with the different language groups, and have an understanding of the subject domain measured. Hambleton, Sireci, and Robin (1999) further suggested that translators should also be proficient with respect to principles of good item writing (as cited in Sireci, Yang, Harter, & Ehrlich, 2006).

The judgmental evidence is insufficient to establish the validity of a test in a new language. So to initiate with a small tryout of the adapted test is sensible before investing considerable resources in a more ambitious field test. Pilot study tests the instrument using a small sample of individuals' representative of the eventual target population and compares the results to results obtained from a source sample. The pilot test should include the test administration and the individual's feedback regarding the test itself, instructions, time limits, etc. In field test, the researcher should use the adapted test on a larger sample of individuals representative of the eventual target population and conduct preliminary statistical analysis like reliability analysis and a classical item analysis. In addition, check for construct equivalence using factor analysis should also be carried out. Moreover, Ellis and Mead suggested that when translation teams can not reach consensus regarding the most appropriate translation of an item, items in both languages should be field tested (as cited in Sireci et al., 2006).

If cross-cultural comparisons are of interest, a statistical linking design is needed to place the test scores from the different versions of the test on a common scale. Similarly, for cross-cultural comparison purposes, equivalence of different language versions of the test also needs to be ensured. Regardless of the interest in cross-cultural comparisons of scores from the two language versions of the test, and the related research generated by that concern, there is a need to ensure that the test scores of the newly adapted test are valid and reliable (Hambleton & Patsula, 1999).

Sireci et al. (2006) used differential item functioning methodology to evaluate the comparability of translated items at two different points in time - immediately after the initial translation and 4 years later on revisiting of translation using a more rigorous translation model. The results indicated that the revised translations led to improvements in some but not all items. So improvements in the process of translating survey items, even when based on accepted professional standards, should be statistically evaluated.

It is also important to document all results obtained and prepare a manual for the users of the adapted test. The manual should include specifics regarding the administration of the test, as well as how to interpret the test scores. Training the users of the test is also important. Although documentation and a manual will assist users of the adapted instrument, training will further assist them. Researchers should also remain vigilant to potential flaws in their adapted tests, and this means that ongoing monitoring of adapted tests is needed. Re-investigation and reevaluation of the reliability and validity of test scores should be ongoing.

For the current study all the steps required in test adaptations will be taken in consideration. The purpose of providing a well standardized clinical and educational research tool for children will be weighed against the intensity of differences in language and culture between the two countries (the country from which test originated and Pakistan). Decision of going for test translation or adaptation will also take account of empirical evidence gathered from procedures like pre-testing of original test. If required, appropriate translation design and translation team will be thought of. Judgmental validity of the adapted version will be confirmed through field testing; and statistical equivalence and psychometrics for the adapted version will be established and documented.

Beside all the technicalities, test adaptation is a much common practice among psychologists and researchers of the developed countries. This practice results from the awareness of researchers about the role of test adaptations to avoid duplication of efforts and to facilitate cross-cultural studies. Some examples of the studies that involve test adaptation are briefly described below.

Keung and Leung (1990) conducted a study for the adaptation of the Family Environment Scale (FES; Moos & Moos, 1986) for the Chinese children and adolescents in Hong Kong. The FES scale intercorrelations were consistent, and the factor pattern of the used eight scales was explicable in terms of the characteristics of the Chinese culture. Results also showed that middle class families in general had a more positive family social environment than the working class families.

Glaub and Kamphaus (1991) conducted a study to construct a non-verbal adaptation (short form) of the Stanford-Binet Fourth Edition to use in the evaluation of hearing impaired, speech/language disabled, and limited English proficient children. The reliability coefficient for the new composite scores was .95, and the validity of this composite, as estimated by its correlation with the test composite from the full battery, was .91.

Scott, Lucio, and Reyes-Lagunes (1994) developed a Spanish (Mexican) translated version of Minnesota Multiphasic Personality Inventory-2 (1989) by utilizing the concept of transliteration rather than literal translation of individual scale items. Sample data were collected from 929 male and 1245 female Mexican college students. The instrument was evaluated by comparing Mexican student's profiles to MMPI-2 college student values from the United States. Results indicated that besides small differences on scale L, the two groups were remarkably similar. It was also concluded that the instrument was appropriate for use in the college population.

Growing awareness for the importance of cross-cultural studies; and lack of resources and expertise for test construction have motivated researchers of developing countries to undertake test adaptations. In one such study Kamat (1935) made an attempt to adapt the Stanford-Binet Scale in Marathi and Kannada languages in India (as cited in Gardezi, 1994).

Chowdhury (1989) developed a Bangali adaptation of Spielberg's State-Trait Anxiety Inventory, Form X. Cross-language equivalence testing for both trait anxiety-TA and state anxiety-SA was done on 443 subjects. A normative score pattern of TA and SA of 1912 subjects along with pathological scores of 1805 neurosis patients was also provided.

Akkok and Askar (1989) conducted a study for the adaptation and standardization of the teacher version of the Child Behavior Profile for boys aged 6- 11 to Turkish boys aged 7- 12 years. Data were obtained from 48 referred and 294 non-referred boys. Comparison of referred and non-referred samples showed significant differences on all behavior problem scores, except the 'anxious scale'. The median of the internal consistency of the scale scores was 0.75.

Kaushik and Sheikh (1990) developed a Hindi adaptation of the Social Avoidance and Distress Scale (SAD) developed by Watson and Friend (1969). The method of 'back translation' was used in adaptation procedures. Very high correlations between Hindi and English forms of the test were found. All the items of the Hindi SAD Scale were found to be homogeneous and valid. The test-retest reliability was fairly high and the adapted scale was validated by correlating the scores on SAD Scale with the scores of the subjects on State-Trait Anxiety Inventory, Self-Concept Inventory and the Sociability factor of the Self-Concept Scale.

Malda et al. (2008) have adapted the Kaufman Assessment Battery for Children, second edition (KABC-II) for 6 to 10 year old Kannada-speaking children of low socioeconomic status in Bangalore, India. They followed a judgmental approach of translating, piloting and modifying the subtests that resulted in adaptation of test instructions, item content of both verbal and non-verbal tests, and item order of few tests. Through this adaptation Malda et al. has pointed out the process and problems of adapting a westernized test for a non-westernized country such as controlling cultural and familiarity related bias.

Adaptation of Wechsler Intelligence Scales. All tests included in Wechsler series of intelligence scales are widely used throughout the world due to their psychometric strength and comprehensive interpretative system. So Wechsler scales have been adapted in many countries and cultures. Considering WISC, its third edition (WISC-III) has been adapted for more than 16 countries and cultures including United Kingdom, Canada, France, Netherlands, Germany, Australia, Switzerland, Sweden, Lithunia, Solvenia, Greece, Japan, South Korea and Taiwan (Panicker et al., 2006). Similarly WISC-IV also has many adapted versions including UK, Australian, Spanish, French and Canadian (Weiss as cited in Weiss, Prifitera, Saklofske, & Holdnack, 2006).A Germen language adaptation of WISC-IV (HAWIK-IV; Wechsler, 2007) and Indian adaptation of WISC-IV (WISC-IV $^{\text{INDIA}}$; Wechsler, 2012) has also been published.

Studies Based on Adaptation of WISC-III and WISC-IV. Some of the studies that carried out adaptation of 3rd and 4th editions of WISC-IV are briefly described below.

Sans (1984) did his study on hundred pupils belonging to different socio-cultural levels whose ages ranged from twelve to thirteen. These pupils have been given a translated and adapted version of WISC to analyze the difficulty and relevance of the items included in the scale. They observed that the difficulty criterion was not fully met. For example in Information and Vocabulary subtests, few items were shown to be more difficult, while some other turned out to be too easy for the order they have been placed in. Besides, there are few words too sensitive to the subject's cultural differences while some items turned out too difficult due to the fact that the required information is less frequent in the environment.

Ueno and Nakatani worked in the project aimed at standardization of the third edition WISC (WISC-III; Wechsler, 1991) in Japan by the Institute of Psychological Aptitude in 1998. Vocabulary subtest has the highest level of items changed which in Japan was 93% of the whole original items. For the Information subtest, Japanese psychologists replaced almost half of the American items by local historical persons, famous indigenous individuals and local geographical places. The time limit for some timed subtests and some items was shortened (e.g., Coding and Symbol Search) which was shortened from 120 seconds to 90 seconds. In the adaptation process, Japanese psychologists were highly sensitive to their environment with respect to the issue of disabled people. For this reason, items related to missing organs in the human body or parts of animals in the Picture Completion test were replaced with other things (as cited in Khaleefa, 2006).

Gregoire (2001) found out factor structure of French adaptation of the WISC-III as the standardization of the French version was conducted on 1,120 participants with an age range of 6 to 16 years. For assessing 4-factor solution found on U.S. version, the same exploratory and confirmatory factor analysis was performed. It was found that 4-factor solution did not fit the data and was very unstable across the age ranges. The 3-factor solution seemed a better one as having Verbal Comprehension, Perceptual Organization and Processing Speed factors.

Panicker and Hirisave (2005) have developed an Indian adaptation of WISC-III UK along with the Indian norms for the test. This adaptation has a test-retest reliability ranging from .55 to .90 and split-half reliability ranging from .75 to .95. The criterion validity varies from .25 to .41. Whereas, the discriminant validity was also established (as cited in Panicker et al., 2006).

Hussain (2005) adapted the Third Edition of the Wechsler Intelligence Scale (WISC-III; Wechsler, 1991) in Sudan. Most changes were done in the subtests of Vocabulary and Information. About 65% of the original items were replaced by new items in the Vocabulary Subtest, while half of the original items of information subtests were replaced by local historical persons, famous indigenous individuals and local geographical places. Few changes were made in performance tests like Picture completion as well. The Sudanese children performed less in speed subtests. For that reason a suggestion was made for increasing the time for speed subtests by 30 seconds (as cited in Khaleefa, 2006).

An Australian adaptation of WISC-IV came as WISC-IV Australian (2003). Norms were also updated to match the current Australian data along with replacement of many outdated items (WISC-IV Australian, 2003). Similarly, Wechsler Intelligence Scale for Children, Fourth UK Edition (WISC-IV UK) was published in 2004 with the new UK norms and standardization sample (Wechsler, 2004). WISC-IV Canadian (2004) was normed and standardized using a Canadian sample of 1,1000 children (50% male & 50% female) ages 6 to 16 years 11 months.

WISC-IV Spanish (2005) is a comprehensive adaptation of the original English edition and incorporates language that represents the U.S. Spanish-speaking population from many countries of origin. In some cases, an existing WISC-IV item was translated directly into Spanish. In other cases, where no direct translation was possible (where an English word has multiple meanings in Spanish) the original item was replaced with a completely new Spanish item. WISC-IV Spanish comprises the same 10 core subtests as the original WISC-IV, but has four of the five supplemental subtests, as Word Reasoning has been excluded. Subtests are grouped into the same four indices as in original edition.

Petermann and Petermann (2007) have handled the German language adaptation of WISC-IV as HAWIK-IV (Wechsler, 2007). The German norms for the said adaptation were also developed and the standardization sample was comprised of 1650 children from Germany, Austria, and Switzerland.

Dang, Weiss, Pollack, and Nguyen (2011) have worked on adaptation of WISC-IV for Vietnam. Being a non-western non-English speaking country, the study was conducted following the various cultural adaptation and the standardization steps.

WISC-IV $^{\text{INDIA}}$ (Wechsler, 2012) is the recent adaptation of WISC-IV as published by the NCS Pearson, India. This version not only has culturally appropriate subtest content for assessing Indian children but has also provided Indian scaled score, composite score and Full scale IQ based norms.

Many other projects are in progress to develop language or cultural adaptations of WISC-IV. The current study is also an effort to adapt and standardize WISC-IV for Pakistani children and is carried out in collaboration with NCS Pearson, India Pvt. Ltd.

Translation and Adaptation of Tests in Pakistani Culture

In developing countries, need for tests are multi-fold. Tests are needed for the maximum utilization of human resources. Whereas, the rapidly spreading educational facilities in these countries require testing for admission purpose as well as for individual counseling. Similarly, tests are required for job selections in different professional fields (Anastasi, 1990). As in any other developing countries, tests or assessment tools are much needed in Pakistan for clinical, research, educational as well as for employee selection purposes. Due to lack of resources for development of new tests, efforts have been made for the adaptation of many foreign/westernized tests to meet the testing needs in Pakistan.

Adaptation of Non-Cognitive Tests. Measurement of certain constructs for research purposes seems to be the leading cause for adaptation of non-cognitive tests. Whereas, clinical, personality, and psycho-educational assessment has also lead to the adaptation of few tests in local cultural or linguistic context. Some of the studies conducted in Pakistan involving adaptation of non-cognitive assessment tools are briefly described below.

Raza and Sheikh (1991) conducted a study to translate and adapt Children's Personality Questionnaire (CPQ) of the IPAT series into Urdu. Local norms based on a sample comprising 292 boys and 281 girls, across 14 bipolar dimensions of personality were also developed in their study.

Anila, Khan, and Pervaiz (1991) adapted Caldwell and Bradley's Home Observation for Measurement of the Environment (HOME) Inventory (infant version) for use in Pakistan. The inventory appeared to be internally consistent (KR-20 = .82). Significant differences were found on the total and subscales of HOME Inventory scores for SES and father's occupation, and some of the subscales of father's and mother's education, mother's occupation, birth order of the child, and family type. No significant difference was found in gender, age, number of male and female siblings, language used in the family, current child caregiver, and family size.

Rahman and Saleem (1992) indigenized Jones's Irrational Belief Test (IBT; 1968) in order to use it effectively with Pakistani population to assess irrational beliefs in a valid and reliable way for Rational Emotive Therapy (RET) practitioners and researchers. Local norms were also developed for Psychiatric, drug addict, and normal Pakistani population. Results showed that IBT (Urdu Version) was a reliable and valid instrument.

Khan (1992) conducted study to translate and adapt Gordon Occupational Checklist (GOCL; Gordon, 1967) an interest inventory in Urdu for the matric students of Pakistan. It was named *App Ki Dilchaspian* (AKD). The inventory was administered to 300 students (150 boys, 150 girls) of class VIII. The results indicated that AKD is a reliable instrument to measure vocational interest of the Pakistani students.

Manzoor (2000) has adapted the Mini Marker Set (MMS; Saucier, 1994) in his research in order to assess the personality characteristics along the Big Five dimensions. The reliability and empirical evidence for the adapted version was ascertained with an independent sample of 195 college students.

Chishti (2002) designed a research to adapt, translate, and validate the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) for Pakistani culture.

The first study consisted of adaptation, translation and cross language validation. The reliability and the empirical equivalence was established. Data analysis suggested that the adapted version of the NEO-PI-R operates in the same way as the original one and measures the five factors with an equally comparable effectiveness. In study II, the discriminant and the convergent validities were established. The study also probed into the mean difference between Pakistani Air Force Cadets and American students.

Loona has translated School Social Behavior Scale (Merrell, 1993) into Urdu in 2002. Translated version was finalized through multi-step process of translation, back-translation and conduction of committee approach. Later, psychometric characteristics of the translated version were also assessed.

Hanif and Pervez (2003) translated and adapted Teacher Stress Inventory (TSI) originally developed by Fimian (1984) into Urdu language for the measurement of levels and sources of stress among school teachers in Pakistan. The inventory was translated by using "back translation" technique and was administered to 120 school teachers from Islamabad, Rawalpindi and Chakwal. The results indicated that Pakistani adapted version of TSI was a reliable and valid scale that can be used for measuring levels and sources of work stress for Pakistani school teachers.

Alla-ud-Din (2003) translated the Ambivalent Sexism Inventory (ASI; Fiske & Glick, 1994) into Urdu with an aim to establish relationship between ambivalent sexism, gender stereotyping, and sexual harassment among professionals. Original inventory was translated, back-translated and then finalized through committee approaches. Researcher also established psychometric characteristics of the translated inventory.

Naqvi (2007) translated, adapted and validated Eysenck Personality Questionnair (EPQ-Junior) in order to explore the patterns of delinquency and personality traits of adolescents in child labor. The reliability of the instrument was found to be satisfactory.

Khan (2008) has translated Mental Health Inventory (MHI; Veit & Ware, 1983) in Urdu for research purposes. She aimed to explore relationship of parental mental health with children behavioral problems and role of moderating factors in it.

Zahid and Pervaiz (2009) conducted a study to translate, adapt and validate the Children's Action Tendency Scale (Deluty, 1979) for Pakistan. After initial test translation, committee approach and translation review strategy was applied to finalize the adaptations and translation. Adaptations were made in two items including item 2 and 6. Translated version was administered on a sample of 88 children for establishing reliability and validity. Two separate studies were conducted to establish convergent and discriminant validity of the translated version.

Rizvi (2009) has adapted and translated Male Role Norms Inventory-Revised (MRNI-R; Levant, 2007); and Fatherhood Scale (FS; Dick, 2004) for exploring father's masculinity ideology and their adolescent's perception. She has made changes in few items of both measures and also established their psychometric properties.

Ikram (2009) has translated Work-life Balance Scale (Kopelman, 1993) for her research exploring relationship between work-life balance and personality factors. Similarly, Yusuf (2009) has translated Social Impact Scale (Fife, 1995) and the Stress Inventory (Nagan, 2007) for the research purpose. Translation and back-translation technique was used and reliability for all sub-scales of the two measures was also established.

Shahid (2010) has translated Multidimensional Scale of perceived Social Support (Zinet, Dahlem, Zinet, & Farley, 1988) for exploring the perception of social support, marital satisfaction, and coping strategies among infertile couples.

Ashraf (2010) has adapted and translated Controlling Behavior Scale (Archer & Kevan, 2003) in order to study women experience of husband's controlling behavior and their self-esteem. Researcher followed the Brislin (1976) approach for test translation and also modified one item of the scale. This was follow by a try-out on 50 wives to assess the translated scale's functioning.

Shamama-Tus-Sabah (2010) has translated Chaos Scale (Matheny et al., 1995) in order to study relationship between chaos at home and school children's cognitive ability and socio-emotional adjustment. Researcher translated and back-translated the scale following the Brislin (1976) procedure but no adaptive changes were made. Psychometrics for the translated version were also established.

Akhter, Hanif, Tariq, and Atta (2011) have translated Parenting Style Dimensions Questionnaire (PSDQ; Robinson, Mandleco, Olsen, & Hart, 1995) into Urdu in order to determine parenting styles for the research purpose. They have also established psychometric properties of the translated questionnaire.

Loona and Kamal (2011) have translated and adapted the Disruptive Behaviour Disorder Rating Scale (Pelham, Gnagy, Gremslade, & Milich, 1992) for assessment of childhood behavioral problems in Pakistan. Back-translation technique was implied and the adapted version was validated through exploratory factor analysis (EFA). EFA resulted in four-factor structure.

Adaptation of Cognitive Tests. These tests play important roles in clinical settings, personnel selection, and psycho-educational assessment. Some of the studies involved in adaptation of these tests are as following.

Tareen (1982) translated and adapted Wechsler Intelligence Scale for Children, Revised (WISC-R) in Urdu. It was a language adaptation along with few changes in order of the administered items due to changes in their difficulty level as a result of translation into Urdu. Some items were only translated while others were adapted culturally and then translated. For example, the Information subtests item: Name the two countries that border the United State? was adapted and translated as: پاکستان کے دو سرحدی ہمسایہ ممالک کے نام کیا ہیں؟

Few original versions' items were completely replaced by new ones. For example, the item: What are hieroglyphics? was replaced with the Urdu item: کتابت کسے کہتے ہیں؟

Ahmad and Aziz (1993) had adapted Columbia Mental Maturity Scale (CMMS; Burgemeister, Blum, & Lorge, 1972) for Pakistan. The study also aimed at exploring the validity of the test for school children in Pakistan. The subjects consisted of 270 children of nursery through grade five, aged 3 to 10 years in urban schools. The reliability and validity indices as obtained in the study characterized the CMMS-P as a useful tool for children in Pakistan, especially in the school setting.

Rana (1995) undertook a study to explore the learning disabilities in Pakistani school children. For that purpose she translated WISC-R in Urdu. But that translation was just used for the specific research purpose and no norms development or psychometric analysis was done.

Kausar (2007) developed indigenous norms for WISC subtests of Information and Arithmetic for public and private school children of 6^{th} and 7^{th} class. She did a gender based comparative study. Results showed that there was no statistically significant difference in performance of boys and girls on subtest of Information but boys performed significantly better than girls on the subtest of Arithmetic. It was also found that many factors like parental education level effects intelligence.

Ambreen (2008) adapted Verbal Comprehension Index subtests of WISC-IV UK and also developed its norms for Pakistani children. Considering the inappropriate difficulty level or cultural irrelevance, five items in the vocabulary tests were replaced by new words while order of administration of two items was also changed. Similarly in the information subtest, four items were replaced by new items. Scaled score, composite score, percentile, and age-equivalent norms were developed.

Noor and Najma (2009) have translated Wechsler Memory Scale (WMS- III; Wechsler, 1997) in Urdu for their comparative study of memory deficits in younger and older adults. After translation the subtests were pre-tested and then scores on original and translated versions were compared and assessed for consistency.

All these studies have shown that test adaptation and translation is a common trend in Pakistani researches. Important issue is to what extent these adaptations and translation has followed the international guidelines or standards (see International Test Commission, 2010, 2012). Review of the current literature indicates that though some of the important guidelines have been followed, few others were totally ignored. For instance, in almost all test translations and/or adaptations proper translation design was followed (mostly back-translation design) and care was also taken to validate the adapted test on the local population. But psychometric analyses of these adapted/translated instruments remain limited to reliability and validity establishment, proper item analysis was almost ignored. Similarly, use of statistical techniques to establish equivalence of the adapted tests with original version was also ignored. Furthermore, this literature review has also pointed out that relatively fewer studies had carried out adaptation of cognitive or intelligence tests.

Rationale for the Present Study

The present research is aimed for the adaptation, translation, and standardization of Wechsler Intelligence Scale for Children – Fourth edition (WISC-IV) in Pakistan so that intelligence of Pakistani children can be assessed more effectively and reliably.

Intelligence testing or intellectual assessment is an important component of vocational, psycho-educational, and neuropsychological testing. Intelligence test scores can provide an indication of current cognitive abilities, aptitude and potential performance at various occupations, and they can also indicate whether skills are markedly below expectations for academic and occupational background, perhaps as a consequence of heredity or any acquired injury or illness.

Like many other developing countries, intelligence testing is gaining attention in Pakistan for educational, clinical, and research purposes. Many intelligence tests are in use in Pakistan by different organizations and testing services. Some efforts have also been made to develop indigenous intelligence tests, but due to several reasons these efforts can not be considered as satisfactory. Firstly, most of these tests are non-verbal in nature while in educational research measurement of verbal intellectual ability is immensely importance. Secondly, most of these tests are not standardized adequately, so due to limited normative data, they are not applicable or interpretable throughout the country. Moreover, most of them are not following the recent intelligence testing trends. In the presence of such difficulties like having limited resources and expertise, it is always advisable to adapt already existing well-established tests against local criteria (Hambleton & Pastula, 1999).

Test adaptation and/or translation is beneficial as it limits duplication of efforts in test construction; save test developmental cost; help in achieving fairness in assessment; and facilitate comparative studies across cultures. Specially, for cross-cultural comparison adaptation of test is almost essential. Many tests have been adapted and translated for testing purposes in Pakistan (see for example, Anila, Khan, & Pervaiz, 1991; Chishti, 2002; Zahid & Pervaiz, 2009) but very few of them are intelligence tests.

Moreover, the intelligence or cognitive ability tests that have been adapted in our country can not be considered psychometrically sound due to limited adherence to the standard test adaptation guidelines and inadequate psychometric evaluation or standardization. For example, in a study Rana (1995) aimed a cognitive investigation of learning disabilities in Pakistani school children. For this purpose many tests were used including WISC-R. These tests were translated and adapted according to the requirements but no attention was given regarding establishment of equivalence with the original version or validation of adapted version against local criterion. So there is a need of a comprehensive intellectual assessment tool in Pakistan. Especially for children the importance for a comprehensive, well-standardized intelligence assessment tool is multifold. Intellectual ability needs to be assessed to predict child's learning potential and it is of great utility in clinical and research fields as well. WISC-IV with its comprehensive interpretative system and psychometric strength can be of utility in all the above mentioned fields if it is adapted properly in Pakistan.

WISC is one of the most widely used tests for children in the world especially in English language speaking countries. It is the language and the cultural loadedness of its content that limit its use in non-English speaking countries. Many of the subtests included in WISC-IV are measures of crystallized intelligence which is cultural specific. So, cultural consideration has become much important in measuring and interpreting these subtests. Moreover, cultural demographic differences act as contextual mediators of cognitive performance and skill acquisition (Weiss et al., 2006). For instance, Panicker et al. (2006) compared the WISC-III ^{UK} adaptation scores of Indian primary school children when Indian and UK norms were used. The children scores showed a drop of 16, 21, and 20 points on Verbal, Performance, and Full scales, respectively when UK norms were applied. To compensate for that limitation all revisions of WISC has been extensively translated or adapted throughout the world.

Considering WISC-IV, there have been many translations and adaptations of the, and norms have been established for a number of countries and languages, for example, Spanish (U.S., Spain, and Mexico), French (France and Canada), German (Germany, Austria, and Switzerland), English (Canada, and United Kingdom), Welsh, Dutch, Japanese, and Chinese (Sanchez-Escobedo, Hollingworth, & Fina, 2011). In Pakistan, clinicians, researchers, and educationists can not benefit effectively from WISC due to language, cultural, and ethnic differences. So its adaptation and translation following all the adaptation guidelines would help in assessing intellectual ability of children effectively.

Ambreen and Kamal (2011) adapted the Verbal Comprehension Index subtests of WISC-IV. The adapted tests proved quite reliable and valid for Pakistani children but the language still seemed to restrict its application with children who are not good in English. Poor performance of most children included in the sample on the subtest of Vocabulary gave support for the notion that translation of the subtests in Urdu would enhance the applicability of the test in Pakistan. Moreover, Urdu translation of items may result in improving the response time of the children. This would also result in ease in the performance of children who are good in English as it is still their second language. Many psychologists argued that due to linguistic variations between countries, many items of intelligence tests may be are biased to specific environment or cultural setting (e.g., Jensen, 1988; Hambleton, 1994). So in order to ensure adequate applicability in the local environment rigorous translation along with adaptation of the intelligence tests is needed.

Moreover, adaptation of only verbal part of the test is not enough for achieving adequate applicability of any test. It is a general view that only verbal tests need adaptation due to language problems and cultural loadedness of their content, while perceptual or performance tests have universality in nature. But researches proved that even performance test have cultural biases that could influence the resulting interpretations. For example, in the adaptation process, Japanese psychologists were highly sensitive to their environment with respect to the issue of disabled people. For this reason, items related to missing organs in the human body or parts of animals in the Picture Completion test of WISC-III were replaced with other things. Perhaps, after World War II the issue of the handicapped became very sensitive in Japan (as cited in Khaleefa, 2006). Similarly, it is also important to deal with administration issues, like time limit for the tests or test items with care. So, the time limit for some subtests and some items was shortened (e.g., Coding and Symbol Search) from 120 seconds to 90 seconds in the Japanese adaptation of WISC-III (Ueno & Nakatanias as cited in Khaleefa, 2006). While time limit for few subtest items was increased from 120 to 150 seconds for the Sudanese adaptation of WISC-III (Hussain as cited in Khaleefa, 2006). So, extensive adaptation and translation of all the subtests of WISC-IV is needed to get maximum benefit from it.

In addition to the adaptation and translation, adequate standardization of the test is very important for the indigenization of any test in a particular culture. So validation and local norms development is essentially required for getting reliable and valid results from a test. So, considering the above discussion, the current study is set to adapt and translate the Wechsler Intelligence Scale for Children – Fourth edition (WISC-IV) subtests in Pakistan along with adequate standardization of the adapted version.

OBJECTIVES AND RESEARCH DESIGN

Objectives and Research Design

Objectives of the Research

The present research is conducted to achieve the following objectives:

1. To adapt and/or translate all the subtests of WISC-IV for Pakistani children.
2. To establish reliability and validity evidence for the adapted subtests of WISC-IV.
3. To develop norms for the adapted subtests of WISC-IV in Pakistan.
4. To explore the relationship of age, gender, geographical region, and parental education level with children's subtests scores.

Research Design

The above mentioned objectives were achieved through an iterative process of four years. This process was completed through three studies. The adaption and the psychometric evaluation of the adapted WISC-IV was carried out in the first two studies. While after finalization of the WISC-IV ^{PAK} its standardization was carried out in the third study. Each study with its specific aims and objectives was completed through various phases and is discussed in separate chapters for descriptive clarity. The research design with its various studies is further elaborated below.

Study I - Adaptation of the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV) for Pakistan. This study was mainly concerned with adaptation and/ or translation of the WISC-IV subtests. It involved various qualitative and quantitative techniques to achieve the aim of adapting WISC-IV. The study was completed through two phases and is described in chapter III.

Phase I: Pre-testing of the WISC-IV. The initial phase of the study aimed at pre-testing of the original instrument in order to have empirical evidence of the need to adapt and/or translate it.

Phase II: Translation and Adaptation of WISC-IV. The later phase of this study involved various judgmental or qualitative steps required to adapt and translate the instructions and/or items of all WISC-IV subtests.

Study II - Psychometric Evaluation of the Adapted Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV ^{PAK}). As mentioned in the study title, study II is concerned with exploration of psychometric strengths and weaknesses of the adapted WISC-IV subtests. The aim of this exploration is to finalize the subtests of WISC-IV ^{PAK} for the standardization stage of research process. This study also involved various steps and during these steps various statistical or psychometric techniques are used to improve the items of the subtests. Study II was completed through three phases and is described in chapter IV.

Phase I: Tryout I. In the first phase of this study the newly adapted subtests of WISC-IV were field tested on a small sample of 33 children in order to assess the item functioning and comprehensibility.

Phase II: Tryout II. In this phase all the subtests of the adapted WISC-IV were tried out on a relatively larger sample of 88 children. The aim was to improve the psychometric strength of the subtests by identifying, and then reordering and/or replacing weak items.

Phase III: Tryout III. A third tryout was conducted during this phase on a sample of 110 children in order to assess the psychometric strength of the modified subtests that functioned unsatisfactorily in the tryout II.

Study III - Standardization of Urdu adaptation of Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV ^{PAK}). As the title implies, this study was aimed at standardization WISC-IV ^{PAK} on Pakistani population. Standardization process mainly involved derivation of norms along with provision of reliability and validity evidence. A secondary aim of the study was to explore the influence of age, gender, geographical area, and parental education level on the intelligence scores of the children. This study was completed through four phases and is described in chapter V.

The entire standardization process including establishment of psychometric properties and norms development was carried out on a normative sample of 800 children. A stratified random sampling design was implied for selecting normative sample from different areas of Pakistan. The sample was stratified on age, gender, geographic area, and parental education level.

Phase I: Establishment of Reliability Evidence for the WISC-IV ^{PAK}. The reliability evidence including subtest temporal stability and internal consistency evidence was established during this phase. For establishing test-retest reliability a separate sample of 34 children was used.

Phase II: Validation of WISC-IV ^{PAK}. The construct validity (convergent and discriminant) of the newly adapted test was established during the phase II. This phase also involved the cross validation of factorial structure of WISC-IV ^{PAK}.

Phase III: Development of Norms for the WISC-IV ^{PAK} in Pakistan. This phase was primarily concerned with the process involved in development of norms for the WISC-IV ^{PAK}. Two types of norms were developed including standard scores (scaled scores, composite scores) and test-age equivalent norms.

Phase IV: Exploring relationship of few Demographic Variables with Children's Scores. This phase explored the association of age, gender, geographical area, and parental education level with intelligence in terms of index scores and FSIQ of children to achieve the secondary aim of the study III.

STUDY I- ADAPTATION OF WECHSLER INTELLIGENCE SCALE FOR CHILDREN, FOURTH EDITION (WISC-IV) FOR PAKISTAN

Study I - Adaptation of the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV) for Pakistan

Study I includes all the steps involved in the proper adaptation and/or Urdu translation of all the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV) subtests. The study was completed through two phases. Each phase itself involved various steps.

Objectives of Study I

Main objectives of the study I are as following:

1. To identify the problems in the original WISC-IV subtests regarding the linguistic comprehensibility, difficulty level, and cultural relevance.
2. To adapt those item(s) of all the subtests that has been identified as being culturally inappropriate.
3. To translate items and other content of all the items if required and recommended.

Phase I: Pre-testing of WISC-IV

It was concerned with pre-testing of the original WISC-IV subtests on Pakistani children to assess their cultural appropriateness and comprehensibility. The findings of this phase would give empirical evidence that WISC-IV needs adaptation and/or translation to be used effectively in Pakistan.

Sample. Twelve children from the age range of 6-16 years 11 months participated in pre-testing including four girls and eight boys. Four children were taken from each of the three age groups of 6-8 years, 9-12 years, and 13-16 years. The children were selected at convenience and were studying in different government and private schools and colleges of Rawalpindi and Islamabad.

Instrument. These include the Informed Consent Form (attached in Appendix A) and Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV) along with its Record Form, Response Booklet 1, Response Booklet 2, and Scoring keys. This version of WISC-IV has been constructed (by NCS Pearson, India) for use in South Asian countries and its administration manual has items written in English and Hindi. This version is not yet finalized and standardized. All the subtests used for administration were in English (see Appendix B1, B2, & B3 for WISC-IV Record Form in English, Response Booklet 1 & 2, respectively).

Following the nomenclature and subtest categorization of all other versions of WISC-IV (see for example WISC-IV UK; Wechsler, 2004), WISC-IV South Asia also have 15 subtests including ten core and 5 supplemental subtests. These subtests have been categorized into four index scores. The index scores are as following:

1. Verbal Comprehension Index- VCI with three core subtests including Similarities (SI), Vocabulary (VC), and Comprehension (CO); and two supplemental subtests including Information (IN) and Word Reasoning (WR).
2. Perceptual Reasoning Index- PRI with three core subtests including Block Design (BD), Picture Concept (PCn), and Matrix Reasoning (MR); and one supplemental subtest of Picture Completion (PCm).

3. Working Memory Index- WMI has two core subtests Digit Span (DS) and Letter-Number-Sequencing (LN); and one supplemental subtest of Arithmetic (AR).
4. Processing Speed Index- PSI again with two core subtests including Coding (CD) and Symbol Search (SS); one supplemental subtest of Cancelation (CA). For detailed description of indices see page 16 and 17 of chapter I.

Procedure. After taking informed consent from the participants and their parents subtests were administered individually on the children. The administration did not follow all the standardized procedures for example, discontinuation rule was not followed in order to have maximum information on all the items of the subtests (according to discontinuation rule examiner has to stop the item administration for a particular subtest on having specified number of consecutive incorrect responses from the examinee). Problems in content of items, instructions or administration, and cultural appropriateness of the items were noticed during the administration. A feedback session proceeded the pre-test administrations to have children's view on difficulty, comprehensibility, and relevance of the items and subtests.

Results and Discussion of the Phase-I. The pre-testing was aimed to assess the cultural appropriateness of the original WISC-IV subtests. Considering the findings of Ambreen and Kamal (2010) regarding linguistic problems of VCI subtest in Pakistan, a small sample was taken to assess original test's functioning. Responses of students, observations taken during the pre-testing, and the student's feedback were analyzed. Considering the test completion time, administration time for the core subtest seemed to be 90-120 minutes while WISC-IV administration and scoring manual (Wechsler, 2004) indicated 65-80 minutes as administration time for the core subtests.

Verbal Subtests. Many language, cultural, and familiarity based problems in the content of the items and instructions were identified after assessing the responses of the students and their views in the feedback sessions held after the administrations. Language related problems were identified in almost all the verbal subtests and in case of comprehension subtests length of few items was also pointed out to be problematic. Children also commented negatively on vocabulary subtest as having much difficult English words. The culture based problem was indicated only in item 28 (regarding winter solstice) of the information subtest. Familiarity related problems were indicated most frequently. In comprehension subtest the item 11 regarding mid-day meals was considered less familiar by the participating children. Whereas, in information subtest five items (item no. 18, 21, 28, 29, & 32) were identified as being less familiar to the children. In these items knowledge about less familiar or unfamiliar persons and tasks has been assessed. Similarly, in arithmetic subtest names of the people in the content of the items were considered less familiar by the children.

Such problems have also been indicated in various other test adaptation studies. For instance, Malda et al. (2008) also indicated that cognitive test constructed in the westernized countries may require several language, culture, and familiarity-recognizibility driven adaptations (also construct and theory driven) for their effective use in non-westernized countries.

Non-Verbal Subtests. All the non-verbal subtests items of WISC-IV seemed to be functioning well except 'processing speed index (PSI) subtests'. But language comprehensibility of instructions to the examinee seemed quite poor. Processing Speed Index subtests (Coding A & B; Symbol Search A & B; & Cancellation R & S) seem to require changes in the time limits especially for the subtest of Symbol Search (Symbol Search B). Most student performance was judged to be below average on the PSI subtests according to the UK norms.

This finding and identified problem in subtests involving speed is also pointed out in few other studies involving adaptation of cognitive tests (for example Khaleefa, 2006; Malda et al., 2008). Khaleefa (2006) compared the adaptation process of the WISC-III in Sudan and Japan and observed that in Japan, the time limit for some subtests was shortened from 120 to 90 seconds, by contrast, in Sudan it was increased from 120 to 150 seconds. Overall, this phase indicated a strong need to translate the content of WISC-IV in Urdu, and it also identified few items to be needing adaptation or modification.

Phase II: Translation and Adaptation of WISC-IV

It involved a priori procedures (judgmental procedures) for adaptation and translation of the WISC-IV subtests. These procedures include multiple-forward translation; committee approaches and expert reviews; and then modifications in translation, item-content and administration (wherever required).Phase II was carried out through various steps.

Step I: Translation of the WISC-IV Subtests in Urdu. This step was concerned with translation of required test material (items and examinee directed instructions) in Urdu as suggested in the previous phase.

Translation Committee. Translation for all subtests was done by a 12 membered committee of bilinguals having the qualification of M.Phil in Psychology. Besides being bilinguals care was taken that all translators must have good knowledge of the cultural context and are familiar with the test/item construction rules in general.

Procedure. The required test material was given to the committee members in a way that every member had to translate at least one verbal and one nonverbal subtests. They were asked to translate individually the examinee directed instruction and item content (for all verbal subtests except vocabulary) into Urdu. They were further asked to keep the meaning and difficulty of the translations as close to the original content as possible.

For Vocabulary subtest items, three replacement options in Urdu (for each of the original English word) were taken from 5 individuals (2 educationists & 3 from National Language Authority of Pakistan with minimum qualification of Masters). They were instructed to give these replacement options considering the closeness of meaning and difficulty level of the replacement options with the original words of the vocabulary subtest.

Result. The translated material along with the original content was re-collected from the translation committee for compilation. In total four translations of the verbal subtests and three translations of the nonverbal subtests were congregated.

Step II: Committee Approach. Various committee meetings were held to evaluate the quality of the compiled Urdu translations and to decide about the best translation for all the fifteen subtests of WISC-IV.

Adaptation Committee. A four membered adaptation committee was constructed for this purpose. Three committee members were having doctoral degrees in Psychology, while the fourth member was M.Phil in Psychology. They all are university faculty members teaching to the students of at least Masters level.

Procedure. A total of 11 committee approaches were conducted including eight for verbal subtests and three for non-verbal subtests. During the course of these 11 committee approaches, all the translations of the subtests were thoroughly evaluated by the members along with the researcher. Then the committee decided about the best translations for both items and instructions. For some of the items and test instructions even these selected translations underwent minor modifications by the committee members to have improved verbal fluency.

Committee also reviewed the items and other content of the subtests in order to identify any cultural, informational, and/or familiarity based adaptive changes required in their content or presentation format. They suggested adaptive changes in items of comprehension, information, and arithmetic subtests (see details about the changes in the result section). In identification of items needing adaptation committee also considered the findings of the pre-testing, most of the problematic items pointed out in the previous phase underwent adaption on committee's review. The non-verbal (performance) subtest materials were all viewed by the committee. All the pictures were explored for identifying any culturally inappropriate object depiction. Most of the pictures were judged to be appropriate to be used with Pakistani children except one in picture completion subtest.

Adaptive changes led to the replacement and modification of few subtest items, while sampled response options for few subtests items also underwent addition and/or replacement. For item replacement WISC-IV ^{UK} Administration and Scoring Manual (Wechsler, 2004) and Pakistan's adaptation of Verbal Comprehension Index Subtests (VCI-P; Ambreen, 2008) of WISC-IV ^{UK} was consulted by the committee.

Results. Beside finalization of the subtest translations committee approaches also resulted in various adaptive changes in the subtests. These changes included culture and familiarity based adaptations and resulted in replacement of few items. For instance, item 11 (mid-day meals) of the comprehension subtest was replaced by the comprehension subtest item 11 of WISC-IV ^{UK} (Wechsler, 2004). Similarly, five items of the information subtests were replaced. Out of these three items including item 18 (hieroglyphics), item 21 (Kali Daasa), and item 32 (Confucius) were replaced with the information subtest items of VCI-P (Ambreen, 2008), while item 28 (winter solstice) and 29 (Dubai and Mumbai) were replaced by new items. In arithmetic subtest names of persons used in 15 items were replaced by more common Pakistani names. The adaptive changes also included addition (in item 8 of information subtest- coins) and replacement of various sampled response options (including response options of item 11 of comprehension; and item 8, 21, 28, 29, and 32 of information subtest).

Considering performance subtests, most items were considered appropriate by the committee members for use in Pakistan with one exception. A minor change has been suggested (to NCS Pearson, India) in the item number 14 of picture completion subtest. This picture displays face of a woman with 'Bindia' on the forehead and this was considered inappropriate for Pakistani culture.

Step III: Construction of Urdu Vocabulary Subtest. This step involved construction of list of Urdu words for Urdu vocabulary subtest and then construction of Urdu vocabulary subtest with help of these enlisted words.

Construction of Urdu Vocabulary Items. List of original vocabulary words along with three replacement options were given to a panel of 10 teaching faculty of Masters level with minimum qualification of M.Phil in Psychology. They were asked to choose the best replacement word for each original word (considering closeness of both difficulty and meaning of the replacement word with the original word). Most frequently endorsed replacement options were finally selected to construct list of Urdu words for vocabulary subtest. The minimum criteria for selection was the replacement word chosen as appropriate by at least half (50%) members of the panel. See Appendix C1 & C2 for Initial and Final List of Urdu Replacement Words for Vocabulary Subtest.

This list of Urdu replacement words was then used by the researcher to write and arrange items for the new vocabulary subtest. Urdu Corpus (Urdu 5000 Most frequently Used Words; Ijaz, 2007) was also consulted for deciding about the order of administration of the words.

Vocabulary Sampled Response Options Compilation. New vocabulary items require new sampled response option for suitable scoring. For this purpose, responses were collected from various sources including children, linguistic experts, previous samples responses, and various Urdu dictionaries to compile Urdu sampled response options.

Participants. A total of 50 children and 3 linguistic experts participated in the response compilation. The participating children were from the age group of 6 to 16 years 11 months and included both boys (n=27) and girls (n=23). Whereas, the linguistics experts belonged to the National Language Authority of Pakistan.

Procedure. All the newly constructed vocabulary items were presented to the children and linguistic experts, and they were asked to give multiple responses to those items. Collected responses on all the vocabulary items were then sorted into accurate, relatively correct, and incorrect categories. For further compilation of vocabulary sampled response options many other sources were also considered including Urdu dictionaries and translations of sampled response options of vocabulary items of original subtests. Sample response options for all Urdu vocabulary items were then finalized into the respective categories by a three member committee with the minimum qualification of M.Phil in Psychology.

Results. After construction of new vocabulary items, the response option compilation process resulted in finalization of new Urdu 2-point, 1-point and 0-point sampled responses for all the vocabulary subtest items. This concluded the Urdu vocabulary subtest development process.

Step IV: Expert Review. In this step all the translated and adapted subtests were reviewed by an expert committee for further finalization.

Committee Members. The expert committee was comprised of four members. The verbal translated subtests were reviewed by two bilingual experts (Ph.D. in Psychology and having thorough experience of test construction and/or administration).

Whereas, the non-verbal translated subtests were reviewed by two personnel from National Language Authority of Pakistan (Ph.D. in Urdu/Persian language).

Procedure. All translated subtests were presented to the experts after proper composing and arrangement in the form of a test manual. They reviewed all the contents of the subtests in detail. The reviewed content included the Urdu items and examinee directed instructions; English and Urdu sampled response options; and the English instructions for test administration. They also suggested minor improvement in the wording and arrangement of the subtests.

Results. Keeping in view the suggestions of the experts, necessary changes were made in subtest content. At few places translation/wording of instructions were modified for verbal fluency and improved comprehensibility. Whereas, at some other places minor re-arrangement of content was carried out for visual facilitation. Another important suggestion of expert committee came for the Digit Span subtest. Besides previous committee's decision experts have shown some concern on administration of digits in English for Digit Span subtest. They were of the view that digit presentation should be tried out in both English and Urdu before finalization of its administration medium.

Step V: Sub-study on Digit Span Subtest. On expert committee's suggestion a sub-study was conducted to empirically observe the effect of language on subtest that involves presentation of digits (digit span).

Sample. Twenty children from the age group of 6 to 16 years and 11 months participated in this step ($n = 10$ for both boys and girls).

Procedure. Digit span forward and backward was administered to the children in English initially. The responses were recorded along with the administration time. Then the items of digit span were presented in Urdu (i.e. digits were called in Urdu). The children's responses and administration time was again recorded to have a comparison of children's performance.

Results and Discussion. The sub-study was conducted to see the influence of English or Urdu language on performance of children on Digit Span subtest. This exploration was based on the fact that in simple memory tasks involving digits, differences in phonological length of digit names in different languages may prove critical. For example, the digit four in English is a four lettered word but in Urdu it is a three lettered word, similarly the digit seven's English name is five lettered, whereas its Urdu name is three lettered. This difference in phonological length of digit names in two languages can interplay with the memory span of the children resulting in decrease or increase in their performance on tasks involving memory and attention. So performance of children in terms of correct reproductions or response time may differ when digits are presented in Urdu or English.

This probable difference in performance can be explored through comparison of acquired mean scores of children on the digit span subtest in the two administrations (i.e. on presenting digits in English and Urdu). The aim was to decide which language of digit presentation is more appropriate for Pakistani children.

Table 1

Comparison of Mean Raw Scores of Children on Digit Span in English & Urdu (N = 20)

	English		Urdu		t	p	95%Confidence Interval		d
	M	*SD*	*M*	*SD*			*UL*	*LL*	
Digit span (DS)	17.90	3.65	15.7	2.87	2.15	.038	4.27	.128	0.67
DS-Forward	10.35	2.23	8.75	1.55	2.63	.012	2.83	.369	0.83
DS-Backward	7.55	1.79	6.95	1.87	1.03	.308	1.77	-.57	0.32
LDSF	6.75	1.29	5.90	1.02	2.30	.027	1.59	.104	0.73
LDSB	4.20	1.00	3.95	1.05	.769	.447	.908	-.40	0.24

Note. LDSF= Longest Digit Span Forward; LDSB= Longest Digit Span Backward; df = 38

Table 1 shows mean differences of children's scores on digit span subtest when it is administered in English and Urdu digits (on paired sample t-test). Results indicated that children performed significantly better when Digit Span is administered in English.

This finding is inconsistent with several previous test adaptation based researches (see for example Malda et al., 2008) that have followed Baddeley's phonological model. According to the Baddeley's phonological loop model (Cowan, Baddeley, Elliott, & Norris, 2003), the number of items that can be stored in the memory varies with their phonological length (such as the number of syllables).But the task familiarity and practice issue might explain this significantly high performance, according to the teachers and the children themselves nowadays Urdu counting is less frequently used in Pakistan especially in school settings. Interestingly, these performance differences were non-significant for Digit Span backwards so may be level of attention is also important (DS backward is considered more difficult so children might have attended the presented stimuli equally in both languages). Overall, the sub-study resulted in decision of administering Digit Span subtest with digits in English language.

Step VI: Cognitive Interviews. This step involved interviewing children regarding the linguistic comprehensibility of the instructions and items of adapted WISC-IV sub-tests. The objective was not to assess the children responses but to judge the level of understanding children show about the content of sub-tests instructions and items.

Participants. Three boys and two girls belonging to the age group of 6 to 14 years were interviewed for this purpose.

Procedure. Subtest instructions were presented to each child individually and they were asked to reproduce whatever they have understood from the said instructions. Similarly, all the translated and/or adapted items of comprehension, arithmetic, and word reasoning subtests were presented to the children to assess their level of understanding about the items. Their feedback on the test content was taken afterwards.

Results. All the children reproduced the subtest instructions and the items of the verbal subtests in the correct manner indicating level of comprehensibility of the test content as appropriate. In the feedback session only one 6 years old girl commented on some part of the content as being difficult to understand including last four items of comprehension subtest (high difficulty items) and instruction of symbol search subtest. All other children regarded the test to be interesting and engaging.

This step concludes the efforts to adapt WISC-IV subtest for use in Pakistan (see Appendix D1 for List of Items suggested for Adaptation along with the Replaced Items, and D2 for List of Changes in Content of WISC-IV Items or Sampled Responses based on Pre-testing/Expert Opinion).

The study I was completed with efforts to develop Microsoft Word files for the publication of the adapted WISC-IV Manual, Record Form, and Response Booklet 1. These word files have all the child directed instructions and sub-test items written in Urdu (Urdu InPage, 2000 was used), while the examinee directed instructions and sampled responses (except for vocabulary subtest) were kept in English. The final copy (Microsoft Word file) of the adapted test and record Form was sent to NCS Pearson India Private Limited, Bangalore, India for further processing. This was followed by cycles of content editing and formatting till publication of WISC-IV (Urdu Standardization Edition) Administration and Scoring Manual along with Record Form (Urdu) and Response Booklet 1 (see Appendix E1 for Record Form-Urdu and E2 for Response Booklet 1 in Urdu). Response Booklet 2 was kept as original.

During all the steps taken for adaptation of the subtest the test adaptation guidelines were kept in mind. This process started with taking the decision of translating and adapting the test on empirical grounds (through pre-testing). Then the selection of translators and consultations with educationists, psychologists, linguistic experts, and testing experts were carried out to take into account all cultural, linguistic, and contextual consideration. This was all in line with the international guidelines for translating and adapting tests (see for example International Test Commission, 2010, 2012). But as indicated in ITC guidelines this judgmental evidence of appropriate test adaptation is not enough, all the adapted items needs to be field tested for psychometric evaluation. So the study I was concluded with availability of WISC-IV ^{PAK} for further field testing.

STUDY II- PSYCHOMETRIC EVALUATION OF THE ADAPTED WECHSLER INTELLIGENCE SCALE FOR CHILDREN, FOURTH EDITION (WISC-IV ^{PAK})

Study II – Psychometric Evaluation of the Adapted Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV ^{PAK})

This study involved field testing of newly adapted/translated test and was aimed at detailed psychometric evaluation of WISC-IV ^{PAK}. The study was completed through three phases.

Objectives of the Study II

1. To establish evidence that newly adapted subtests are functioning well.
2. To explore the psychometric properties of all subtests of WISC-IV ^{PAK}.
3. To conduct item-level analysis of all WISC-IV ^{PAK} subtests.

Phase I: Try-out I: Preliminary Field Testing

It was conducted as a preliminary field test to assess the functioning of adapted/translated instructions and items of all the WISC-IV ^{PAK} subtests through test administration and feedback sessions.

Sample. Thirty three students from the age group of 6 to 16 years 11 months (n = 3 for each of the 11 age groups of one year) participated in the tryout I. The sample was divided in two gender groups of boys (n = 20) and girls (n = 13). All the participants were students of various government schools and colleges of the Rawalpindi and Islamabad. These schools were selected due to the similarity in their curriculum, medium of instruction, and level of cultural and financial diversity of the students.

Instrument and Materials. These include the Informed Consent Form and WISC-IV PAK kits. Each kit includes WISC-IV (Urdu Standardization Edition) Administration and Scoring Manual; Record Form (Urdu); Response Booklet 1 and 2; and a Stimulus Booklet along with block design box and scoring keys.

Informed Consent Form. It includes topic along with the purpose of the research. It also has the information regarding rights of the participants about maintenance of confidentiality and right to withdraw information during course of the research. On the basis of information provided duly signed informed consent is taken from the participants (see Appendix A for the Informed Consent Form).

Wechsler Intelligence Scale for Children, 4th Edition- Pakistan (WISC–IV PAK). The WISC-IV has 10 core and five supplemental subtests that can be summed up into four indices, and one Full Scale IQ. It can be administered to children from the age range of 6 to 16 years and 11 months. The four composite indices (with adaptive changes in subtests of three indices) are:

Verbal Comprehension Index (VCI). It requires verbal conceptualization, stored knowledge access and oral expression. The VCI is a measure of verbal concept formation. Due to having verbal content, many items of its subtests got adaptation. The subtests included in this index are: Similarities; Vocabulary (adapted); Comprehension (adapted); Information (adapted); and Word Reasoning.

Perceptual Reasoning Index (PRI). It requires visual perception, organization and reasoning with visually presented, nonverbal material to solve the kinds of problems that are not school taught. The PRI is a measure of non-verbal and fluid reasoning. It includes sub-tests: Block Design; Picture Concepts; Matrix Reasoning; and Picture Completion (adapted).

Working Memory Index (WMI). It requires working memory processes applied to the manipulation of orally presented verbal sequences. It involves attention, concentration, mental control, and reasoning. Its subtests are: Digit Span; Letter-Number Sequencing; and Arithmetic (adapted).

Processing Speed Index (PSI). It is a measure of processing speed and assesses children's abilities to focus attention and quickly scan, discriminate between, and sequentially order visual information. The subtests included are: Coding; Symbol Search; and Cancellation

There are in total 15 subtests including 10 core and 5 supplemental subtests. Items and instructions of all these subtests have been translated into Urdu and a brief description of these subtests is given below. The subtest task definitions are reported as cited in Flanagan and Kaufman (2004):

1. Block Design (BD): The examinee is asked to reassemble a set of 10 modeled or printed two-dimensional geometric patterns using red and- white blocks within a given time limit. It has 14 items and each item can be scored from 0 to 7.

2. Similarities (SI): The examinee is required to explain that in what way two words that represent common objects or concepts are similar. It has 23 items and every item can have a score of 2, 1, or 0.

3. Digit Span (DS): The examinee is required to repeat numbers verbatim as stated by the examiner in forward or reverse order. It has 16 items in total (with two trials in each item) that can be scored as 2, 1, or 0.

4. Picture Concept (PCn): The examinee is requested to pick one picture, from among two or three rows of pictures presented, to form a group with a common characteristic. It has 28 dichotomously scored items.

5. Coding (CD): The examinee is asked to copy symbols that are paired with either geometric shapes or numbers using a key within the given time.

6. Vocabulary (VC): The examinee is required to name pictures or provide definitions for words. This subtest has all newly adapted verbal items and sampled response categories. It has 36 items including 4 picture and 32 verbal items with scoring categories of 2, 1, and 0.

7. Letter-Number Sequencing (LN): The examinee is presented with a number and letter sequence and is asked to recall numbers in ascending order and letters in alphabetical order. It has 10 items (three trials in each) with scoring categories of 3, 2, 1, and 0.

8. Matrix Reasoning (MR): The child is asked to complete the missing portion of a picture matrix by selecting one of five response options. It has 35 dichotomously scored items.

9. Comprehension (CO): The examinee is required to answer a series of questions based on his or her understanding of general principles and social situations. It has 21

items including one adapted/replaced item (item 11). It has three scoring categories of 2, 1, and 0.

10. Symbol Search (SS): The examinee is requested to scan a search group and indicate the presence or absence of a target symbol(s) within a given time limit.

11. Picture Completion (PCm): The examinee is required to view a picture and name the essential missing part of the picture within a specified time limit. It has 38 dichotomously scored items with item 14 as an adapted item.

12. Cancellation (CA): The examinee is asked to scan both a random and a nonrandom arrangement of pictures and mark target pictures within a specified time limit.

13. Information (IN): The examinee is required to answer questions related to a wide range of general-knowledge topics. It has 33 dichotomously scored items including five replaced items (item 18, 21, 28, 29, & 32).

14. Arithmetic (AR): The examinee is required to mentally solve a variety of orally presented arithmetic problems within a given time limit. It also has 33 dichotomously scored items with adapted names in 15 of its items.

15. Word Reasoning (WR): The examinee is required to identify a common concept being described by a series of clues. It has 24 dichotomously scored items.

Procedure. After taking informed consent, all the adapted and/or translated subtests were individually administered on the children. Administration followed all the standardized procedures including the starting age, reverse and discontinuation rules for all subtests except for Processing Speed Index (PSI) subtests (discontinuation rule was not followed in order to estimate the average completion time taken by the students for these subtests). After administration, feedback of the children was also taken regarding difficulty, comprehensibility and cultural appropriateness of the items and instruction of the adapted subtests. Children's responses to all items were recorded carefully not just for scoring purpose but also for identification of any response that can be added to sampled correct or incorrect responses being true or a common erroneous response for our culture.

Results and Discussion. These results are not only based on the responses of the children on newly adapted subtests of WISC-IV but also on observations taken during administration and feedback session. Preliminary statistical techniques like compilation of response frequencies and reliability analyses of the subtests was done in order to assess the item functioning.

During administration of subtests it was observed that few Urdu words in similarities subtest cannot be properly comprehended by the children (especially of age 10-13 years). It was also observed that children proceeded very slowly on Processing Speed Index (PSI) subtest as most of the children did not even completed half of the subtests in the allowed time.

Response Frequencies. Compilation of response frequencies on all subtests indicated that almost all subtests are functioning well. Though a detailed evaluation of the response frequencies indicated poor functioning of some of the items. The details are as following:

Considering the verbal subtests, all the items of the similarities subtest indicated satisfactory correct response rate except for item 16 on which only two children responded correctly. For vocabulary subtest poor response rate was observed from item 21 onwards on vocabulary subtest. Beside that one initial item (considered as having low difficulty) was also indicated to be problematic, 13 out of 33 students responded incorrectly to item 7 (clock) of vocabulary subtest. As this item implies the reverse scoring rule, so it may require reordering or modification of translation. In comprehension subtest all items were indicated to be functioning well including the adapted item 11 as almost 58% children responded it correctly. Considering adapted items of information subtests, though item 18 and 21 was responded correctly by 3 and 4 students respectively, but item 29 being a high difficulty item was responded correctly by only 2 students; while item 28 and 32 were not responded correctly by any student. Whereas, in arithmetic subtest all items were indicated to be functioning well considering their order of difficulty except one. On item 21 of arithmetic, only 10 students gave correct response, so almost 70% students responded it incorrectly. In Word Reasoning all items seemed to be functioning well considering their order of difficulty.

In case of nonverbal subtests, all PRI and WMI subtests were functioning well as children responded satisfactorily on these subtests. But on timed subtest such as coding and symbol search, besides showing interest most of the students did not even reach half of the items in due time raising some concerns about time limit of these subtests.

As a priori procedures of test adaptation involve iterative cycles of translations/ adaptations, pilot-testing/tryouts and then modifications (Malda et al., 2008) so try-out I results led to few more changes in the items of Urdu WISC-IV subtests. Item 21 of 'arithmetic' subtest was replaced with a new item as the concept of paisa to rupee conversion did not seem much familiar to the Pakistani children. Similarly after assessing the response rate, two items of the newly developed 'vocabulary' subtest (item 7 & 36) were also replaced considering its order and difficulty level (Urdu words were replaced with new Urdu words with close meaning but different difficulty level). In terms of language, one change was made in Urdu translation of one of the items (item 12) of 'similarities' subtest. Furthermore, on the bases of children's response patterns, response options were added in 1-point sampled response category of two (item 9 & 15) comprehension items (see Appendix D1 &D2 for details).

Reliability Analysis. Alpha coefficients for all the Index scores and subtest score were found to be satisfactory. The alpha coefficients for Verbal comprehension Index (VCI) subtests ranged from .89 (information subtest) to .94 (vocabulary subtest), while it ranged from .85 (block design) to .91 (Picture completion) for Perceptual Reasoning Index (PRI). For Working Memory Index (WMI) reliability coefficient ranged from .79 (digit span) to .88 (arithmetic). Considering the index scores the reliability coefficients were .97, .91, .88, and .77 for VCI, PRI, WMI, and PSI, respectively. This indicated that all the subtests seem to have satisfactory psychometric strength to be used further for assessing the intellectual level of Pakistani children.

Considering International Test Commission (ITC) guidelines for test translation and adaptation (2010), adaptation of any test constructed in a developed country to a developing country requires a long process. Adaptation process should include proper cultural and language considerations, expert view, field testing, appropriate statistical design for establishing its equivalence with original version, norms development, and proper documentation of all the procedures and changes involved. So despite of the encouraging findings of try-out I further piloting (or tryouts) with larger samples is required for considering this Urdu translated version of WISC-IV reliable and valid enough for assessing the intelligence level of Pakistani children.

Phase II: Tryout II: Item Analyses

Tryout II was aimed to improve the psychometric strength of the WISC-IV ^{PAK}. Detailed item analysis of adapted WISC-IV subtests was the prime objective for conduction of this tryout.

Sample. It comprised of 88 students from age group of 6 – 16 years 11 months (n= 8 for each of the 11 age groups). Sample was equally divided into the gender groups. The sample was selected from randomly selected schools and colleges of capital territory that are under the administration of Federal Directorate of Education (FDE), Islamabad. Official permission was taken from Federal Directorate of Education for this purpose (See Appendix F1 for the List of Randomly Selected Educational Institutes under Administrative Control of Federal Directorate of Education, Pakistan and F2 for Permission Letter from Federal Directorate of Education, Islamabad).

Procedure for selection of Schools/Colleges. First of all a list of schools and colleges under the administrative control of FDE was taken, then through random table generator a total 20 schools and colleges were selected. After grant of permission to have test administrations in these institutes, meetings were arranged with the principals or institutional heads to take their consent and to schedule test administrations.

Instrument and Materials. These include the Informed Consent Form and WISC-IV ^{PAK} kits were used for test administrations. Each kit includes WISC-IV (Urdu Standardization Edition) Administration and Scoring Manual; Record Form (Urdu); Response Booklet 1 and 2; and a Stimulus Booklet along with block design box and scoring keys. The WISC-IV ^{PAK} Administration and Scoring Manual used in this tryout has some changes. One similarities subtest item (item 12) has been modified, while two vocabulary (item 7 & 36) and one arithmetic subtest (item 21) items have been replaced. Moreover, the item 9 and 15 of comprehension subtest have one added response option in new WISC-IV ^{PAK} manual (For other details see instrument description of Phase I: Tryout I).

Procedure. First of all, administrative heads of all permitted schools and colleges were contacted for consented planning of test administrations in their respective institute. After taking informed consent from the participants all 15 adapted subtests of WISC-IV were administered following all the standardized procedures. A team of five trained test examiners was involved in administration and scoring of the subtest. After recording and scoring of responses all the Record Forms were rechecked by the researcher for administration and scoring errors.

Data sheet for further analysis was constructed under the guidance of NCS Pearson India Private Limited, Bangalore. Detailed item analysis of WISC-IV subtests was done with the help of various software including IBM SPSS (PASW) Statistics 19, SAS (PROC UNIVARIATE), and WINSTEPS (for Rasch and Partial-credit IRT Models).

Results and Discussion of Tryout II. Detailed item analysis of adapted WISC-IV subtests was the prime objective for conduction of this tryout. Both test and item statistics were to be established for the whole sample (N=88) and for the three age groups of 6-8 years (n=24), 9-12 years (n=32), and 13-16 years (n=32). Analyses based on responses of students on 15 subtests of WISC-IV Urdu Standardization Edition were carried out with the help of IBM SPSS (PASW) Statistics 19, SAS (The UNIVARIATE Procedure), and WINSTEPS (for Rasch and Partial-credit IRT Models).

Univariate Analysis of WISC-IV ^{PAK} Subtests. This analysis was conducted to explore the distributions of scores on all the subtests in order to determine characteristics of the data. This will further facilitate in deciding about retaining items later on. IBM SPSS (PASW) Statistics 19 was used to run this analysis.

The analysis includes examining mean scores of all subtests along with the standard deviations, median, and range of scores. The analysis also includes examination of skew statistics (with an associated standard error) which is considered as a good indicator of normality of distribution. Skew in any score distribution can indicate the test/subtest being easy or difficult for any particular sample. In other words, it may indicate the high or low performance of individuals on a particular subtest. Field (2009) has asserted that the value of skewness should be zero in a normal distribution. The more distant is the value from zero, the more is the possibility that the scores are not normally distributed. The informative value of skew statistics can be increased by converting it into a Z-skewness score by dividing it to the standard error of skewness associated with it:

$$Z \text{ - skewness} = S - 0/SE\text{skewness}$$

These z-skewness scores can be compared with the known values for the normal distribution i.e. a z-skewness score above 1.96 is significant at $p < .05$, above 2.58 is significant at $p < .01$, and above about 3.29 is significant at $p < .001$. In small samples looking for values greater than 1.96 is good enough; but in larger samples this criterion should be increased to the 2.58 and above.

Table 2

Psychometric Properties and Score Distribution of Adapted WISC-IV Subtests (N= 88)

Sub-tests/ Variables	*M (SD)*	Median	Range Mini.	Range Max.	Skew (Std. Error)
BD	25.40 (12.76)	26	6	55	0.43 (0.257)
SI	16.01 (8.19)	14	2	38	**0.75 (0.257)**
DS	17.09 (4.27)	16.50	8	28	0.28 (0.257)
PCn	14.45 (3.67)	14.50	5	25	-0.08 (0.257)
CD	46.16 (12.23)	45	22	77	0.39 (0.257)
VC	27.58 (11.37)	24	12	60	**0.87 (0.257)**
LN	16.31 (4.64)	17	5	26	**-0.56 (0.257)**
MR	16.01 (5.65)	15.50	6	27	0.15 (0.257)
CO	19.32 (6.95)	20	6	35	0.07 (0.257)
SS	21.38 (8.05)	21.50	1	40	-0.07 (0.257)
PCm	19.82 (5.93)	20	6	35	-0.09 (0.257)
CA	69.97 (23.04)	68	30	126	0.28 (0.257)
IN	16.07 (5.14)	15	7	29	**0.66 (0.257)**
AR	20.68 (4.91)	22	9	30	-0.48 (0.257)
WR	10.26 (3.12)	10	4	19	0.38 (0.257)

Note. BD= Block Design; SI= Similarities; DS= Digit Span; PCn= Picture Concept; CD= Coding; VC= Vocabulary; LN= Letter-Number-Sequencing; MR= Matrix Reasoning; CO= Comprehension; SS= Symbol Search; PCm= Picture Completion; CA= Cancelation; IN= Information; AR= Arithmetic; WR= Word Reasoning.

Table 2 shows means, standard deviations, and range of scores on the subtests of WISC-IV along with the skew statistics (and associated standard error of skewness). Positive values of skewness indicate a pile-up of scores on the left of the distribution (difficult test), whereas negative values indicate a pile-up on the right (easy test). The skew values for similarities, vocabulary, letter-number-sequencing, and information subtest (values in bold) indicates that they have highly skewed score distributions.

Z-skewness scores for all the subtests were also calculated and most subtests found to have non-significant skew values (score distribution is close enough to be considered as normal distribution) except similarities, vocabulary, LNS, and information subtests. For similarities Z skewness score is 2.91 so there is a significant positive skew ($p < 0.05$) indicating a low performing sample for this particular subtest. For vocabulary Z-score for skewness is found to be 3.38 thus indicating highly significant positive skew ($p < .001$) that in turn suggests a very difficult subtest for this sample. In contrast with these two subtests, the Z-score of -2.17 for LNS indicates a negatively skewed ($p < .05$) score distribution suggesting a high performing sample (relatively easy subtest) for LNS. Score distribution for information was also positively skewed (Z-skewness = 2.56; $p < .05$) suggesting it to be another difficult subtest for the sample.

To further confirm the deviations from normality the Kolmogorov–Smirnov test and Shapiro–Wilk test were run for all the subtests. They compare the scores in the sample to a normally distributed set of scores with the same mean and standard deviation. If the test is non-significant ($p > .05$) it tells us that the distribution of the sample is not significantly different from a normal distribution (i.e. it is probably normal). If, however, the test is significant ($p < .05$) then the distribution in question is significantly different from a normal distribution (i.e. it is non-normal). For smaller samples (*N*= 7-2000) Shapiro-Wilk test is considered more appropriate test of normality (Shapiro & Wilk, 1965 as cited in Guido, 2009; Park, 2008).

Box plot is another way of viewing distribution of scores being normal or not. The gap and distance of upper and lower extreme ends of box plot with 25^{th} and 75^{th} quartile indicates presence of skew. It also indicates clearly if any outlier exists in the score distribution. See Appendix G1 - G15 for Box plots of all subtests of WISC-IV ^{PAK}.

Table 3

Tests of Normality for Similarities, Vocabulary, Letter-Number-Sequencing, and Information Subtests (*N*= 88)

Subtests	**Kolmogorov–Smirnov test**			**Shapiro–Wilk test**		
	D	*df*	*p*	*W*	*df*	*p*
Similarities	0.151	88	.000	0.927	88	.000
Vocabulary	0.156	88	.000	0.914	88	.000
LNS	0.150	88	.000	0.950	88	.002
Information	0.116	88	.005	0.967	88	.002

Table 3 shows test statistics for K-S (Kolmogorov–Smirnov test; denoted by D) and S-W (Shapiro–Wilk test; denoted by W) for the four subtests indicated to have significant skew statistics in the previous table. Both tests confirm significant differences of their score distribution from normal distribution ($p < .001$ for similarities, vocabulary, and LNS subtest; $p < .01$ for information subtest). Normality tests for other 11 subtests also indicated skewed population in few other subtests. For example, Block Design (D (88) =0.132, $p < .01$; W (88) = 0.951, $p < .01$), Matrix Reasoning (D (88) =0.128, $p < .01$;

W (88) = 0.958, $p < .01$), and Arithmetic subtests (*D* (88) =0.117, $p < .01$; *W* (88) = 0.967, $p < .05$) also seem to have significant non-normal score distribution. Whereas, score distribution for Word Reasoning subtest show significant K-S statistics ($D = .122$, $p < .01$), but non-significant S-W statistics.

For seven subtests that have shown non-normal distribution on Z-skewness scores, K-S statistics, or S-W statistics score distribution was also explored for the three age groups of 6-8 years (*n*=24), 9-12 years (*n*=32), and 13-16 years (*n*=32).

Table 4

Psychometric Properties and Score Distribution for the Three Age Groups on Adapted WISC-IV Subtests (N= 88)

Subtests Variable	Age groups	*M (SD)*	Median	Range		Skew (Std. Error)
				Mini.	Max.	
BD	6-8 yr	13.71 (6.73)	14	6	34	1.22 (0.472)
	9-12 yr	24.75 (8.73)	26	9	51	0.73 (0.414)
	13-16yr	34.81 (12.2)	35.50	9	55	-0.45 (0.414)
SI	6-8 yr	8.91 (3.05)	8	2	16	0.15 (0.472)
	9-12 yr	15.94 (7.08)	14	7	34	0.99 (0.414)
	13-16yr	21.41 (7.88)	21.50	6	38	0.13 (0.414)
VC	6-8 yr	17.08 (2.55)	18	12	20	-0.48 (0.472)
	9-12yr	25.88 (8.40)	23	16	55	1.50 (0.414)
	13-16yr	37.16 (10.2)	37	18	60	0.17 (0.414)
LNS	6-8 yr	11.88 (3.93)	11	5	18	0.09 (0.472)
	9-12 yr	16.69 (3.64)	17	7	21	-1.03 (0.414)
	13-16 yr	19.25 (3.38)	20	9	26	-0.92 (0.414)
MR	6-8 yr	11.21 (4.17)	11	6	23	1.03 (0.472)
	9-12 yr	16.72 (4.80)	16	10	27	0.45 (0.414)
	13-16 yr	18.91 (5.21)	20	8	27	-0.47 (0.414)

IN	6-8 yr	10.92 (1.82)	11.50	7	14	-0.48 (0.472)
	9-12 yr	16.31 (3.80)	15	11	27	1.05 (0.414)
	13-16 yr	19.69 (4.83)	19	10	29	0.25 (0.414)
AR	6-8 yr	15.13 (3.75)	15.50	9	22	-0.07 (0.472)
	9-12 yr	21.91 (3.16)	22	4	17	-0.37 (0.414)
	13-16yr	23.63 (3.61)	23.50	16	30	-0.40 (0.414)

Note. BD= Block Design; SI= Similarities; VC= Vocabulary; LNS= Letter-Number-Sequencing; MR= Matrix Reasoning; IN= Information; AR= Arithmetic; n = 24 for 6-8 years age group; n = 32 for 9-12 years, and 13-16 years age groups.

Table 4 shows means, standard deviations, and range of scores on the subtests of WISC-IV along with the skew statistics (and associated standard error of skewness) for the three age group of 6-8 years, 9-12 years, and 13-16 years. Statistics for only those seven subtests that have shown non-normal distribution for the overall sample have been tabulated. For the age group of 6-8 years results indicate significantly skewed score distribution for Block Design and Matrix Reasoning subtests, while for 9-12 years age group significant skew is indicated for score distributions of Similarities, Vocabulary, LNS, and Information subtest. Whereas for 13- 16 years old age group skew statistics and Z-skewness score suggests significant non-normal (negatively skewed) distribution for only LNS subtest. See Appendix H1 - H7 for Schematic plots of these subtests to have a visual comparison of age groups based score distribution for these subtests of WISC-IV.

Reliability Analysis for Adapted WISC-IV Subtests. Reliability coefficient is considered as most important test statistics for any test/subtest. It refers to the extent to which the test is internally consistent and is likely to produce consistent results. Alpha and split-half coefficients of all the subtests of adapted WISC-IV were calculated for the whole sample and for the three age groups of 6-8 years, 9-12 years, and 13-16 years.

Table 5

Reliability Coefficients of the WISC-IV Subtests for the Full Sample and for the Three Age Groups

Subtests	Reliability Coefficients			
	Full Sample (n=88)	6 – 8 years (n= 24)	9 – 12 years (n= 32)	13–16 years (n= 32)
Block Design	**0.74**	**0.66**	**0.75**	**0.80**
	0.83	0.79	0.83	0.87
Similarities	**0.83**	**0.63**	**0.89**	**0.89**
	0.91	0.68	0.94	0.96
Digit Span	**0.70**	**0.65**	**0.68**	**0.75**
(Forward)	0.89	0.86	0.87	0.93
Digit Span	**0.65**	**0.57**	**0.68**	**0.70**
(Backward)	0.83	0.76	0.88	0.82
Picture Concept	**0.71**	**0.66**	**0.74**	**0.72**
	0.75	0.72	0.76	0.76
Vocabulary	**0.82**	**0.51**	**0.90**	**0.89**
	0.83	0.61	0.92	0.86
Letter-Number	**0.72**	**0.74**	**0.73**	**0.68**
-Sequencing	0.89	0.91	0.88	0.87
Matrix Reasoning	**0.85**	**0.82**	**0.83**	**0.88**
	0.90	0.91	0.89	0.91
Comprehension	**0.74**	**0.52**	**0.82**	**0.87**
	0.82	0.78	0.82	0.85
Picture	**0.76**	**0.71**	**0.76**	**0.79**
Completion	0.82	0.80	0.82	0.84
Cancellation	**0.92**	-	-	-
	0.92	-	-	-
Information	**0.79**	**0.58**	**0.84**	**0.87**
	0.89	0.86	0.86	0.94
Arithmetic	**0.78**	**0.80**	**0.76**	**0.79**
	0.80	0.86	0.77	0.76
Word Reasoning	**0.66**	**0.47**	**0.73**	**0.73**
	0.68	0.51	0.64	0.82

Note. Alpha and Split-half reliability coefficients are displayed for every subtest where Alpha coefficients are in boldface.

Table 5 presents internal consistency coefficients of all the subtests of adapted WISC-IV for the full sample and for the three age groups of 6-8 years, 9- 12 years, and 13-16 years. For the full sample, split-half reliability coefficients of all the subtests are in acceptable range (i.e. $r > 0.75$) except for Word Reasoning subtest. Alpha coefficients (in boldface) of all subtests also seem to have moderate to high values for the full sample (i.e. $\alpha > 0.70$) except for Word Reasoning and Digit Span (Backward). Similarly, for the age group of 9- 12 and 13- 16 split-half reliabilities for all subtests have moderately high coefficients ($r > 0.75$) with the exception of Word Reasoning again. Alpha Coefficients (in boldface) for all subtests also seem moderate for the above mentioned age groups (i.e. $\alpha > 0.70$) except for LNS (for 13-16 years) and Digit Span (for 9-12 years).

Most adapted WISC-IV subtests indicate good reliabilities for the older age groups but for the younger age group (6-8 years) the table indicates low alpha coefficients ($\alpha < 0.60$) for most of the verbal subtests including Vocabulary, Comprehension, Information, and Word Reasoning; and for Digit Span (Backward). Similarly the split-half reliabilities also exhibit low coefficients for Vocabulary and Word reasoning subtest ($r < 0.65$). Length of the test has a strong influence on reliability coefficients and in case of WISC-IV subtests due to increasing order of difficulty the younger children can attempt much lesser no of items than other two age groups that may result in relatively low reliability coefficients. Moreover, the sample size is also smaller for younger group than the other two age groups.

Beside these reasons, the low reliabilities for these subtests may be resulting from presence of weak items having inappropriate difficulty level or low discriminatory power and/or poor item fit. So in order to explore these item statistics detailed item analyses based on both Classical Test Theory (CTT) and Item Response Theory (IRT) was conducted.

Item Analysis of WISC-IV ^{PAK} Subtests. Item analysis is a process which examines student responses to individual test items in order to assess the quality of those items and of the test as a whole. Item analysis is especially valuable in improving items, but it can also be used to eliminate ambiguous or misleading items in a single test administration.

Classical Test Theory (CTT) Based Item Analysis. CTT based item analysis has circular dependency so both the person statistic (i.e., observed score) is (item) sample dependent, and the item statistics (i.e., item difficulty and item discrimination) are (examinee) sample dependent. Two item statistics, item difficulty and item discrimination were assessed for the whole tryout II sample and for the three age groups of 6-8 years, 9-12 years, and 13-16 years.

Item Difficulty (p-value or pv) is the passing rate or proportion of the students that correctly answered the item (Zimmaro, 2003) and it ranges from 0 to 1. For any item with dichotomous response categories item mean (MN) represents the item difficulty, but for polytomous items with partial crediting scoring system proportion of the students responding to the item correctly as getting a credit of 1 or 2 represents p-value (pv). Accordingly, "pv0 / MN0", "pv1 / MN1", "pv2 / MN2", and "pv3 / MN3" are the item difficulties of the subtest based on all cases (full sample), age group of 6-8 years, 9-12 years, and 13-16 years, respectively.

Item Discrimination is the point-biserial (PB) relationship between how well students did on the item and their total test score. A highly discriminating item indicates that the students who had high tests scores got the item correct whereas students who had low test scores got the item incorrect (Zimmaro, 2003).Item discrimination also ranges from 0 to 1. Higher the value, higher is the discriminatory power of the item. The corrected item-total correlation coefficient for any item represents item discrimination. In the following analysis "PB0", "PB1","PB2", and "PB13" are the corrected item-total correlation for all cases and for ages 6-8, 9-12, and 13-16 years, respectively.

Item Response Theory (IRT) Based Item Analysis. Beside the high informative value of CTT based item statistics, item misfit diagnosis based on IRT was also conducted as it is not sample dependent. For interpretation of Misfit diagnosis "standardized residuals" and "Mean squares" are important. So Infit Mean-square (IN-MSQ), Outfit Mean-square (OUT-MSQ), Infit Z-standardized (ZSTD), Outfit Z-standardized (ZSTD) were estimated for all the items of the subtests. IRT requires large samples for reliable estimation of its parameters so analysis was run on all cases only (not on age groups).

Misfit is an indicated item that cannot fit into the overall structure of the test (Yu, 2012). Out-fit is an outlier-sensitive fit statistic, while In-fit is an inlier-pattern-sensitive fit statistic. *Mean-square* is the chi-square statistic divided by its degrees of freedom. Consequently, its expected value is close to 1.0. Ideally, values greater than 1.0 (under fit) indicate un-modeled noise or other source of variance in the data, while values less than 1.0 (over fit) indicate that the model predicts the data too well. *Z-Standardized* report the statistical significance (probability) of the chi-square (mean-square) statistics occurring by chance when the data fit the Rasch model. The values reported are unit-normal deviates, in which .05% 2-sided significance corresponds to 1.96. Over fit is reported with negative values. The criterion to judge the Mean-square fit statistics as indicated in WINSTEPS Users Guide and Manual is as following:

Interpretation of parameter-level mean-square fit statistics:	
>2.0	Distorts or degrades the measurement system.
1.5 - 2.0	Unproductive for construction of measurement, but not degrading.
0.5 - 1.5	Productive for measurement.
<0.5	Less productive for measurement, but not degrading. May produce misleadingly good reliabilities and separations.

The general principle to diagnose fit is to investigate outfit before infit, mean-square before t-standardized, high values before low or negative values. Moreover, high outfit mean-squares may be the result of a few random responses by low performers (outliers), while high infit mean-squares indicate that the items are mis-performing for the people on whom the items are targeted, which is a bigger threat to validity (WINSTEPS User Guide). Similarly it has also been stated that Zstd (standardized residual) is an index of model fit. Model fit takes the overall structure into consideration. If we remove some "misfit" items and re-run the IRT analysis the distribution will look more normal, but there will still be items with high residuals. Due to having verbal content, many items of its subtests got adaptation. Because of this, the "model fit" approach is not a good way to examine item fit, so 'Mean squares' should be emphasized as index of item fit (Yu, 2012).

Item Analysis for Block Design Subtest. Block design is a 'Perceptual Reasoning Index' subtest in which examinee is required to replicate a set of modeled or printed two dimensional geometric patterns using red and white blocks within a specified time limit (in increasing difficulty level of items).

Table 6

Item Difficulties (MN) and Corrected Item-total Correlation Coefficients (PB) of the Block Design Subtest Items for all the Cases and for the Three Age Groups

Obs	NAME	MN0	PB0	MN1	PB1	MN2	PB2	MN3	PB3
1	BD_1	1.99	0.07	2.00	.	1.97	0.16	2.00	.
2	BD_2	1.95	0.01	1.92	-.17	1.97	0.16	1.97	-.18
3	BD_3	1.93	0.02	1.96	-.23	1.97	0.31	1.88	0.19
4	BD_4	3.59	0.38	2.50	0.42	4.00	.	4.00	.
5	BD_5	3.18	0.46	2.33	0.32	3.25	0.36	3.75	0.44
6	BD_6	2.82	0.59	1.50	0.48	3.25	0.36	3.38	0.70
7	BD_7	2.41	0.65	0.67	0.51	2.75	0.47	3.38	0.56
8	BD_8	2.23	0.61	0.50	0.58	2.63	0.46	3.13	0.44
9	BD_9	1.84	0.70	0.17	0.56	1.56	0.57	3.38	0.58
10	BD_10	1.55	0.73	0.17	0.56	0.88	0.66	3.25	0.63
11	BD_11	0.76	0.61	0.00	.	0.13	0.24	1.97	0.68
12	BD_12	0.66	0.59	0.00	.	0.16	0.47	1.66	0.60
13	BD_13	0.30	0.41	0.00	.	0.13	0.49	0.69	0.35
14	BD_14	0.19	0.34	0.00	.	0.13	0.49	0.41	0.27

Note. MNO = item means for all cases (*n*=88); MN1 = item means for 6-8 years (*n*=24); MN2 = item means for 9-12 years (*n*=32); MN3 = item means for 13-16 years (*n*=32); PB0= corrected item-totals for all cases; PB1= corrected item-totals for 6-8 years; PB2= corrected item-totals for 9-12 years; PB3= corrected item-totals for 13- 16 years.

Table 6 present item means and corrected item-total coefficients for all cases and for the three age groups of 6-8 years, 9-12 years, and 13-16 years. Item means are not much meaningful as for block design the difficulty of item is determined by the provided completion time limit and number of blocks to be used. Considering all cases, PB0 or corrected item-total coefficients of item 3 to item 14 range from 0.34 to 0.73 indicating good discriminatory power of the items. Zimmaro (2003) has indicated that items having discrimination index > 0.2 can be considered as fairly good items. The indicated low discrimination index for item no. 1, 2 and 3 is may be due to less variability of responses to these items as the basal age items are needed to be correctly reproduced for continuation of the subtest. This trend of no, low or negative PB coefficient for initial items is also evident in the three age groups and is accepted due to nature of subtest administration and scoring.

Table 7

Misfit Indices for items of Block Design Subtest of Adapted WISC-IV (N=88)

	Infit		Outfit	
Item	**MNSQ**	**ZSTD**	**MNSQ**	**ZSTD**
1	1.02	0.35	0.82	0.25
2	1.10	0.37	2.51	1,62
3	1.10	0.41	2.70	2.10
4	0.85	-0.26	0.30	-0.23
5	1.13	0.74	1.27	0.65
6	0.88	-0.60	0.37	-0.13
7	1.00	0.09	0.89	0.39
8	1.21	0.90	9.90	5.52
9	1.14	0.74	0.80	0.32
10	0.75	-1.17	0.48	0.01
11	0.79	-0.68	0.25	-0.32
12	0.67	-1.25	0.16	-0.49
13	0.97	0.00	0.19	-0.43
14	1.03	0.21	0.17	-0.47

Note. MNSQ = Item Mean Squares; ZSTD = Standardized Residuals

All the four fit indices of IRT (Rasch analysis) are shown in Table 7. Considering Infit Mean squares all the values fall in acceptable range (Productive items MNSQ = 0.5-1.5 as cited in WINSTEPS User Guide) indicating all items having good fit. Considering outfit, Mean Squares values for item 2, 3, and 8 are exceeding the acceptable range. But as indicated before outfit index is not considered a big threat to item validity as high outfit mean squares results from few random responses of low performing examinees (out-liers). Removing those random responses might improve the item fit (WINSTEPS User Guide). So considering the nature of the items and subtest this high outfit index may be ignored. Over all, item analysis did not lead to any change in item content or item order and subtest is considered ready for standardization.

Item Analysis of Similarities Subtest. Similarities is a 'Verbal Comprehension Index' subtest and examinee is required to describe how two words having common objects or concepts are similar. Items are required to be administered in increasing order of difficulty. This subtest involved translation and adaptation so rigorous item analysis is required.

Table 8

Item Difficulties (pv) and Corrected Item-total Correlation Coefficients (PB) of the Similarities Subtest Items for all the Cases and for the Three Age Groups

Obs.	Name	pv0	PB0	pv1	PB1	pv2	PB2	pv3	PB3
1	SI_1	0.92	0.26	0.75	0.07	0.97	0.15	1.00	.
2	SI_2	0.97	0.19	0.91	0.31	1.00	.	1.00	.
3	SI_3	0.98	0.30	0.95	0.21	1.00	0.22	1.00	.
4	SI_4	0.96	0.27	1.00	0.50	0.90	0.29	1.00	0.34
5	SI_5	0.87	0.37	0.66	0.54	0.96	0.04	0.93	0.31
6	SI_6	0.92	0.42	0.83	0.25	0.93	0.54	0.96	0.30
7	SI_7	0.71	0.50	0.29	0.38	0.81	0.28	0.93	0.31
8	SI_8	0.36	0.76	0.04	0.44	0.25	0.65	0.71	0.70
9	SI_9	0.64	0.74	0.29	0.28	0.68	0.63	0.87	0.66
10	SI_10	0.54	0.60	0.37	0.34	0.50	0.56	0.71	0.59
11	SI_11	0.30	0.69	0.00	.	0.28	0.56	0.56	0.63
12	SI_12	0.31	0.78	0.00	.	0.34	0.76	0.53	0.75
13	SI_13	0.29	0.75	0.04	0.44	0.34	0.81	0.43	0.71
14	SI_14	0.22	0.64	0.00	.	0.18	0.66	0.43	0.51
15	SI_15	0.25	0.70	0.04	0.44	0.18	0.67	0.46	0.67
16	SI_16	0.14	0.60	0.00	.	0.09	0.58	0.31	0.56
17	SI_17	0.21	0.59	0.00	.	0.15	0.50	0.43	0.52
18	SI_18	0.10	0.51	0.00	.	0.09	0.62	0.18	0.45
19	SI_19	0.14	0.60	0.00	.	0.09	0.63	0.31	0.55
20	SI_20	0.11	0.59	0.00	.	0.12	0.74	0.18	0.52
21	SI_21	0.13	0.62	0.00	.	0.06	0.55	0.31	0.66
22	SI_22	0.05	0.34	0.00	.	0.00	..	0.15	0.34
23	SI_23	0.00	.	0.00	.	0.00	.	0.00	.

Note. pv0 = p-value for all cases (*n*=88); pv1 = p-value for 6-8 years (*n*=24); pv2 = p-value for 9-12 years (*n*=32); pv3 = p-values for 13-16 years (*n*=32); PB0= corrected item-totals for all cases; PB1= corrected item-totals for 6-8 years; PB2= corrected item-totals for 9-12 years; PB3= corrected item-totals for 13- 16 years.

Table 8 shows item difficulty and item discrimination values of similarities subtest items for all cases and for the three age groups. A p-value of 1.00 and 0.00 for basal and ceiling items for any particular age group is accepted because of the administration and scoring rules of the subtest but item difficulties clearly indicate that items need to be reordered according to the "pv0" column. Furthermore, for ages 6-8, item difficulty (pv1) also seems to increase suddenly after item 6 (from .83, to .66 to .37 even if reordered). Considering PB coefficients some of the items seem to have low discrimination index (PB < 0.20) but again this can be expected considering less variability in responses to the basal items and items having high difficulty

Table 9

Misfit Indices for items of Similarities Subtest of Adapted WISC-IV (N=88)

Item	Infit		Outfit	
	MNSQ	ZSTD	MNSQ	ZSTD
1	1.05	0.25	0.60	-0.25
2	0.77	-0.19	0.27	-0.77
3	1.05	0.26	0.55	-0.33
4	1.17	0.55	0.47	-0.37
5	0.88	-0.55	5.10	2.58
6	1.49	2.68	1.49	2.23
7	1.23	1.09	0.86	0.28
8	0.93	-0.24	0.58	-0.64
9	0.80	-1.30	0.70	-1.17
10	1.60	2.70	2.02	1.42
11	1.25	1.19	0.71	-0.35
12	0.75	-1.33	0.42	-1.37
13	0.83	-0.82	0.53	-0.78

14	1.04	0.27	0.67	-0.32
15	0.90	-0.42	0.65	-0.37
16	0.88	-0.45	0.38	-0.58
17	1.25	1.17	0.66	-0.25
18	0.94	-0.08	0.30	-0.67
19	0.82	-0.76	0.53	-0.25
20	0.73	-0.94	0.22	-0.89
21	0.77	-0.94	0.24	-0.81
22	1.01	0.15	0.43	-0.37
23	1.00	0.00	1.00	0.00

Note. MNSQ= Item Mean Squares; ZSTD= Standardized Residuals

Table 9 presenting fit indices for similarities subtest indicates misfit of item 5 and 10 due to having higher Mean Square infit and/or outfit values. These values are exceeding accepted range for productive items (i.e. 0.5 – 1.5) as mentioned in WINSTEPS user guide.

Overall item analysis indicates re-ordering of the subtest items along with re-viewing or improving scoring rules for few items to improve item fit and to attain relatively gradual increase in item difficulty. Test needs to be tried out again after making changes (See List of the Similarities Subtest Re-ordered Items along with the Difficulty Index and Original Order in Appendix I).

Item Analysis of Digit-Span Subtest. Digit-Span Forward and Backward requires examinee to repeat numbers verbatim as stated by the examiner. No change was incorporated in this subtest for adaptation of WISC-IV.

Table 10

Item Difficulties (MN) and Corrected Item-total Correlation Coefficients (PB) of the Digit-Span (Backward) Subtest Items for all the Cases and for the Three Age Groups

Obs.	NAME	MN0	PB0	MN1	PB1	MN2	PB2	MN3	PB3
1	DSB_1	2.00	.	2.00	.	2.00	.	2.00	.
2	DSB_2	1.99	-.09	2.00	.	2.00	.	1.97	-.09
3	DSB_3	1.23	0.57	0.71	0.68	1.34	0.54	1.50	0.44
4	DSB_4	0.93	0.68	0.38	0.61	1.19	0.76	1.09	0.55
5	DSB_5	0.49	0.69	0.13	0.50	0.53	0.65	0.72	0.71
6	DSB_6	0.23	0.57	0.00	.	0.22	0.45	0.41	0.71
7	DSB_7	0.02	0.23	0.00	.	0.03	0.13	0.03	0.31
8	DSB_8	0.01	0.22	0.00	.	0.00	.	0.03	0.31

Note. MNO = item means for all cases (n=88); MN1 = item means for 6-8 years (n=24); MN2 = item means for 9-12 years (n=32); MN3 = item means for 13-16 years (n=32); PB0= corrected item-totals for all cases; PB1= corrected item-totals for 6-8 years; PB2= corrected item-totals for 9-12 years; PB3= corrected item-totals for 13- 16 years.

Table 11

Item Difficulties (MN) and Corrected Item-total Correlation Coefficients (PB) of the Digit-Span (Forward) Subtest Items for all the Cases and for the Three Age Groups

Obs.	NAME	MN0	PB0	MN1	PB1	MN2	PB2	MN3	PB3
1	DSF_1	2.00	.	2.00	.	2.00	.	2.00	.
2	DSF_2	2.00	.	2.00	.	2.00	.	2.00	.
3	DSF_3	1.93	0.16	1.88	0.31	1.97	0.11	1.94	0.02
4	DSF_4	1.65	0.54	1.46	0.63	1.72	0.34	1.72	0.63
5	DSF_5	1.30	0.64	0.96	0.61	1.28	0.67	1.56	0.57
6	DSF_6	0.83	0.66	0.38	0.59	0.91	0.60	1.09	0.67
7	DSF_7	0.31	0.59	0.08	0.34	0.25	0.54	0.53	0.65
8	DSF_8	0.18	0.49	0.00	.	0.16	0.47	0.34	0.55

Note. MNO = item means for all cases (n=88); MN1 = item means for 6-8 years (n=24); MN2 = item means for 9-12 years (n=32); MN3 = item means for 13-16 years (n=32); PB0= corrected item-totals for all cases; PB1= corrected item-totals for 6-8 years; PB2= corrected item-totals for 9-12 years; PB3= corrected item-totals for 13- 16 years.

Table 10 and 11 show item means and corrected item-totals for the digit-span subtest. Item means are again not meaningful for digit-span backward and forward items as the item order and item difficulties are determined on the bases of increase in number of digits to be repeated by the examinee. Item-total coefficients for all cases as well as for the three age groups indicate satisfactory discriminatory power of the items except basal items and items having high difficulty for any particular age group.

Misfit diagnosis was not required due to nature of items and its scoring rules. So in conclusion Digit-span is also ready for standardization.

Item Analysis of Picture Concept Subtest. Picture Concept is another subtest of 'Perceptual Reasoning Index' and it requires examinee to choose one picture, from among two or three rows of pictures presented, to form a group with a common characteristic. This subtest also requires item administration in increasing order of difficulty but do not include any adaptive changes for adapted WISC-IV.

Table 12

Item Difficulties (MN) and Corrected Item-total Correlation Coefficients (PB) of the Block Design Subtest Items for all the Cases and for the Three Age Groups

Obs	NAME	MN0	PB0	MN1	PB1	MN2	PB2	MN3	PB3
1	PCn_1	1.00	.	1.00	.	1.00	.	1.00	.
2	PCn_2	1.00	.	1.00	.	1.00	.	1.00	.
3	PCn_3	1.00	.	1.00	.	1.00	.	1.00	.
4	PCn_4	0.90	0.26	0.63	-.07	1.00	.	1.00	.
5	PCn_5	0.98	0.17	0.92	0.07	1.00	.	1.00	.
6	PCn_6	0.90	0.31	0.79	0.13	0.88	0.42	1.00	.
7	PCn_7	0.84	0.37	0.71	0.40	0.84	0.31	0.94	0.18
8	PCn_8	0.81	0.27	0.58	0.26	0.91	-.06	0.88	0.12
9	PCn_9	0.77	0.32	0.63	0.33	0.88	0.19	0.78	0.28
10	PCn_10	0.86	0.43	0.63	0.43	0.97	-.03	0.94	0.36
11	PCn_11	0.78	0.35	0.58	0.42	0.88	0.29	0.84	0.03
12	PCn_12	0.72	0.40	0.50	0.45	0.72	0.11	0.88	0.38
13	PCn_13	0.84	0.41	0.71	0.55	0.88	0.22	0.91	0.31
14	PCn_14	0.60	0.53	0.29	0.49	0.81	0.20	0.63	0.65
15	PCn_15	0.63	0.41	0.38	0.46	0.66	0.25	0.78	0.20
16	PCn_16	0.20	0.19	0.17	-.03	0.22	0.36	0.22	0.16
17	PCn_17	0.45	0.25	0.38	0.23	0.41	0.18	0.56	0.28
18	PCn_18	0.22	0.49	0.00	.	0.31	0.56	0.28	0.38
19	PCn_19	0.22	0.38	0.08	-.11	0.19	0.56	0.34	0.34

20	PCn_20	0.27	0.47	0.08	0.15	0.31	0.41	0.38	0.52
21	PCn_21	0.13	0.36	0.08	0.15	0.16	0.47	0.13	0.44
22	PCn_22	0.10	0.33	0.00	.	0.03	0.52	0.25	0.25
23	PCn_23	0.05	0.25	0.00	.	0.06	0.50	0.06	0.00
24	PCn_24	0.07	0.36	0.00	.	0.09	0.52	0.09	0.26
25	PCn_25	0.09	0.31	0.00	.	0.16	0.23	0.09	0.42
26	PCn_26	0.02	0.17	0.00	.	0.00	.	0.06	0.22
27	PCn_27	0.00	.	0.00	.	0.00	.	0.00	.
28	PCn_28	0.01	0.28	0.00	.	0.03	0.52	0.00	.

Note. MNO = item means for all cases (n=88); MN1 = item means for 6-8 years (n=24); MN2 = item means for 9-12 years (n=32); MN3 = item means for 13-16 years (n=32); PB0= corrected item-totals for all cases; PB1= corrected item-totals for 6-8 years; PB2= corrected item-totals for 9-12 years; PB3= corrected item-totals for 13- 16 years.

Table 12 has presented item means and corrected item-total coefficients for the item of picture concept subtest. Item difficulty index for all cases as well as for all age groups indicate need for re-ordering of the items. Regarding discrimination index, the PB0 (for all cases) indicate satisfactory coefficients for most items except for few including item no. 5, 16, and 26. But table also indicate several low and negative PB coefficients for the three age groups. Zimmaro (2003) argued that the discrimination index is not always a measure of item quality. There is a variety of reasons an item may have low discriminating power for example extremely difficult or easy items will have low ability to discriminate but such items are often needed to adequately sample content and objectives. Similarly an item may show low discrimination if the test measures many different content areas and cognitive skills.

Table 13

Misfit Indices for items of Picture Concept Subtest of Adapted WISC-IV (N=88)

Item	Infit		Outfit	
	MNSQ	ZSTD	MNSQ	ZSTD
1	1.00	0.00	1.00	0.00
2	1.00	0.00	1.00	0.00
3	1.00	0.00	1.00	0.00
4	1.03	0.21	1.10	0.38
5	1.17	0.46	0.49	-0.19
6	1.11	0.48	0.62	-0.40
7	0.99	-0.01	0.81	-0.23
8	1.14	0.83	1.23	0.64
9	1.05	0.35	1.26	0.74
10	0.82	-0.76	0.66	-0.50
11	0.98	-0.05	1.99	2.07
12	0.99	-0.02	1.01	0.14
13	0.92	-0.37	0.65	-0.62
14	0.82	-1.67	0.76	-0.78
15	0.96	-0.28	1.31	1.03
16	1.21	1.25	2.27	2.24
17	1.28	2.29	1.37	1.39
18	0.82	-1.14	0.55	-1.08
19	0.96	-0.19	1.06	0.29
20	0.87	-0.94	0.76	-0.55
21	0.89	-0.44	1.11	0.38
22	0.92	-0.24	1.02	0.26
23	0.96	0.05	0.62	-0.13
24	0.85	-0.35	0.35	-0.81
25	1.06	0.30	0.60	-0.42

26	1.27	0.59	0.45	-0.22
27	1.00	0.00	1.00	0.00
28	0.51	-0.38	0.03	-1.36

Note. MNSQ= Item Mean Squares; ZSTD= Standardized Residuals

Table 13 show fit indices for picture concept subtest. Infit Mean squares for all the items are in accepted range for productive items (i.e. 0.5 -1.5). Similarly, most items also showed satisfactory outfit Mean squares except for item 11 and 16. Re-ordering of items according to difficulty level may lead to improve the outfit index. Over all, subtest is considered ready for standardization with re-ordering of items (See Appendix I for re-ordered items along with the difficulty index and original order).

No CTT or IRT based item analysis was run for coding subtest. Coding subtest do not imply any dichotomous or partial credit scoring system. Total score is the number of shapes or symbols correctly coded within the given time limit of 120 seconds. Respondent's scores and the gradual increase in score with the increasing age indicate proper functioning of the subtest. Though as noticed in the previous tryout, most of the respondents could not reach even the half of the subtest within the allowed time limit. But this issue may be resolved with the presence of local/national norms for interpreting respondent's performance.

Item Analysis of Vocabulary Subtest. Vocabulary is a 'Verbal Comprehension Index' subtest in which all original items were replaced with new items so again this requires stringent item analysis.

Table 14

Item Difficulties (pv) and Corrected Item-total Correlation Coefficients (PB) of the Vocabulary Subtest Items for all the Cases and for the Three Age Groups

Obs.	Name	pv0	PB0	pv1	PB1	pv2	PB2	pv3	PB3
1	VC_1	1.00	.	1.00	.	1.00	.	1.00	.
2	VC_2	1.00	.	1.00	.	1.00	.	1.00	.
3	VC_3	1.00	.	1.00	.	1.00	.	1.00	.
4	VC_4	1.00	.	1.00	.	1.00	.	1.00	.
5	VC_5	1.00	0.08	1.00	0.27	1.00	-.07	1.00	.
6	VC_6	1.00	0.10	1.00	-.10	1.00	-.07	1.00	.
7	VC_7	1.00	0.16	1.00	0.41	1.00	.	1.00	0.06
8	VC_8	1.00	0.29	1.00	0.28	1.00	0.23	1.00	.
9	VC_9	0.98	0.29	0.95	-.11	1.00	0.19	1.00	.
10	VC_10	1.00	0.30	1.00	0.30	1.00	0.16	1.00	0.39
11	VC_11	0.53	0.46	0.37	0.35	0.46	0.33	0.71	0.25
12	VC_12	0.73	0.60	0.25	0.25	1.84	0.45	1.00	0.07
13	VC_13	0.42	0.74	0.00	.	0.37	0.56	0.78	0.56
14	VC_14	0.14	0.44	0.04	0.08	0.09	0.57	0.28	0.27
15	VC_15	0.82	0.48	0.62	0.52	0.87	0.47	0.93	0.38
16	VC_16	0.51	0.74	0.16	0.13	0.46	0.67	0.81	0.62
17	VC_17	0.28	0.63	0.00	.	0.18	0.49	0.59	0.43
18	VC_18	0.28	0.70	0.00	.	0.28	0.69	0.50	0.62
19	VC_19	0.62	0.68	0.12	0.21	0.75	0.54	0.87	0.53
20	VC_20	0.46	0.76	0.04	0.17	0.40	0.63	0.84	0.57
21	VC_21	0.07	0.48	0.00	.	0.03	0.20	0.18	0.51
22	VC_22	0.28	0.58	0.00	.	0.28	0.69	0.50	0.26
23	VC_23	0.22	0.70	0.00	.	0.21	0.74	0.40	0.63
24	VC_24	0.53	0.76	0.00	.	0.62	0.62	0.84	0.56
25	VC_25	0.00	.	0.00	.	0.00	.	0.00	.
26	VC_26	0.07	0.57	0.00	.	0.03	0.62	0.18	0.60
27	VC_27	0.15	0.52	0.00	.	0.09	0.58	0.34	0.32

28	VC_28	0.18	0.68	0.00	.	0.09	0.56	0.40	0.65
29	VC_29	0.22	0.79	0.00	.	0.06	0.64	0.56	0.79
30	VC_30	0.04	0.38	0.00	.	0.00	.	0.12	0.40
31	VC_31	0.10	0.52	0.00	.	0.06	0.69	0.21	0.40
32	VC_32	0.19	0.71	0.00	.	0.03	0.61	0.50	0.63
33	VC_33	0.15	0.69	0.00	.	0.06	0.69	0.37	0.64
34	VC_34	0.06	0.46	0.00	.	0.00	.	0.18	0.50
35	VC_35	0.09	0.43	0.00	.	0.03	0.62	0.21	0.28
36	VC_36	0.06	0.50	0.00	.	0.03	0.61	0.15	0.45

Note. pv0 = p-value for all cases (n=88); pv1 = p-value for 6-8 years (n=24); pv2 = p-value for 9-12 years (n=32); pv3 = p-values for 13-16 years (n=32); PB0= corrected item-totals for all cases; PB1= corrected item-totals for 6-8 years; PB2= corrected item-totals for 9-12 years; PB3= corrected item-totals for 13- 16 years.

Table 14 indicates p-values and corrected item-totals for items of vocabulary subtest. The pv0 clearly indicate re-ordering of many item, moreover for ages 6-12 years, item difficulty (pv1, pv2) increase suddenly after item 10 (from 1.0, to .63 to .38 for age 6-8). So, more items are required from the range of .40 to .90 in order to compensate for these jumps in difficulty level. Considering discrimination though the "PB0" coefficients seem satisfactory, but "PB1", "PB2", and "PB3" of the CTT results show that many of the corrected item-total correlations are negative or lower than .20.

Table 15

Misfit Indices for items of Vocabulary Subtest of Adapted WISC-IV (N=88)

	Infit		Outfit	
Item	MNSQ	ZSTD	MNSQ	ZSTD
1	1.00	0.00	1.00	0.00
2	1.00	0.00	1.00	0.00
3	1.00	0.00	1.00	0.00

4	1.00	0.00	1.00	0.00
5	1.07	0.31	4.08	2.07
6	1.33	0.78	3.02	1.62
7	0.89	-0.13	3.58	1.89
8	0.93	-0.14	0.50	-0.34
9	1.23	0.79	0.74	-0.02
10	1.27	1.55	1.81	1.33
11	2.21	4.73	2.86	2.98
12	0.98	-0.08	0.94	-0.07
13	0.92	-0.33	0.80	-0.02
14	1.43	1.61	1.78	1.00
15	1.12	0.75	1.66	1.39
16	0.84	-0.85	0.80	-0.34
17	1.34	1.61	0.91	0.15
18	0.87	-0.64	0.76	-0.28
19	0.96	-0.16	0.81	-0.60
20	0.69	-2.16	0.55	-1.80
21	0.89	-0.20	0.33	-0.62
22	1.29	1.49	0.97	0.14
23	0.78	-1.07	0.43	-0.97
24	0.70	-1.67	0.49	-1.38
25	1.00	0.00	1.00	0.00
26	0.64	-1.16	0.16	-0.60
27	1.36	1.46	1.01	0.37
28	0.77	-1.07	0.41	-0.65
29	0.49	-3.04	0.21	-1.18
30	0.80	-0.32	0.45	-0.18
31	1.02	0.16	0.47	-0.19
32	0.77	-1.05	0.31	-0.47
33	0.70	-1.33	0.23	-0.89
34	0.91	-0.10	0.25	-0.82

35	1.28	0.91	0.48	-0.35
36	0.91	-0.15	0.21	-0.53

Note. MNSQ= Item Mean Squares; ZSTD= Standardized Residuals

Showing fit indices of vocabulary subtest item Table 15 indicates that following the criterion mentioned in WINSTEPS user guide most items of vocabulary fall within the productive item range of 0.5 – 1.5 in case of Infit Mean Squares except item 11. But considering outfit Mean Squares almost 17 items are out of the range for productive items. Many of the outfit mean squares even exceed the value of 2 (items that are non-productive but not degrading for the test).

Over all, based on the findings of table 14 and 15 re-ordering of vocabulary items is required along with re-viewing of scoring rules for few items. Results also suggested inclusion of few new items to replace weak items. Poor reliability coefficients for 6-8 years old students in the previous section also support further changes in vocabulary subtest. So, new vocabulary subtest has two additional items which may replace weak items of vocabulary after further analysis (See Appendix I for re-ordered vocabulary subtest items).

Item Analysis for Letter-Number-Sequencing Subtest. In this subtest examinee is read a number and letter sequence and is required to recall numbers in ascending order and letters in descending order. No adaptive changes were incorporated in this performance subtest.

Table 16

Item Difficulties (MN) and Corrected Item-total Correlation Coefficients (PB) of the Letter-Number-Sequencing Subtest Items for all the Cases and for the Three Age Groups

Obs.	NAME	MN0	PB0	MN1	PB1	MN2	PB2	MN3	PB3
1	LN_1	2.95	0.03	2.92	0.24	3.00	.	2.94	-.14
2	LN_2	2.99	0.10	2.96	0.00	3.00	.	3.00	.
3	LN_3	2.64	0.58	2.21	0.55	2.66	0.46	2.94	0.47
4	LN_4	2.26	0.72	1.54	0.79	2.31	0.68	2.75	0.32
5	LN_5	2.22	0.79	1.25	0.88	2.44	0.79	2.72	0.44
6	LN_6	1.99	0.75	0.96	0.76	2.28	0.63	2.47	0.63
7	LN_7	0.92	0.60	0.04	0.12	0.91	0.48	1.59	0.61
8	LN_8	0.25	0.40	0.00	.	0.09	0.22	0.59	0.44
9	LN_9	0.06	0.21	0.00	.	0.00	.	0.16	0.21
10	LN_10	0.03	0.21	0.00	.	0.00	.	0.09	0.29

Note. MNO = item means for all cases (n=88); MN1 = item means for 6-8 years (n=24); MN2 = item means for 9-12 years (n=32); MN3 = item means for 13-16 years (n=32); PB0= corrected item-totals for all cases; PB1= corrected item-totals for 6-8 years; PB2= corrected item-totals for 9-12 years; PB3= corrected item-totals for 13- 16 years.

Table 16 shows item means and corrected item total correlations for all cases and for the three age groups on the subtest of Letter-Number –Sequencing. Item means indicate a gradual decline in correct response rate with increasing length of the items. This subtest is also based on partial crediting scoring system. Respondent gets a score of 3 for correct answer on all three trial of an item, score of 2 or 1 on correctly answering two or one out of three trials for an item, and a score of zero for incorrect answers to all the three trials of an item. Length of every L-N-S proceeding item increases (number of digits and/or letter increases) and length of the item or the number of letters and digits asked to be reproduced in proper sequence determines the difficulty level of the items. So in this case may CTT based item means (MN) or pvs and PBs shown in this table as well as IRT based item fit indices are of little significance.

Table 17

Misfit Indices for items of Letter-Number-Sequencing Subtest of Adapted WISC-IV (N=88)

Item	Infit		Outfit	
	MNSQ	ZSTD	MNSQ	ZSTD
1	1.68	1.22	9.90	9.91
2	1.28	0.58	1.06	0.36
3	1.31	1.26	0.91	0.19
4	0.86	-0.72	2.54	3.06
5	0.52	-2.68	1.70	1.21
6	0.91	-0.38	0.85	-0.19
7	0.66	-2.12	0.70	-0.57
8	0.79	-0.81	0.36	-0.76
9	1.23	0.59	0.96	0.28
10	0.57	-0.50	0.10	-1.19

Note. MNSQ= Item Mean Squares; ZSTD= Standardized Residuals

Misfit indices shown in table 17 indicates item 1 being misfit as both infit and outfit Mean Squares for this item are out of range of the values suggested for productive items by WINSTEPS user guide. This misfit in first item suggests inclusion of practicing items or improvement in instructions for the examinee so that they can understand the task well before responding to the scoring items. Beside that the subtest seems to require no change in items or item order, so it is also ready for standardization.

Item Analysis of Comprehension Subtest. Comprehension is another "Verbal Comprehension Index' subtest that includes adaptive changes in its content for adapted WISC-IV so rigorous item analysis is required to judge item functioning.

Table 18

Item Difficulties (pv) and Corrected Item-total Correlation Coefficients (PB) of the Comprehension Subtest Items for all the Cases and for the Three Age Groups

Obs	NAME	pv0	PB0	pv1	PB1	pv2	PB2	pv3	PB3
1	CO_1	1.00	0.14	1.00	-.08	1.00	.	1.00	.
2	CO_2	0.97	0.40	0.91	0.17	1.00	.	1.00	0.40
3	CO_3	0.81	0.51	0.62	0.12	0.81	0.52	0.96	-.05
4	CO_4	0.88	0.44	0.70	0.22	0.90	0.46	1.00	0.41
5	CO_5	0.87	0.37	0.79	0.07	0.87	0.26	0.93	0.38
6	CO_6	0.67	0.53	0.45	0.23	0.71	0.69	0.78	0.27
7	CO_7	0.90	0.34	0.79	0.13	0.96	0.08	0.93	0.37
8	CO_8	0.89	0.47	0.66	0.29	0.96	0.21	1.00	-.14
9	CO_9	0.69	0.56	0.54	0.30	0.65	0.59	0.84	0.54
10	CO_10	0.86	0.57	0.58	0.28	0.93	0.28	1.00	0.58
11	CO_11	0.75	0.62	0.37	0.23	0.81	0.45	0.96	0.40
12	CO_12	0.57	0.66	0.16	0.15	0.62	0.68	0.84	0.36
13	CO_13	0.48	0.67	0.08	0.18	0.50	0.62	0.78	0.54
14	CO_14	0.34	0.32	0.08	0.32	0.50	0.30	0.37	0.15
15	CO_15	0.63	0.70	0.25	0.39	0.68	0.69	0.87	0.45
16	CO_16	0.14	0.41	0.04	0.12	0.06	0.20	0.31	0.42
17	CO_17	0.06	0.38	0.00	.	0.00	.	0.18	0.43
18	CO_18	0.15	0.47	0.00	.	0.12	0.33	0.31	0.46
19	CO_19	0.14	0.49	0.00	.	0.12	0.47	0.28	0.46
20	CO_20	0.09	0.46	0.00	.	0.06	0.37	0.18	0.51
21	CO_21	0.06	0.40	0.00	.	0.00	.	0.18	0.51

Note. pv0 = p-value for all cases (n=88); pv1 = p-value for 6-8 years (n=24); pv2 = p-value for 9-12 years (n=32); pv3 = p-values for 13-16 years (n=32); PB0= corrected item-totals for all cases; PB1= corrected item-totals for 6-8 years; PB2= corrected item-totals for 9-12 years; PB3= corrected item-totals for 13- 16 years.

Table 18 shows item difficulties (pv) and item discrimination values (PB) for all cases and for the three age groups. Results indicate that items need to be reordered according to the "pv0" column. PB0 indicates satisfactory discrimination values for all cases, but for the three age groups discrimination index indicates problems in few items other than the basal or high difficulty items. For example, "PB1" and "PB2" of the CTT results show that items 5 and 7 function poorly for ages 6-12 and item 3, 8, and 14 function poorly for ages 13-16.

Table 19

Misfit Indices for items of Comprehension Subtest of Adapted WISC-IV (N=88)

	Infit		Outfit	
Item	MNSQ	ZSTD	MNSQ	ZSTD
1	1.03	0.26	0.52	-0.16
2	0.88	-0.35	0.43	-0.82
3	1.10	0.68	2.10	2.16
4	0.97	-0.10	0.52	-0.38
5	1.34	1.93	1.66	1.68
6	1.11	0.79	1.49	2.83
7	1.31	1.70	1.77	1.64
8	1.12	0.83	1.22	1.17
9	1.06	0.49	1.04	0.32
10	0.97	-0.19	0.94	-0.39
11	0.93	-0.41	0.98	-0.04
12	0.81	-1.37	0.77	-1.58
13	0.84	-1.06	0.76	-1.16
14	1.50	2.65	1.36	1.25

15	0.74	-1.99	0.72	-2.18
16	0.96	-0.09	0.84	-0.03
17	0.84	-0.46	0.30	-0.23
18	0.87	-0.49	0.47	-0.95
19	0.74	-1.05	0.40	-1.06
20	0.75	-0.96	0.34	-0.98
21	0.72	-0.63	0.21	-0.98

Note. MNSQ= Item Mean Squares; ZSTD= Standardized Residuals

Table 19 indicates that infit and outfit Mean squares for all items fall with in the accepted range of item fit as suggested in WINSTEP user guide except for item 3 and item 5 (outfit Mean square > 1.5); and item 17 (outfit Mean square < 0.5).

Overall, after reordering of items along with reviewing of scoring rules for items showing poor discrimination, comprehension subtest is considered ready for standardization (see Appendix I for re-ordered comprehension items).

Item Analysis of Matrix Reasoning Subtest. Matrix reasoning subtest requires the examinee to complete the missing portion of picture matrix by selecting one of the five response options. This performance subtest includes no adaptive changes and all original items were retained.

Table 20

Item Difficulties (MN) and Corrected Item-total Correlation Coefficients (PB) of the Matrix Reasoning Subtest Items for all the Cases and for the Three Age Groups

Obs.	Name	MN0	PB0	MN1	PB1	MN2	PB2	MN3	PB3
1	MR_1	1.00	.	1.00	.	1.00	.	1.00	.
2	MR_2	1.00	.	1.00	.	1.00	.	1.00	.

3	MR_3	1.00	.	1.00	.	1.00	.	1.00	.
4	MR_4	0.94	0.32	0.79	0.34	1.00	.	1.00	.
5	MR_5	0.99	0.08	0.96	-0.04	1.00	.	1.00	.
6	MR_6	1.00	.	1.00	.	1.00	.	1.00	.
7	MR_7	0.85	0.27	0.63	0.13	0.94	-0.15	0.94	0.31
8	MR_8	0.68	0.44	0.21	0.10	0.84	-0.08	0.88	0.58
9	MR_9	0.80	0.44	0.63	0.36	0.78	0.33	0.94	0.46
10	MR_10	0.77	0.38	0.54	0.03	0.78	0.24	0.94	0.52
11	MR_11	0.63	0.51	0.25	0.27	0.69	0.43	0.84	0.30
12	MR_12	0.83	0.40	0.63	0.45	0.91	0.08	0.91	0.36
13	MR_13	0.51	0.51	0.29	0.59	0.44	0.25	0.75	0.52
14	MR_14	0.51	0.61	0.38	0.37	0.50	0.75	0.63	0.67
15	MR_15	0.42	0.66	0.17	0.51	0.38	0.54	0.66	0.68
16	MR_16	0.70	0.29	0.54	0.54	0.69	-.07	0.84	0.23
17	MR_17	0.43	0.58	0.25	0.59	0.50	0.65	0.50	0.48
18	MR_18	0.43	0.76	0.13	0.68	0.44	0.62	0.66	0.79
19	MR_19	0.34	0.72	0.08	0.65	0.31	0.68	0.56	0.68
20	MR_20	0.40	0.44	0.17	-0.00	0.47	0.38	0.50	0.51
21	MR_21	0.33	0.59	0.04	0.15	0.38	0.42	0.50	0.69
22	MR_22	0.28	0.56	0.13	0.64	0.22	0.41	0.47	0.55
23	MR_23	0.28	0.54	0.00	.	0.41	0.56	0.38	0.47
24	MR_24	0.17	0.51	0.00	.	0.22	0.55	0.25	0.47
25	MR_25	0.11	0.25	0.04	0.30	0.19	0.44	0.09	0.01
26	MR_26	0.10	0.38	0.04	0.57	0.06	0.27	0.19	0.35
27	MR_27	0.13	0.33	0.13	0.64	0.19	0.48	0.06	0.29
28	MR_28	0.08	0.42	0.00	.	0.06	0.38	0.16	0.46
29	MR_29	0.05	0.18	0.04	0.57	0.06	0.05	0.03	0.22
30	MR_30	0.09	0.36	0.04	0.57	0.13	0.47	0.09	0.20
31	MR_31	0.06	0.23	0.08	0.65	0.06	0.27	0.03	0.18
32	MR_32	0.02	0.15	0.04	0.30	0.00	.	0.03	0.25
33	MR_33	0.05	0.36	0.00	.	0.06	0.41	0.06	0.37
34	MR_34	0.00	.	0.00	.	0.00	.	0.00	.
35	MR_35	0.02	0.16	0.00	.	0.03	0.13	0.03	0.18

Note. MNO = item means for all cases (n=88); MN1 = item means for 6-8 years (n=24); MN2 = item means for 9-12 years (n=32); MN3 = item means for 13-16 years (n=32); PB0= corrected item-totals for all cases; PB1= corrected item-totals for 6-8 years; PB2= corrected item-totals for 9-12 years; PB3= corrected item-totals for 13- 16 years.

Item difficulty values indicated in table 20 suggests re-ordering of items according to the MN0 column of the table. PB coefficients indicated low and negative values for many items other than basal items and items having very high difficulty for any particular age group. For example, item 5 and 32 seems to function poorly for the overall sample; item 7, 8 is functioning for age group of 6-12 years; while item 10, 20, and 21 seem to be working poorly for age group of 6-8 years.

Table 21

Misfit Indices for items of Matrix Reasoning Subtest of Adapted WISC-IV (N=88)

	Infit		Outfit	
Item	MNSQ	ZSTD	MNSQ	ZSTD
1	1.00	0.00	1.00	0.00
2	1.00	0.00	1.00	0.00
3	1.00	0.00	1.00	0.00
4	0.96	0.00	0.30	-0.66
5	1.14	0.44	1.15	0.49
6	1.00	0.00	1.00	0.00
7	1.20	0.93	2.67	1.82
8	1.06	0.47	1.52	1.31
9	1.11	0.65	0.75	-0.26
10	1.27	1.49	1.05	0.28
11	1.06	0.49	1.02	0.18
12	1.03	0.20	1.23	0.54
13	1.08	0.64	1.23	0.84

14	0.94	-0.42	0.77	-0.78
15	0.81	-1.32	0.66	-1.04
16	1.24	1.61	3.66	4.07
17	0.96	-0.23	0.96	-0.01
18	0.61	-3.13	0.44	-2.05
19	0.64	-2.72	0.44	-1.50
20	1.27	1.72	1.18	0.60
21	0.90	-0.63	0.69	-0.62
22	0.93	-0.41	0.70	-0.49
23	0.97	-0.17	0.72	-0.45
24	0.85	-0.86	0.49	-0.77
25	1.23	1.04	1.11	0.39
26	1.01	0.13	0.42	-0.72
27	1.08	0.43	0.87	0.06
28	0.73	-0.98	0.32	-0.85
29	1.06	0.29	1.57	0.82
30	0.98	0.01	0.46	-0.58
31	1.10	0.39	0.65	-0.11
32	0.97	0.17	0.93	0.30
33	0.74	-0.56	0.21	-0.87
34	1.00	0.00	1.00	0.00
35	1.03	0.26	0.49	-0.23

Note. MNSQ= Item Mean Squares; ZSTD= Standardized Residuals

Table 21 present IRT based item fit indices for the items of Matrix Reasoning. Infit Mean Squares for all items indicate satisfactory item fit. But outfit Mean Squares pointed out misfit in few items including item 7, 8, 16, and 29.

Overall, with re-ordering of items according to the item means Matrix Reasoning subtest is also considered ready for standardization (see Appendix I for re-ordered items).

Symbol Search and Cancellation Subtests. Both Symbol Search and cancelation subtests are included in 'Processing Speed Index' and do not follow dichotomous or partial credit scoring system followed for scoring of other subtests of WISC-IV. Both these subtests are timed test and no adaptive changes were made in them for adapted WISC-IV. No item analysis was run for both of these two subtests.

Item Analysis of Picture Completion Subtest. Picture completion is the supplemental subtest of 'Perceptual Reasoning Index'. Examinees are required to identify the missing parts of the pictures within a specific time limit. No change was made in this subtest for the adaptation of WISC-IV in Pakistan.

Table 22

Item Difficulties (MN) and Corrected Item-total Correlation Coefficients (PB) of the Picture Completion Subtest Items for all the Cases and for the Three Age Groups

Item No.	MN0	PB0	MN1	MN2	MN3
1	1.00	-	1.00	1.00	1.00
2	1.00	-	1.00	1.00	1.00
3	0.99	0.03	1.00	1.00	0.97
4	0.92	0.38	0.71	1.00	1.00
5	0.97	0.23	0.92	0.97	1.00
6	0.84	0.41	0.63	0.84	1.00
7	0.91	0.39	0.79	0.91	1.00
8	0.85	0.39	0.63	0.88	1.00
9	0.77	0.56	0.42	0.84	0.97
10	0.91	0.31	0.79	0.97	0.94
11	0.83	0.38	0.63	0.84	0.97
12	0.91	0.23	0.88	0.88	0.97
13	0.80	0.42	0.50	0.84	0.97

14	0.64	0.47	0.29	0.72	0.81
15	0.65	0.24	0.46	0.56	0.88
16	0.68	0.43	0.33	0.81	0.81
17	0.49	0.26	0.29	0.47	0.66
18	0.42	0.39	0.13	0.50	0.56
19	0.45	0.47	0.08	0.56	0.63
20	0.44	0.48	0.13	0.44	0.69
21	0.84	0.59	0.46	1.00	0.97
22	0.56	0.42	0.21	0.72	0.66
23	0.27	0.32	0.08	0.28	0.41
24	0.11	0.33	0.04	0.13	0.16
25	0.31	0.40	0.08	0.50	0.28
26	0.40	0.52	0.25	0.41	0.50
27	0.43	0.40	0.21	0.50	0.53
28	0.15	0.37	0.00	0.28	0.13
29	0.17	0.37	0.04	0.19	0.25
30	0.07	0.27	0.00	0.06	0.13
31	0.24	0.45	0.00	0.28	0.38
32	0.19	0.37	0.13	0.19	0.25
33	0.22	0.47	0.04	0.28	0.28
34	0.20	0.41	0.04	0.31	0.22
35	0.05	0.30	0.00	0.03	0.09
36	0.06	0.24	0.04	0.00	0.13
37	0.06	0.30	0.00	0.09	0.06
38	0.03	0.24	0.00	0.03	0.06

Note. MNO = item means for all cases (n=88); MN1 = item means for 6-8 years (n=24); MN2 = item means for 9-12 years (n=32); MN3 = item means for 13-16 years (n=32); PB0= corrected item-totals for all cases; PB1= corrected item-totals for 6-8 years; PB2= corrected item-totals for 9-12 years; PB3= corrected item-totals for 13- 16 years.

Table 22 has the estimated pv and PB values for all items of picture completion subtest. Item difficulty index suggests re-ordering of many items, while item discrimination values for the overall sample indicates moderate to high discriminatory power of almost all items of picture completion subtest.

Table 23

Misfit Indices for items of Picture Completion Subtest of Adapted WISC-IV (N=88)

	Infit		Outfit	
Item	MNSQ	ZSTD	MNSQ	ZSTD
1	1.00	0.00	1.00	0.00
2	1.00	0.00	1.00	0.00
3	1.09	0.40	1.49	0.76
4	0.93	-0.14	0.41	-0.82
5	1.08	0.33	0.41	-0.32
6	1.02	0.18	0.67	-0.62
7	0.87	-0.41	0.53	-0.63
8	1.09	0.49	0.68	-0.57
9	0.81	-0.17	0.61	-1.00
10	0.94	-0.13	0.95	0.14
11	1.07	0.41	0.80	-0.31
12	1.11	0.49	1.40	0.76
13	0.97	-0.10	1.03	0.22
14	1.00	0.05	0.90	-0.32
15	1.26	1.89	3.16	5.37
16	1.01	0.14	1.06	0.29
17	1.31	2.59	1.32	1.43
18	1.08	0.75	1.13	0.59
19	0.98	-0.18	1.00	0.07
20	0.97	-0.28	0.88	-0.46
21	0.64	-1.99	0.34	-1.72
22	1.09	0.77	1.04	0.24

23	1.13	1.06	1.01	0.16
24	0.92	-0.25	1.31	0.66
25	0.99	-0.08	1.23	0.80
26	0.87	-1.19	0.80	-0.81
27	1.08	0.78	1.01	0.13
28	0.97	-0.10	0.70	-0.44
29	0.96	-0.14	0.85	-0.15
30	1.17	0.58	0.57	-0.35
31	0.93	-0.50	0.70	-0.75
32	0.98	-0.07	0.90	-0.07
33	0.87	-0.82	0.63	-0.86
34	0.94	-0.36	0.78	-0.40
35	0.93	-0.02	0.40	-0.56
36	1.02	0.17	0.97	0.25
37	0.98	0.08	0.47	-0.50
38	1.00	0.17	0.47	-0.33

Note. MNSQ= Item Mean Squares; ZSTD= Standardized Residuals

IRT misfit indices based on infit and outfit Mean Squares shown in Table 23 also suggest satisfactory item fit for all the items except for item 15 having a outfit mean square value that is > 2. So overall, picture completion subtest is decided to be ready for standardization after re-ordering of the items (see Appendix I for re-ordered items of picture completion subtest).

Item Analysis for Information Subtest. It is a supplemental subtest of 'Verbal Comprehension Index'. Few of the information subtests item underwent adaptive changes along with translation of all the items in Urdu that asks for a detailed item analysis for assessing functioning of translated and replaced items.

Table 24

Item Difficulties (MN) and Corrected Item-total Correlation Coefficients (PB) of the Information Subtest Items for all the Cases and for the Three Age Groups

Obs	NAME	MN0	PB0	MN1	PB1	MN2	PB2	MN3	PB3
1	IN_1	1.00	.	1.00	.	1.00	.	1.00	.
2	IN_2	1.00	.	1.00	.	1.00	.	1.00	.
3	IN_3	1.00	.	1.00	.	1.00	.	1.00	.
4	IN_4	1.00	.	1.00	.	1.00	.	1.00	.
5	IN_5	1.00	.	1.00	.	1.00	.	1.00	.
6	IN_6	0.95	0.27	0.88	0.40	1.00	.	0.97	0.33
7	IN_7	0.95	0.31	0.83	0.61	1.00	.	1.00	.
8	IN_8	0.97	0.24	0.92	0.36	1.00	.	0.97	0.33
9	IN_9	0.97	0.21	0.88	0.16	1.00	.	1.00	.
10	IN_10	0.89	0.38	0.67	0.34	0.97	0.11	0.97	0.33
11	IN_11	0.89	0.28	0.71	0.32	0.94	0.17	0.97	-.05
12	IN_12	0.89	0.33	0.71	0.19	0.94	0.20	0.97	0.22
13	IN_13	0.56	0.60	0.13	0.28	0.66	0.48	0.78	0.39
14	IN_14	0.69	0.58	0.13	0.36	0.81	0.31	1.00	.
15	IN_15	0.51	0.67	0.04	0.13	0.53	0.56	0.84	0.40
16	IN_16	0.33	0.63	0.00	.	0.34	0.28	0.56	0.65
17	IN_17	0.45	0.64	0.04	-.10	0.47	0.55	0.75	0.42
18	IN_18	0.19	0.64	0.00	.	0.13	0.46	0.41	0.65
19	IN_19	0.32	0.62	0.00	.	0.38	0.64	0.50	0.44
20	IN_20	0.26	0.66	0.00	.	0.31	0.61	0.41	0.64
21	IN_21	0.19	0.66	0.00	.	0.16	0.73	0.38	0.57
22	IN_22	0.18	0.58	0.00	.	0.06	0.60	0.44	0.44
23	IN_23	0.06	0.26	0.00	.	0.09	0.49	0.06	0.08
24	IN_24	0.24	0.66	0.00	.	0.22	0.55	0.44	0.62
25	IN_25	0.10	0.61	0.00	.	0.03	0.48	0.25	0.69

26	IN_26	0.13	0.55	0.00	.	0.09	0.52	0.25	0.52
27	IN_27	0.06	0.48	0.00	.	0.03	0.33	0.13	0.58
28	IN_28	0.09	0.50	0.00	.	0.09	0.52	0.16	0.52
29	IN_29	0.02	0.27	0.00	.	0.00	.	0.06	0.30
30	IN_30	0.07	0.51	0.00	.	0.00	.	0.19	0.62
31	IN_31	0.08	0.53	0.00	.	0.03	0.23	0.19	0.66
32	IN_32	0.03	0.38	0.00	.	0.03	0.48	0.06	0.38
33	IN_33	0.00	.	0.00	.	0.00	.	0.00	.

Note. MNO = item means for all cases (n=88); MN1 = item means for 6-8 years (n=24); MN2 = item means for 9-12 years (n=32); MN3 = item means for 13-16 years (n=32); PB0= corrected item-totals for all cases; PB1= corrected item-totals for 6-8 years; PB2= corrected item-totals for 9-12 years; PB3= corrected item-totals for 13- 16 years.

Table 24 shows difficulty level in terms of item means and discrimination value in terms of corrected item-total for the information subtest items. MN0 (item means for all cases) suggests changes in order of administration for many information subtest items. Furthermore, even if re-ordered for ages 6-8, item difficulty (MN1) increase suddenly after item 12 (from .67, .71 to .13), this sudden ascending of difficulty is evident in MN0 column too. So more items in .20 to .50 range are needed to have a gradual increase in difficulty level for the subtest. PB0 coefficients are indicated to be in satisfactory range for most of the items, but"PB1", "PB2", and "PB3" of the CTT results indicate poor functioning of many items. For example, item 11 is function poorly for ages 9-12; item 12 is function relatively poorly for all ages; and item 23 is function poorly for ages 13-16.

Table 25

Misfit Indices for items of Information Subtest of Adapted WISC-IV (N=88)

	Infit		Outfit	
Item	MNSQ	ZSTD	MNSQ	ZSTD
1	1.00	0.00	1.00	0.00
2	1.00	0.00	1.00	0.00
3	1.00	0.00	1.00	0.00
4	1.00	0.00	1.00	0.00
5	1.00	0.00	1.00	0.00
6	0.83	-0.37	0.16	-1.30
7	0.54	-1.36	0.09	-1.65
8	0.91	-0.08	0.15	-1.39
9	1.01	0.17	0.29	-0.93
10	0.96	-0.06	0.45	-0.49
11	1.21	0.89	9.90	5.38
12	1.35	1.36	0.85	0.12
13	1.19	1.00	0.92	0.11
14	0.75	-1.26	0.70	-0.24
15	0.87	-0.70	0.55	-0.54
16	1.15	0.85	1.05	0.30
17	1.02	0.14	1.50	0.86
18	0.97	-0.08	0.45	-0.59
19	1.13	0.76	0.86	0.03
20	0.99	0.03	0.65	-0.29
21	0.88	-0.51	0.37	-0.76
22	1.17	0.79	0.56	-0.39
23	1.50	1.47	0.84	0.10
24	1.01	0.14	0.49	-0.57
25	0.63	-1.58	0.17	-1.46

26	1.06	0.33	0.38	-0.78
27	0.70	-0.95	0.14	-1.35
28	0.99	0.04	0.31	-0.92
29	1.07	0.31	0.18	-1.18
30	0.77	-0.75	0.16	-1.37
31	0.75	-0.90	0.19	-1.26
32	0.77	-0.43	0.11	-1.43
33	1.00	0.00	1.00	0.00

Note. MNSQ= Item Mean Squares; ZSTD= Standardized Residuals

Following the criteria of Mean Square values within 0.5 – 1.5 indicating item fit for productive items (as cited in WINSTEPS user guide), Table 25 suggests item 11 as highly misfit item with an outfit Mean Square value of 9.90. Other than that all items are showing satisfactory item fit.

So overall, information subtest is still not recommended to be ready for standardization. Re-ordering of items is needed along with re-viewing of item content or scoring rules for few items. Inclusion of new item(s) for replacing the weak item was also suggested. So new information subtest has re-ordered items along with an added item, while content of three items have under gone slight alterations (see Appendix I for re-ordered items of information subtest along with original order).

Item Analysis of Arithmetic Subtest. It is supplemental subtest of 'Working Memory Index'. Adaptive changes in this subtest include changes in names of persons used in items and replacement of one of the original items along with translation of all the items in Urdu.

Table 26

Item Difficulties (MN) and Corrected Item-total Correlation Coefficients (PB) of the Arithmetic Subtest Items for all the Cases and for the Three Age Groups

Obs	NAME	MN0	PB0	MN1	PB1	MN2	PB2	MN3	PB3
1	AR_1	1.00	.	1.00	.	1.00	.	1.00	.
2	AR_2	1.00	.	1.00	.	1.00	.	1.00	.
3	AR_3	1.00	.	1.00	.	1.00	.	1.00	.
4	AR_4	1.00	.	1.00	.	1.00	.	1.00	.
5	AR_5	0.99	0.15	0.96	0.07	1.00	.	1.00	.
6	AR_6	0.97	0.35	0.88	0.38	1.00	.	1.00	.
7	AR_7	1.00	.	1.00	.	1.00	.	1.00	.
8	AR_8	0.99	0.24	0.96	0.30	1.00	.	1.00	.
9	AR_9	0.99	0.21	0.96	0.38	1.00	.	1.00	.
10	AR_10	0.97	0.35	0.88	0.38	1.00	.	1.00	.
11	AR_11	0.93	0.26	0.75	-.12	1.00	.	1.00	.
12	AR_12	0.92	0.49	0.75	0.57	0.97	0.23	1.00	.
13	AR_13	0.92	0.49	0.75	0.70	1.00	.	0.97	0.08
14	AR_14	0.83	0.58	0.54	0.64	0.94	-.00	0.94	0.50
15	AR_15	0.77	0.55	0.38	0.43	0.91	0.28	0.94	0.01
16	AR_16	0.72	0.56	0.46	0.43	0.72	0.50	0.91	0.44
17	AR_17	0.69	0.53	0.33	0.35	0.81	0.20	0.84	0.38
18	AR_18	0.78	0.55	0.46	0.48	0.91	0.47	0.91	0.13
19	AR_19	0.68	0.65	0.25	0.44	0.88	0.29	0.81	0.60
20	AR_20	0.67	0.61	0.29	0.63	0.78	0.55	0.84	0.15
21	AR_21	0.34	0.45	0.04	-.12	0.44	0.48	0.47	0.29
22	AR_22	0.50	0.59	0.08	0.33	0.56	0.23	0.75	0.55
23	AR_23	0.50	0.66	0.08	0.38	0.59	0.67	0.72	0.42
24	AR_24	0.52	0.59	0.25	0.59	0.59	0.54	0.66	0.49
25	AR_25	0.23	0.49	0.00	.	0.22	0.48	0.41	0.41

26	AR_26	0.10	0.34	0.00	.	0.06	0.23	0.22	0.34
27	AR_27	0.28	0.59	0.04	0.34	0.31	0.58	0.44	0.63
28	AR_28	0.08	0.24	0.04	0.34	0.06	-.11	0.13	0.45
29	AR_29	0.13	0.40	0.00	.	0.03	0.24	0.31	0.44
30	AR_30	0.09	0.31	0.00	.	0.06	0.23	0.19	0.29
31	AR_31	0.02	0.25	0.00	.	0.00	.	0.06	0.37
32	AR_32	0.01	0.16	0.00	.	0.00	.	0.03	0.23
33	AR_33	0.03	0.25	0.00	.	0.03	0.18	0.06	0.33
34	AR_34	0.02	0.15	0.00	.	0.03	0.12	0.03	0.17

Note. MNO = item means for all cases (n=88); MN1 = item means for 6-8 years (n=24); MN2 = item means for 9-12 years (n=32); MN3 = item means for 13-16 years (n=32); PB0= corrected item-totals for all cases; PB1= corrected item-totals for 6-8 years; PB2= corrected item-totals for 9-12 years; PB3= corrected item-totals for 13- 16 years.

Table 26 suggests re-ordering of few items of arithmetic subtest based on the MN0 values or item means. Except for the basal items and items with very high difficulty PB0 coefficients indicated satisfactory discriminatory power of all the items, though in case of age groups PB coefficients for few items have low and negative values indicating poor functioning of the item in that particular age group.

Table 27

Misfit Indices for items of Arithmetic Subtest of Adapted WISC-IV (N=88)

	Infit		**Outfit**	
Item	**MNSQ**	**ZSTD**	**MNSQ**	**ZSTD**
1	1.00	0.00	1.00	0.00
2	1.00	0.00	1.00	0.00
3	1.00	0.00	1.00	0.00
4	1.00	0.00	1.00	0.00
5	1.26	0.57	0.46	-0.46
6	0.93	-0.02	0.23	-0.83

7	1.00	0.00	1.00	0.00
8	0.84	0.05	0.05	-1.77
9	1.03	0.31	0.09	-1.53
10	0.93	-0.02	0.23	-0.83
11	1.93	2.01	2.62	1.45
12	0.72	-0.72	0.47	-0.35
13	0.66	-0.96	1.03	0.36
14	0.88	-0.48	0.68	-0.17
15	1.06	0.39	0.97	0.19
16	1.16	0.91	0.83	-0.11
17	1.17	0.99	1.58	1.06
18	1.04	0.27	1.05	0.31
19	0.88	-0.65	0.61	-0.63
20	0.89	-0.62	1.19	0.51
21	1.18	1.19	1.87	1.64
22	0.97	-0.18	0.93	0.01
23	0.79	-1.50	0.73	-0.45
24	1.01	0.14	1.02	0.21
25	0.87	-0.70	0.58	-0.60
26	1.06	0.31	0.51	-0.41
27	0.72	-1.86	0.44	-1.21
28	0.97	-0.02	1.49	0.77
29	0.89	-0.44	0.37	-0.78
30	1.22	0.88	0.54	-0.35
31	0.66	-0.44	0.08	-1.40
32	0.82	0.10	0.08	-1.29
33	0.00	0.00	0.00	0.00
34	0.00	0.00	0.00	0.00

Note. MNSQ= Item Mean Squares; ZSTD= Standardized Residuals

Item fit indices presented in Table 27 indicates poor item fit for item 11 of arithmetic subtest based on infit and outfit Mean Squares. Problems in item fit were also indicated in item 17 and 21 due to higher outfit Mean Square value than 1.5. Overall, with change in order of few items arithmetic subtest is also recommended to be ready for standardization (see Appendix I for re-ordered items).

Item Analysis of Word Reasoning Subtest. Word reasoning is a supplemental subtest of 'Verbal Comprehension Index'. It do not involve any adaptive change other then translation of its items into Urdu.

Table 28

Item Difficulties (MN) and Corrected Item-total Correlation Coefficients (PB) of the Word Reasoning Subtest Items for all the Cases and for the Three Age Groups

Obs	NAME	MN0	PB0	MN1	PB1	MN2	PB2	MN3	PB3
1	WR_1	0.98	0.14	0.96	-.03	0.97	0.24	1.00	.
2	WR_2	0.82	0.34	0.67	-.05	0.84	0.33	0.91	0.44
3	WR_3	0.89	0.38	0.75	0.33	0.94	0.30	0.94	0.43
4	WR_4	0.98	0.04	0.96	0.33	1.00	.	0.97	-.11
5	WR_5	0.75	0.38	0.50	0.21	0.81	0.30	0.88	0.27
6	WR_6	0.98	0.09	0.96	-.26	0.97	0.24	1.00	.
7	WR_7	0.81	0.08	0.71	-.17	0.81	0.20	0.88	-.13
8	WR_8	0.90	0.39	0.83	0.69	0.91	0.40	0.94	0.29
9	WR_9	0.72	0.37	0.46	0.25	0.81	0.13	0.81	0.40
10	WR_10	0.81	0.35	0.67	0.53	0.84	0.30	0.88	0.20
11	WR_11	0.14	0.29	0.04	0.04	0.13	0.19	0.22	0.29
12	WR_12	0.44	0.50	0.13	0.25	0.44	0.41	0.69	0.33
13	WR_13	0.08	0.35	0.00	.	0.09	0.40	0.13	0.27
14	WR_14	0.09	0.37	0.00	.	0.03	0.40	0.22	0.26
15	WR_15	0.20	0.41	0.04	0.29	0.28	0.39	0.25	0.36
16	WR_16	0.18	0.46	0.00	.	0.19	0.51	0.31	0.31
17	WR_17	0.09	0.32	0.00	.	0.06	0.49	0.19	0.12
18	WR_18	0.16	0.49	0.00	.	0.06	0.28	0.38	0.48
19	WR_19	0.10	0.48	0.00	.	0.06	0.28	0.22	0.55
20	WR_20	0.08	0.32	0.00	.	0.06	0.28	0.16	0.25

21	WR_21	0.03	0.29	0.00	.	0.03	0.26	0.06	0.31
22	WR_22	0.00	.	0.00	.	0.00	.	0.00	.
23	WR_23	0.02	0.26	0.00	.	0.00	.	0.06	0.31
24	WR_24	0.02	0.36	0.00	.	0.00	.	0.06	0.50

Note. MNO = item means for all cases (*n*=88); MN1 = item means for 6-8 years (*n*=24); MN2 = item means for 9-12 years (*n*=32); MN3 = item means for 13-16 years (*n*=32); PB0= corrected item-totals for all cases; PB1= corrected item-totals for 6-8 years; PB2= corrected item-totals for 9-12 years; PB3= corrected item-totals for 13- 16 years.

Table 28 also suggested change in order of administration of few items of Word Reasoning subtests according to the MN0 values. Presence of low and negative PB values also show poor item functioning of that item for any particular age group.

Table 29

Misfit Indices for items of Word Reasoning Subtest of Adapted WISC-IV (N=88)

	Infit		Outfit	
Item	MNSQ	ZSTD	MNSQ	ZSTD
1	1.07	0.32	0.40	-0.42
2	1.04	0.30	0.63	-0.48
3	0.81	-0.78	0.49	-0.55
4	0.99	0.18	9.90	3.77
5	1.03	0.22	0.74	-0.41
6	1.10	0.37	0.92	0.26
7	1.50	2.56	2.75	2,24
8	0.69	-1.28	0.72	-0.12
9	1.02	0.17	1.21	0.59
10	0.91	-0.46	1.65	1.12
11	1.15	0.74	1.55	0.90
12	0.93	-0.47	0.82	-0.60
13	1.00	0.09	0.40	-0.59

14	0.97	-0.02	0.53	-0.41
15	1.09	0.57	0.98	0.16
16	0.92	-0.37	0.60	0.57
17	1.11	0.48	0.42	-0.63
18	0.87	-0.64	0.43	-0.91
19	0.78	-0.87	0.29	-0.99
20	1.07	0.32	0.43	-0.55
21	0.89	-0.08	0.26	-0.60
22	1.00	0.00	1.00	0.0
23	0.85	-0.05	0.25	-0.60
24	0.67	-0.42	0.08	-1.11

Note. MNSQ= Item Mean Squares; ZSTD= Standardized Residuals

Table 29 presents IRT based fit indices for items of word reasoning subtest. Infit Mean Squares indicate proper item fit for almost all items following the mean squares with in 0.5 -1.5 (for productive items) criterion as suggested in WINSTEP user guide. But outfit mean squares identified three items with poor item fit according to the above mentioned criterion. Overall, word reasoning subtest is also recommended to be ready for standardization after adjustment of item order for few items.

In conclusion Try-out II analysis resulted in finalization of 12 out of 15 adapted WISC-IV ^{PAK} subtests. So the subtests of Block Design, Digit Span, Picture Concept, Coding, Letter-Number-Sequencing, Matrix Reasoning, Comprehension, Symbol Search, Picture Completion, Cancellation, Arithmetic, and Word Reasoning are decided to be 'ready for standardization'. Out of these 12, six subtests require changes in the order of administration of items. Whereas, three subtests are decided to be 'not ready for standardization' including Similarities, Vocabulary, and Information subtest. After incorporating the suggested changes, these subtests require another item analysis for getting ready for standardization (Lists of re-ordered items along with their p-value and original item order is attached in Appendix I).

All the required order and content changes were incorporated in the Manual and Record Form after consultation with the committee members. A total of three committee approaches were conducted for finalizing the changes. Item administration order was changed after keenly going through the item analysis, item content of two items of information subtest was altered, and 2 and 1-point sampled responses of few items were also re-categorized. Furthermore, two new items in vocabulary subtest and one new item in information subtest were added in order to improve the psychometrics of items and subtests (see Appendix J1 for List of Added Items for Tryout-III and Appendix J2 for Record Form-Urdu for Tryout III).

Phase III: Tryout-III: Finalization of WISC-IV ^{PAK}

This tryout was planned on the bases of results of the Tryout-II and was aimed to finalize the WISC-IV ^{PAK} (Urdu Standardization Edition) for the standardization study.

Sample. For the three subtests that were decided to be 'not ready for standardization' administrations were conducted on a sample of 110 students (n= 10 for each of the 11 age groups). Sample was divided in two groups of male (n=69; 62.7%) and female (n=41; 37.3%) students. Sample was further divided into two groups of P.E. I (n=47, 42.7%) and P.E. II (n=63, 57.3%) based on average Parental Education Level of less than 14 years, and 14 years or more than 14 years respectively. The sample was again selected from randomly selected schools and colleges of capital territory that are under the administration of Federal Directorate of Education, Islamabad (for selection procedure see sample section of Phase II: Tryout II).

On 50 students out of the whole sample for this tryout other six subtests (that are decided to be 'ready for standardization' but have changes in item order) were also administered. This sample was again divided into groups of male (n=25; 50%) and female (n=25; 50%); and groups of P.E. I (n=20; 40%) and P.E. II (n=30; 60%).

Instrument and Materials. The previously described Informed Consent Form and WISC-IV ^{PAK} (Urdu Standardization Edition) kits were used for test administrations. Each kit includes WISC-IV (Urdu Standardization Edition) Administration and Scoring Manual (with incorporated changes); Record Form (Urdu) with changes for new tryout; and Stimulus Booklet. Only 9 out of 15 subtests were used in this study as only these nine subtests have changes in their order of administration or item content. These include Picture Concept, Matrix Reasoning, Comprehension, Picture Completion, Arithmetic, and Word Reasoning subtest with re-ordered items; and Similarities, Vocabulary, and Information subtests with re-ordered and/or modified items.

Procedure. The required adapted subtests of WISC-IV (that have order or content based changes) were administered to the sampled students following all the standardized administration procedures. A team of five trained test examiners was involved in administration and scoring of the subtest. All Record Forms were rechecked by the researcher for scoring and administration errors before the data entry in the SPSS data sheet.

Results and Discussion of Tryout III. This tryout III was conducted on the bases of results of Tryout II. Tryout recommended that three (similarities, vocabulary, and information) out of 15 adapted WISC-IV subtests are still not ready for standardization. It also suggested many changes in these three subtests including changes in item order; improving the scoring rules for few items to reduce its difficulty level; and addition of two items in vocabulary and one item in information subtests for replacing weak items later on.

So, a Tryout III was conducted in order to assess the influence of these changes on the three subtests and to finalize these subtest for standardization study. Tryout III was conducted on a sample of 110 students and compiled data was run through item analysis based on both CTT and IRT parameters.

Univariate Analysis of the Three WISC-IV ^{PAK} Subtests. This analysis explored the mean, median, and standard deviations of the three subtests along with the score distribution on the tryout III sample. This

Table 30

Psychometric Properties of the Score Distribution on Adapted WISC-IV Subtests (N= 110)

Sub-tests (Variables)	*M (SD)*	**Median**	**Range**		**Skew (Std. Error)**
			Mini.	**Max.**	
Similarities	16.52 (7.52)	15	4	38	0.74 (0.230)
Vocabulary	28.20 (10.87)	26	14	64	1.11 (0.230)
Information	17.26 (5.64)	17	8	33	0.63 (0.230)

Table 30 has summarized the psychometric characteristics of the distribution of scores on the three adapted subtests. The skew statistics and the corresponding Z-skewness scores indicate significantly positive skewness on all the three subtest of similarities, vocabulary, and information ($p < 0.01$; $p < 0.001$; and $p < 0.01$ for the three subtests respectively). Box plot present the score distribution visual in terms of median and quartile. Box plots for the three subtests also confirmed the positively skewed score distribution indicating the subtests to be considered as difficult for the current sample (see Appendix K for the Box Plots of Similarities, Vocabulary, and Information Subtests of WISC-IV ^{PAK}).

Reliability Analysis of the Three Subtests. Alpha and Split-half reliability coefficients were estimated for the current sample in order to establish the internal consistency of these subtests.

Table 31

Reliability Coefficients of the WISC-IV Subtests for the Full Sample and for the Three Age Groups

Subtests	**Reliability Coefficients**			
	Full Sample (*n*=110)	**6 – 8 years (*n*= 30)**	**9 –12 years (*n*= 40)**	**13-16 years (*n*= 40)**
Similarities	**0.83**	**0.63**	**0.89**	**0.89**
	0.85	0.64	0.92	0.88
Vocabulary	**0.79**	**0.64**	**0.77**	**0.89**
	0.86	0.80	0.79	0.94
Information	**0.75**	**0.57**	**0.79**	**0.84**
	0.81	0.75	0.85	0.83

Note. Alpha and Split-half reliability coefficients are displayed for every subtest where Alpha coefficients are in boldface.

Reliability coefficients presented in Table 31 indicate that all the tests are internally consistent. Considering full sample, both the alpha (in boldface) and split-half seem to have moderately high coefficients for the three subtests. Reliability coefficients are higher for the age group of 9-12 years and 13 -16 years than for the age group 6-8 years that exhibits relatively low reliability coefficients. A smaller sample and response to lesser number of items than the older age groups may have contributed in this relatively low reliabilities for the 6-8 years age group. Moreover, different item statistics like difficulty and discrimination level can also influence the test reliability.

Item Analysis of Similarities Subtest. Item analysis of similarities subtest with changes in item administration order involves determination of CTT based item statistics and IRT based item fit indices. Analyses were done with the help of various software including IBM SPSS19, SAS (The UNIVARIATE Procedure), and WINSTEPS (Rasch and Partial-credit IRT Models).

Table 32

Item Difficulties (pv) and Corrected Item-total Correlation Coefficients (PB) of the Similarities Subtest Items for all the Cases (N=110) *and for the Age Group of 6-8 years (N*=30*)*

New Order of Administration	pv0	PB0	Pv1	Previous order
1	0.98	0.26	0.96	1
2	0.97	0.19	0.93	2
3	0.99	0.30	1.00	3
4	0.95	0.27	1.00	4
5	0.94	0.37	0.80	5
6	0.84	0.42	0.73	6
7	0.81	0.76	0.66	8
8	0.76	0.50	0.53	7
9	0.69	0.74	0.40	9
10	0.41	0.60	0.20	10
11	0.32	0.78	0.06	12
12	0.29	0.69	0.00	11
13	0.28	0.64	0.10	14
14	0.27	0.75	0.06	13
15	0.20	0.60	0.10	16
16	0.19	0.70	0.03	15
17	0.18	0.60	0.10	19
18	0.16	0.59	0.00	20
19	0.14	0.51	0.03	18
20	0.12	0.59	0.00	17
21	0.05	0.62	0.03	21
22	0.04	0.34	0.00	22
23	0.00	-	0.00	23

Note. pv0 = p-value for all cases; pv1 = p-value for 6-8 years; PB0= corrected item-totals for all cases.

Table 32 shows items arranged into new order of increasing difficulty with their corresponding p-values and corrected item-total coefficients. The pv0 and PB0 represents item difficulty and discrimination for the whole sample, while pv1 represents difficulty level of items for the age group of 6-8 years (item difficulty for this age group is tabulated because item statistics showed maximum problem for this youngest age group in Tryout II). After re-ordering based on tryout III data analysis, item difficulties seem gradual enough both for the whole sample (pv0) and for sub-group including 6-8 years (pv1). Corrected item totals also seem satisfactory as except for the basal items all PB0 values are > 0.3.

Considering IRT analysis, both misfit indices (infit & outfit) indicate good fit of the item as all MNSQ fall within the range of productive items (Mean Square values are within 0.5 – 1.5 as cited in WINSTEPS User Guide). So re-ordering and reviewing of items after tryout II has resulted in improvement of item statistics for the similarities subtest, and it is also ready for standardization now.

Item Analysis of Vocabulary Subtest. Many changes have been incorporated in vocabulary subtest after Tryout II including addition of two items (to replace weak items) and reviewing of item scoring rules. For Tryout III item analyses of the subtest again involve determination of CTT based item statistics like difficulty and discrimination level; and IRT based item fit indices. Analyses were carried out with the help of IBM SPSS (PASW) Statistics 19, SAS (The UNIVARIATE Procedure), and WINSTEPS (for Rasch and Partial-credit IRT Models).

Table 33

Item Difficulties (pv) and Corrected Item-total Correlation Coefficients (PB) of the Vocabulary Subtest Items for all the Cases (N=110) and for the Age Group of 6-8 years (N=30)

New Order of Administration	**pv0**	**PB0**	**Pv1**	**Previous order**
1	1.00	-	1.00	1
2	1.00	-	1.00	2
3	1.00	-	1.00	3
4	1.00	-	1.00	4
5	1.00	0.02	1.00	5
6	1.00	-	1.00	6
7	1.00	-	1.00	7
8	0.97	0.29	0.90	8
9	0.96	0.25	0.93	9
10	0.97	0.19	0.93	10
11	0.81	0.41	0.66	11
12	0.70	0.62	0.30	12
	0.646	0.27	**0.60**	**14a**
13	0.645	0.56	0.10	15
14	0.636	0.59	0.33	13
15	0.55	0.34	0.56	14b
16	0.53	0.66	0.10	16
17	0.47	0.64	0.10	17
18	0.40	0.68	0.00	18
19	0.35	0.73	0.00	19
20	0.29	0.69	0.03	20
21	0.23	0.59	0.03	25
22	0.20	0.67	0.00	24

23	0.19	0.68	0.06	21
24	0.19	0.70	0.00	23
25	0.16	0.61	0.00	22
26	0.15	0.43	0.00	27
27	0.076	0.58	0.00	26
28	0.073	0.53	0.03	28
29	0.073	0.47	0.00	32
30	0.072	0.56	0.00	30
31	0.05	0.61	0.00	31
32	0.05	0.48	0.00	33
33	0.04	0.33	0.00	29
34	0.02	0.48	0.00	34
35	0.02	0.40	0.00	35
	0.02	**0.45**	**0.00**	**36b**
36	0.00	-	0.00	36a

Note. pv0 = p-value for all cases; pv1 = p-value for 6-8 years; PB0= corrected item-totals for all cases.

Table 33 shows items arranged into new order of increasing difficulty with their corresponding p-values and corrected item-total coefficients. The pv0 and PB0 represents item difficulty and discrimination for the whole sample, while pv1 represents difficulty level of items for the age group of 6-8 years (item difficulty for this age group is tabulated because item statistics showed maximum problem for this youngest age group in Tryout II). After re-ordering based on tryout III data analysis, item difficulties of all items seem gradual enough both for the whole sample (pv0) and for sub-group except 6-8 years (pv1). For this youngest group few items still seem to function differently, for example, item 12, 13, 14 and 15 show differences in difficulty level from the whole sample. Corrected item totals seem satisfactory for most items as except for the basal items all PB0 values are > 0.3.

For tryout III, 38 instead of 36 vocabulary items were administered (two new items). After re-arrangement of the items two items with weak item statistics were to be eliminated so that length of the test can be kept as in original. So, in table 33 the items with no item administration number (items in boldface) were later eliminated as having weak item statistics. Expert committee decided for this elimination.

Considering IRT analysis, both misfit indices (infit & outfit) indicate good fit of the item as for most items MNSQ fall within the range of productive items (Mean Square values are within 0.5 – 1.5 as cited in WINSTEPS User Guide). Only three items display MNSQ that are out of this desirable range. Item 15 (previously item 14b) is with infit MNSQ of 1.74 and outfit MNSQ of 2.14; item 23 (previously item 21) have MNSQ of 1.93 and 1.97 for infit and outfit respectively; whereas for item 33 (previously item 30) the infit MNSQ is 2.0. These relatively nonproductive items (though not degrading) were retained after some changes in scoring rules as suggested by the committee. So overall, re-ordering and reviewing of items after tryout II has resulted in improvement of item statistics for the vocabulary subtest, and it is also decided to be ready for standardization.

Item Analysis of Information Subtest. Few changes have been incorporated in information subtest after Tryout II including addition of one items (to replace weaker item), and reviewing of item scoring rules. For Tryout III item analyses of the subtest again involve determination of CTT based item statistics like difficulty (item means) and discrimination level; and IRT based item fit indices. Analyses were carried out with the help of IBM SPSS (PASW) Statistics 19, SAS (The UNIVARIATE Procedure), and WINSTEPS (for Rasch and Partial-credit IRT Models).

Table 34

Item Difficulties (MN) and Corrected Item-total Correlation Coefficients (PB) of the Information Subtest Items for all the Cases (N=110) and for the Age Group of 6-8 years (N=30)

New Order of Administration	**MN0**	**PB0**	**MN1**	**Previous order**
1	1.00	-	1.00	1
2	1.00	-	1.00	2
3	1.00	-	1.00	3
4	1.00	-	1.00	4
5	1.00	-	1.00	5
	1.00	-	**1.00**	**8**
6	0.97	0.22	0.90	9
7	0.95	0.22	0.83	6
8	0.95	0.24	0.83	10
9	0.95	0.28	0.80	7
10	0.89	0.38	0.60	12
11	0.86	0.41	0.57	11
12	0.75	0.57	0.30	13a
13	0.66	0.61	0.23	13b
14	0.55	0.58	0.13	14
15	0.52	0.65	0.10	15
16	0.45	0.60	0.07	16
17	0.32	0.66	0.07	19
18	0.31	0.54	0.00	20
19	0.31	0.69	0.00	23
20	0.28	0.55	0.00	17
21	0.26	0.63	0.00	18
22	0.23	0.61	0.00	22

23	0.17	0.62	0.00	28
24	0.16	0.57	0.00	21
25	0.16	0.64	0.00	26
26	0.14	0.59	0.00	25
27	0.10	0.54	0.00	27
28	0.09	0.51	0.00	24
29	0.09	0.59	0.00	30
30	0.05	0.40	0.00	29
31	0.05	0.43	0.00	31
32	0.03	0.34	0.00	32
33	0.01	0.10	0.00	33

Note. MN0 =Item means or p-value for all cases; MN1 =Item means or p-value for 6-8 years; PB0= corrected item-totals for all cases.

Table 34 shows items arranged into new order of increasing difficulty with their corresponding item means and corrected item-total coefficients. The MN0 and PB0 represent item difficulty and discrimination for the whole sample, while MN1 represents difficulty level of items for the age group of 6-8 years. After re-ordering based on tryout III data analysis, item difficulties seem gradual enough both for the whole sample (MN0) and for sub-group including 6-8 years (MN1). Corrected item totals also seem satisfactory as except for the basal items most PB0 values are in acceptable range.

For tryout III, 34 items were administered in information subtest and the new item seems to perform well. So to retain the original length of the subtest (33 items) one of the initial item (previously item 8) was decided to be eliminated. This item with no new administration order number in the table was decided to get eliminated by the expert committee for being very easy, non-discriminatory, and thus nonproductive.

Considering IRT analysis, both misfit indices (infit & outfit) indicate good fit of the items as all MNSQ fall within the range of productive items (Mean Square values are within 0.5 – 1.5 as cited in WINSTEPS User Guide). So reordering, addition, and reviewing of items after tryout II has resulted in improvement of item statistics for the information subtest, and now it is also ready for standardization.

Reliability Analysis of Subtests with New Administration Order. Twelve out of 15 subtests of WISC-IV Urdu were considered ready for standardization in tryout II. Among them for six subtests reordering of items was also suggested for improving functioning of items. These reordered subtests are Picture Concept, Matrix Reasoning, Comprehension, Picture Completion, Arithmetic, and Word Reasoning. Reliability of these reordered subtests was re-analyzed on a sub-sample of tryout III ($N = 50$; Boys = 25 & Girls = 25).

Table 35

Reliability Coefficients of Reordered WISC-IV ^{PAK} Subtests ($N = 50$)

Subtests	No. of items	*Alpha*
Picture Concept	28	.78
Matrix Reasoning	35	.89
Comprehension	21	.86
Picture Completion	38	.86
Arithmetic	34	.88
Word Reasoning	24	.80

Table 35 presents alpha correlation coefficients for six WISC-IV Urdu subtests that were decided to be ready for standardization in tryout II but have gone through re-ordering of items. Coefficients clearly indicated considerable improvement in reliability of all the subtests in tryout III. Considering picture concept, matrix reasoning, and picture completion, the alpha coefficients have improved from .71, .85, and .76 respectively in tryout II to .78, .89, and .86 respectively in tryout III. Similarly, reliability coefficients for comprehension, arithmetic, and word reasoning subtests have risen from .74, .78, and .66 respectively for tryout II to .86, .88, and .80 respectively for tryout III. So item re-ordering has served its purpose of improving psychometric strength of the subtests.

Tryout III concludes the adaptation process by finalizing all adaptive changes in subtests of WISC-IV ^{PAK} and it is now ready for standardization. Afterwards, all the changes suggested by the tryout III analyses were incorporated in the Manual and Record Form. The changes included reordering of similarities, vocabulary, and information subtest items; and elimination of extra (nonproductive) items from vocabulary and information subtest to retain their original length. The edited and finalized copy (Microsoft Word file) of WISC-IV ^{PAK} Administration and Scoring Manual, and Urdu Record Form was again sent to NCS Pearson, India for final printing of WISC-IV ^{PAK} that would be used in the standardization study.

Completion of WISC-IV adaptation was achieved through a long process implying various procedures. Following international standards, all subtests went through iterative cycles of piloting and modifications initially through judgmental procedures and then through statistical procedures (see for example, International Test Commission, 2012; Malda et al., 2008). Judgmental procedures included use of suitable translation design and translators; various committee approaches, and expert reviews; and cognitive interviews and feedback sessions with the children. Whereas, statistical procedures included use of both classical test theory techniques (CTT) and the modern item response theory (IRT) techniques. CTT focuses more on test level psychometrics like enhancement of reliability, while IRT focuses on evaluating the quality of tests at the item level (Egberink, 2010). So both item and subtest level psychometrics were evaluated to prepare WISC-IV ^{PAK} for standardization. Use of all these procedures was important as psychological tests including intelligence tests (like WISC) are used to make important decisions, so high quality standards for construction, adaptation, and evaluation of these instruments are necessary (Egberink, 2010).

Any psychological test that has been adapted for a specific culture can not be used effectively in that culture unless it is standardized in the local cultural context. Same is true for newly adapted WISC-IV ^{PAK}, so after completion of the adaptation process WISC-IV ^{PAK} entered the process of standardization.

STUDY III- STANDARDIZATION OF ADAPTED WECHSLER INTELLIGENCE SCALE FOR CHILDREN, FOURTH EDITION (WISC-IV ^{PAK})

Study III – Standardization of the Adapted Wechsler Intelligence Scale for Children, 4th Edition- Pakistan (WISC-IV ^{PAK})

The study aimed at standardization of adapted Wechsler Intelligence Scale for Children, 4^{th} Edition- Pakistan (WISC-IV ^{PAK}) on Pakistani population. Standardization process involved derivation of norms along with provision of reliability and validity evidence.

Objectives of the Study III

Main objectives to be fulfilled during this study are as following:

1. To further establish the reliability evidence of the WISC-IV ^{PAK}.
2. To further establish the validity evidence of the WISC-IV ^{PAK}.
3. To develop norms for the adapted subtests of WISC-IV ^{PAK} in Pakistan.
4. To explore the relationship of age, gender, geographical area, and parental education level with children's subtests scores.

These objectives were achieved through a laborious process that was completed in four phases. Each phase with its specific objective itself involved several research steps. For all the steps/phases of standardization process same sample was utilized and considering the primary aim of this study the sample may be called as normative sample. So, for all phases of standardization study analyses were conducted on the normative sample (except for establishment of test-retest stability evidence).

Description of the Normative Sample

The normative information for WISC- IV ^{PAK} is based on a nation-wide sample of 800 children from the age group of 6 years to 16 years and 11 months. The sample is randomly selected from various government, semi-government, and private educational institutes of Pakistan. The sample is stratified on the variables of age, gender, geographic region, and parental education level. The description of stratification variables and characteristics of normative sample is given below.

Age. The total sample of 800 children is divided into 11 age groups of one year each including: 6:00 - 6:11, 7:00 – 7:11, 8:00 – 8:11, 9:00 – 9:11, 10:00 – 10:11, 11:00 – 11:11, 12:00 – 12:11, 13:00 – 13:11, 14:00 – 14:11, 15:00 – 15:11, and 16:00 – 16:11. Each group was composed of 72 or 73 children (for detail see Table 36).

Gender. The normative sample comprised of equal number of male (n = 400) and female (n = 400) children (see detail in Table 36).

Geographic Region. The normative sample (n = 800) was taken from five regions of Pakistan including the four provinces (Punjab, Sindh, Khyber Pukhtunkhwa-KPK, and Balochistan) and Islamabad the capital territory. Geographic region based strata includes: Punjab (n = 450; 56% of total sample); Sindh (n = 190; 23.75% of total sample); KPK (n = 110; 13.75% of total sample); Baluchistan (n = 40; 5% of total sample); and Islamabad (n = 10; 1.25% of total sample). The proportion of the sample taken from different regions of Pakistan can also be seen clearly in Figure 1. The administrative region of Gilgit-Baltistan was not considered in the sampling plan due to threatening condition of peace and security in its various areas.

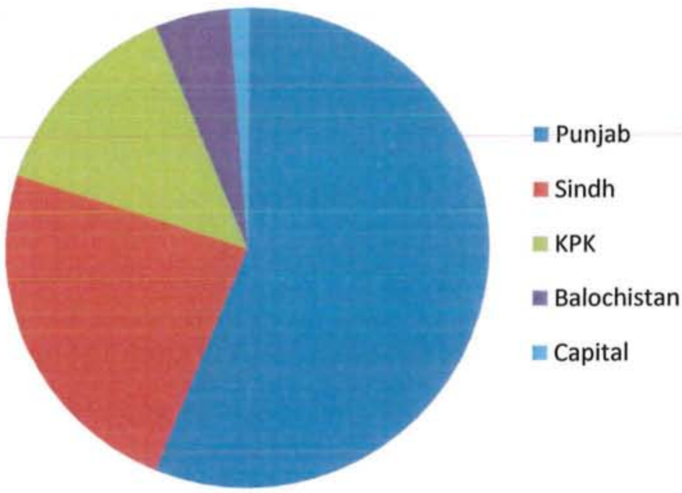


Figure 1 : Normative Sample based on Geographic Region

Normative sample size from these regions has a close correspondence with the total percentage of Pakistani population residing in these areas according to Pakistan Census of 1998 (Pakistan Census Organization). Figure 2 displays this correspondence between sample size and residing population for the five geographical area. Accordingly Punjab having 55.6% of Pakistani population is covered by 56% of total sample; Sindh having 22.99% of Pakistani population is covered by 23.75% of the sample; KPK having 13.4% of Pakistani population is covered by 13.75% of the total sample; Balochistan having 4.96% of Pakistani population is covered by 5% of the total sample; while Capital territory of Islamabad having 0.6% of Pakistani population is covered by 1.25% of the total sample. See Appendix M for **Pakistan's Population by Province/Region since 1951.**

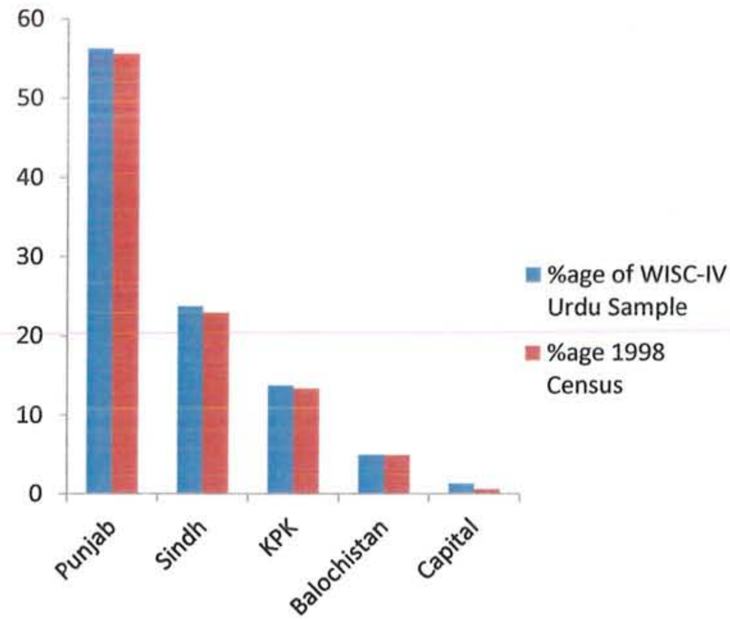


Figure 2: Comparison of Geographic Area based Percentage of WISC-IV Normative Sample with Percentage of Population residing in Geographic Area according to 1998 Census

Within each geographic region data was collected from educational institutes situated in main cities of randomly selected districts. Accordingly, in Punjab data was collected from four districts including: Lahore, Rawalpindi, Multan, and Okara districts. From Sindh data was collected from selected schools of two districts including: Karachi South and Hyderabad districts. In KPK data was collected from selected schools of two districts including: Abbottabad and Kohat districts. In Balochistan data was collected from schools of Quetta District only whereas, in Capital territory data was collected from schools and colleges situated in Islamabad and surrounding semi-urban areas. See Appendix N for List of Selected Schools and Colleges from 9 Randomly Selected Districts of Pakistan along with the Permission Letter for Data Collection and Appendix F1 for List of Randomly Selected Educational Institutes under Administrative Control of Federal Directorate of Education, Pakistan.

Parent Education Level (P.E.). The normative sample was also stratified according to three parent education levels based on average number of years of school completed. Accordingly, P. E. I comprised of children having parents with average education level of less than 10 years (n= 284; 35.5%); P. E. II comprised of children having parents with average education level of less than 14 years (n= 323; 40.37%); and P. E. III comprised of children having parents with average education level of 14 or more than 14 years (n= 193; 24.12%). For WISC-IV (Wechsler, 2003) normative sampling five parental education levels were used for stratification. In Pakistan accurate reporting of parental education is not in practice, even the schools do not have accurate information about student's parental education especially about the education of mothers. So, three parental education levels were decided for stratification in Pakistan.

Characteristics of normative sample along with the subsequent number of children included are given in the Table 36 below. This normative sample was used in various phases of standardization study.

Table 36

Description of Characteristics of Normative Sample (N= 800)

Geographic Region	Districts	Gender		Age Groups in Years											Total
		Male	Female	6	7	8	9	10	11	12	13	14	15	16	
Punjab	Rawalpindi	75	75	14	14	13	14	13	13	13	14	14	14	14	150
	Lahore	60	60	11	11	11	11	11	11	11	10	11	11	11	120
	Okara	40	40	7	7	8	7	7	7	8	8	7	7	7	80
	Multan	50	50	10	9	9	9	9	9	9	9	9	9	9	100
	Sub-total	**225**	**225**	**42**	**41**	**41**	**41**	**40**	**40**	**41**	**41**	**41**	**41**	**41**	**450**
Sindh	Karachi	65	65	11	11	12	12	12	12	12	12	12	12	12	130
	Hyderabad	30	30	6	6	5	5	5	6	6	6	5	5	5	60
	Sub-total	**95**	**95**	**17**	**17**	**17**	**17**	**17**	**18**	**18**	**18**	**17**	**17**	**17**	**190**
KPK	Abbottabad	39	36	8	8	5	9	6	8	6	6	7	6	6	75
	Kohat	16	19	2	2	5	1	4	2	4	4	3	4	4	35
	Sub-total	**55**	**55**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**110**
Balochistan	Quetta	20	20	4	4	4	4	4	4	3	3	4	3	3	40
Capital Territory	Islamabad	5	5	0	1	1	1	1	1	1	1	1	1	1	10
Total		**400**	**400**	73	73	73	73	72	73	73	73	73	72	72	**800**

Table 36 gives a detailed description of the normative sample that was used in the standardization process. As mentioned earlier, in most steps of standardization process this normative sample was used but for establishment of temporal stability (as it involved retesting) a separate relatively small sample was used.

Phase I: Establishment of Reliability Evidence for WISC-IV ^{PAK}

The establishment of reliability evidence for WISC-IV ^{PAK} including temporal stability and internal consistency evidence was the aim of this study. The phase was completed through two steps.

Step I: Establishment of Temporal Stability. This step was concerned with the provision of test-retest stability of WISC-IV ^{PAK} as an important evidence of reliability.

Sample. A total of 34 student participated in the study including both boys (n = 17, 50%) and girls (n = 17, 50%). The sample was further divided into three age groups of 6-8 years (n = 08, 23.5%), 9-12 years (n = 12, 35.3%), and 13-16 years (n = 14, 41.2%). The participants were selected from different education institutes of Rawalpindi and Islamabad including: Islamabad Model School (I-V), G-6/2, Islamabad (n = 6); Army Public School, Askari 10, Rawalpindi (n = 10); Gordon College for Boys, Rawalpindi (n = 8); and Islamabad Model College for Girls, F-6/2, Islamabad (n = 10).

Instruments and Materials. The Informed Consent Form and WISC-IV ^{PAK} (Urdu Standardization Edition) kits were used for test administrations. Each kit includes WISC-IV (Urdu Standardization Edition) Administration and Scoring Manual; Record Form (Urdu); Response Booklet 1 and 2; and a Stimulus Booklet along with block design box and scoring keys.

The WISC-IV ^{PAK} has 10 core and five supplemental subtests that can be summed up into four indices, and one Full Scale IQ. It can be administered on the children from the age range of 6 to 16 years and 11 months. The four composite indices are Verbal Comprehension Index (VCI), the Perceptual Reasoning Index (PRI), the Working Memory Index (WMI), and the Processing Speed Index (PSI).Its Administration and Scoring Manual has all the verbal items and examinee directed instructions in Urdu (for details of the subtests see Study II).

Procedure. After taking permission for test administrations from administrative heads/principals of the concerned schools and colleges, volunteer students of the targeted age groups were selected. Each student was informed individually about the purpose of the research and their consent was taken. After establishing the rapport, the instrument was individually administered to the students in the standardized way. Only age appropriate items were administered except for the cases which require the application of reverse administration rule. After the first administration, same students were contacted again for re-administration of WISC-IV. The test-rest interval was ranging from 21 to 48 days with an average interval of 35 days. All responses of the students were recorded on the Record Form and were carefully scored. The scores taken by the students in both administrations were then analyzed for the estimation of test-retest reliability.

Results of Step I. This step aimed to establish the test-retest reliability evidence of WISC-IV ^{PAK} subtests. For that purpose two sets of data were gathered as test was administered twice to all the participants of the study with an interval of 21 – 48 days between two administrations.

Estimation of Stability Coefficients for all Subtests. Pearson Product Moment Correlations were estimated between the two sets of data for all the 15 subtests along with means and standard deviations. Cohen's *d* was also calculated as an indicator of Standard difference in the mean subtest scores of the two administrations. The analysis was run on the total raw scores of the participants for every subtest.

Table 37

Stability Coefficients of all Subtests of WISC-IV ^{PAK} *along with Means and Standard Deviations for the two testing sessions* ($N = 34$)

	First Testing		Second Testing			
Subtests	*M*	*SD*	*M*	*SD*	*r*	Cohen's *d*
BD	29.76	12.54	33.03	13.45	.91*	0.25
SI	17.71	6.70	19.06	6.97	.77*	0.19
DS	17.91	3.41	18.97	4.19	.84*	0.28
PCn	15.38	3.64	15.74	4.25	.77*	0.09
CD	52.85	13.04	56.35	15.75	.73*	0.24
VC	29.85	8,69	32.15	10.33	.95*	0.24
LN	17.09	4.74	17.32	4.81	.79*	0.04
MR	18.65	5.10	18.85	5.54	.74*	0.03
CO	22.53	7.26	24.18	6.15	.94*	0.24
SS	27.53	6.49	29.68	4.97	.63*	0.37
PCm	22.62	6.05	24.35	5.92	.91*	0.29
CA	74.26	21.54	77.06	17.02	.91*	0.14
IN	17.74	5.06	19.44	5.99	.91*	0.31
AR	21.18	5.25	22.85	6.03	.81*	0.29
WR	12.68	3.43	13.94	3.68	.76*	0.35

Note. BD= Block Design; SI= Similarities; DS= Digit Span; PCn= Picture Concept; CD= Coding; VC= Vocabulary; LN= Letter-Number-Sequencing; MR= Matrix Reasoning; CO= Comprehension; SS= Symbol Search; PCm= Picture Completion; CA= Cancelation; IN= Information; AR= Arithmetic; WR= Word Reasoning. * $p < .001$.

Table 37 shows the stability coefficients of all WISC-IV ^{PAK} subtests. Mean raw scores and standard deviations on all subtests for the two testing sessions along with the standard difference is also indicated. Results indicate that the stability coefficients for all the subtests are significant ($p < .001$) and satisfactory. Among these stability coefficients of Vocabulary, Comprehension, Information, Block Design, and cancellation subtests are excellent (above .90). While the stability coefficients for Digit Span and Arithmetic are also indicated to be good (in the .80s). The stability coefficients for rest of the subtest are also satisfactory (as in the .70s) except Symbol Search having a significant but relatively low correlation coefficient of .63. Overall, the raw scores of participants on the subtests of VCI and PRI seem to be more stable over time than that of WMI and PSI subtests.

The data also indicate that the mean test-retest scores are higher for second administration than the first one for all subtests with mostly small effect sizes ranging from 0.03 for Matrix Reasoning to 0.37 for Symbol Search.

Step II: Establishment of Internal Consistency Evidence. Internal consistency of any measurement tool is must to ensure its reliability. So, the aim of this step was to explore the internal consistency of newly adapted subtests of WISC-IV ^{PAK}.

Sample. As mentioned in the research design and initial standardization study description, the normative sample (N = 800) has been used for establishing internal consistency of the adapted subtests (for details see Table 36 and Description of Normative Sample section of this study). Besides its basic stratification, for comparative analyses sample was further divided into three age groups including: 6-8 years (n = 219, 27.37%), 9-12 years (n = 291, 36.37%), and 13-16 years (n = 290, 36.25%).

Instrument and Materials. The Informed Consent Form and WISC-IV ^{PAK} (Urdu Standardization Edition) kits were used for test administrations (for detail see step I of this phase).

Procedure. Administrative heads/principals of the selected schools and colleges were contacted to schedule the testing sessions and a work plan was developed. On reaching every school, students of the targeted age groups were sampled from different classes before the start of testing sessions. Each student was informed individually about the purpose of the research and their consent was taken. After establishing the rapport, the instrument was individually administered to the students in the standardized way. Only age appropriate items were administered except for the cases which require the application of reverse administration rule as specified in WISC-IV administration and scoring manual. All responses of the students were recorded on the record form and were carefully scored.

Before the start of this study a team of nine females with the minimum qualification of Masters in Psychology were selected and trained for test administrations under the supervision of the researcher. After test administrations and scoring by the team members all the Record Forms went through re-examination by the researcher for possible administration or scoring errors. The scores taken by the students were then entered in the data sheet and were put through multiple analyses for the estimation of internal consistencies and other parameters.

Results of Step II. The step was concerned with the determination of internal consistency of WISC-IV PAK subtests as an evidence of reliability. For that purpose alpha coefficients and item total correlations for all the subtests were computed.

Alpha Coefficients of WISC-IV PAK Subtests. Cronbach Alpha coefficient is considered as most preferred statistic for estimating internal consistency. It is thought of as average of all possible split-half correlations (as corrected by Spearman-Brown formula) for any particular set of items (as cited in Cohen & Swerdlik, 2005). So, alpha coefficients were computed for all the WISC-IV subtests and index scores. The coefficients were estimated for the full sample as well as for the three age groups of 6 – 8 years, 9 – 12 years, and 13 – 16 years. Alpha coefficient is not a good method of estimating internal consistency for speed tests so the coefficients for processing speed index and its constituent subtests were not estimated.

Table 38

Alpha Reliability Coefficients of the WISC-IV ^{PAK} Subtests and Index Scores for the Full Sample and for the Three Age Groups

Subtest/ Index	No. of Items	Alpha Coefficients			
		6 - 8 years (n = 219)	9 - 12 years (n = 291)	13 - 16 years (n = 290)	Full Sample (n = 800)
BD	14	.68	.77	.77	.85
SI	23	.68	.86	.88	.91
DS	16	.66	.76	.82	.82
PCn	28	.80	.82	.74	.86
VC	36	.61	.85	.87	.92
LN	10	.77	.78	.73	.83
MR	35	.81	.85	.87	.90
CO	21	.78	.78	.83	.90
PCm	38	.83	.84	.82	.90
IN	33	.69	.87	.86	.92
AR	34	.85	.78	.78	.91
WR	24	.61	.80	.80	.87
VCI Score	80	.87	.93	.94	.97
PRI Score	77	.82	.86	.85	.91
WMI Score	26	.79	.83	.86	.88

Note. BD= Block Design; SI= Similarities; DS= Digit Span; PCn= Picture Concept; VC= Vocabulary; LN= Letter-Number-Sequencing; MR= Matrix Reasoning; CO= Comprehension; PCm= Picture Completion; IN= Information; AR= Arithmetic; WR= Word Reasoning; VCI Score = Verbal Comprehension Index Score (Sum of raw scores of Similarities, Vocabulary, & Comprehension Subtests); PRI Score = Perceptual Reasoning Index Score (Sum of raw scores of Block Design, Picture Concept, & Matrix Reasoning Subtests); WMI Scores = Working Memory Index Score (Sum of raw scores of Digit Span & Letter-Number-Sequencing subtests).

Table 38 shows the alpha coefficients for Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), and Working Memory Index (WMI) of adapted

WISC-IV along with the coefficients for their constituent subtests. The coefficients indicate moderate (in .60 or .70) to high (in .90s) internal consistency for all the subtests. Whereas, high values of all the coefficients for the index scores indicate high internal consistency of WISC-IV ^{PAK} (from .79 to .97). The coefficients for subtests computed on whole sample range from .82 (for Digit Span) to .92 (for Vocabulary). For the age group of 6-8 years, it ranged from .61 (for Vocabulary) to .85 (for Arithmetic), whereas for the age group of 9-12 years the coefficients are indicated to be from .76 (for Digit Span) to .87 (for Information). For the age group of 13- 16 years the coefficient varies from a moderate level of .73 (for Letter-Number-Sequencing) to a good value of .88 (for similarities).

Overall, the verbal comprehension subtests show higher level of internal consistency than the perceptual reasoning or working memory subtests. Moreover, for the age group of 6-8 years most subtests seem less internally consistent than the older age groups. The standardized administration of subtests includes following of the starting age rule and the discontinuation rules that may result in less number of attempted items on the part of younger age group. This less number of attempted items together with low variability in responses might have resulted in relatively low internal consistency coefficients for 6-8 years old participants.

Item-total Correlations of the WISC-IV ^{PAK} Subtests. Item-total correlations of all the items of the subtests were also computed as another evidence of internal consistency of these subtests. Item total correlation were not computed for coding, symbol search and cancelation subtests due to the nature and format of their items. The tables below display item-total correlation coefficients for eight core subtests of WISC-IV PAK. For the item-total correlation coefficients of the supplemental subtests see Appendix O.

Table 39

Item-total Correlations of Constituent Subtests of Verbal Comprehension Index (VCI) of WISC-IV PAK (N= 800)

	Similarities		Vocabulary		Comprehension	
Item No.	*R*	*p*	*r*	*p*	*r*	*p*
1	.29	.000	-	-	.30	.000
2	.16	.000	.05	.147	.52	.000
3	.46	.000	-	-	.51	.000
4	.32	.000	.04	.230	.58	.000
5	.56	.000	.15	.000	.61	.000
6	.55	.000	.17	.000	.58	.000
7	.70	.000	.16	.000	.61	.000
8	.64	.000	.40	.000	.53	.000
9	.66	.000	.30	.000	.65	.000
10	.78	.000	.35	.000	.66	.000
11	.76	.000	.50	.000	.69	.000
12	.75	.000	.70	.000	.71	.000
13	.74	.000	.71	.000	.73	.000
14	.68	.000	.70	.000	.71	.000
15	.65	.000	.53	.000	.63	.000
16	.72	.000	.74	.000	.61	.000
17	.70	.000	.67	.000	.50	.000
18	.64	.000	.80	.000	.60	.000
19	.60	.000	.75	.000	.50	.000
20	.66	.000	.74	.000	.47	.000
21	.49	.000	.67	.000	.49	.000
22	.40	.000	.70	.000		

23	.27	.000	.60	.000
24			.70	.000
25			.53	.000
26			.60	.000
27			.52	.000
28			.60	.000
29			.52	.000
30			.57	.000
31			.48	.000
32			.53	.000
33			.53	.000
34			.46	.000
35			.35	.000
36			.32	.000

Table 39 presents item-total correlation coefficients for the items of similarities, vocabulary, and comprehension subtests of WISC-IV ^{PAK} (the core subtests that constitute the Verbal Comprehension Index). Results indicate that all the items of similarities and comprehension subtests have significant correlation ($p < .001$) with their total scores that ensure their internal consistency. Same is the case for the subtest of vocabulary as all verbal items (from item no. 5 to 36) display significant item total correlation ($p < .001$) with the total scores. But the initial four items (picture items) of vocabulary subtest seem to have problems. Although item number 2 and 4 show non-significant positive coefficients but for picture item number 1 and 3 correlation coefficient can not even be calculated indicating a low or total lack of variability in the responses of these items. This kind of result can be expected from items on the WISC-IV subtests as for proceeding forward respondents have to take a perfect or non-zero score on the starting items for any age group. The items in these subtests are ordered in a way that starting items have lowest level of difficulty so that most or almost all respondent can respond to them correctly to proceed further.

Table 40

Item-total Correlations of Constituent Subtests of Perceptual Reasoning Index (PRI) of WISC-IV^{PAK} (N= 800)

	Block Design		Picture Concept		Matrix Reasoning	
Item No.	*R*	*p*	*r*	*p*	*r*	*p*
1	.26	.000	.14	.000	-	-
2	.21	.000	.21	.000	.03	.337
3	.25	.000	.35	.000	.13	.000
4	.40	.000	.22	.000	.12	.000
5	.59	.000	.52	.000	.14	.000
6	.65	.000	.47	.000	.22	.000
7	.74	.000	.45	.000	.41	.000
8	.74	.000	.50	.000	.53	.000
9	.79	.000	.58	.000	.36	.000
10	.80	.000	.51	.000	.53	.000
11	.70	.000	.54	.000	.56	.000
12	.68	.000	.57	.000	.56	.000
13	.52	.000	.64	.000	.58	.000
14	.30	.000	.59	.000	.66	.000
15			.58	.000	.67	.000
16			.51	.000	.52	.000
17			.58	.000	.64	.000
18			.52	.000	.73	.000
19			.57	.000	.63	.000
20			.45	.000	.72	.000

21	.45	.000	.63	.000
22	.44	.000	.63	.000
23	.40	.000	.61	.000
24	.37	.000	.56	.000
25	.33	.000	.46	.000
26	.27	.000	.51	.000
27	.28	.000	.44	.000
28	.19	.000	.47	.000
29			.44	.000
30			.45	.000
31			.36	.000
32			.36	.000
33			.30	.000
34			.23	.000
35			.16	.000

Table 40 presents item-total correlations of block design, picture concept, and matrix reasoning subtests (the three core subtests that constitute the Perceptual Reasoning index-PRI). The block design and picture concept subtest show strong internal consistency as having positive and significant ($p < .001$) item-total correlation coefficients for all the items. Similarly, the items of matrix reasoning subtests also have significant correlations ($p < .001$) with the total subtest score except for the two starting items. All the examinees responded correctly to the first item of the test whereas item two also have very low variability in response resulting in having poor correlations with the total scores.

Table 41

Item-total Correlations of Constituent Subtests of Working Memory Index (WMI) of WISC-IV^{PAK} (N= 800)

	Digit Span		Letter-Number-Sequencing	
Item No.	r	p	r	p
1	.06	.097	.24	.000
2	.07	.040	.40	.000
3	.20	.000	.72	.000
4	.51	.000	.79	.000
5	.67	.000	.85	.000
6	.72	.000	.85	.000
7	.71	.000	.77	.000
8	.62	.000	.61	.000
9	.23	.000	.39	.000
10	.23	.000	.25	.000
11	.57	.000		
12	.71	.000		
13	.76	.000		
14	.67	.000		
15	.50	.000		
16	.37	.000		

Table 41 displays internal consistency evidence for digit span and letter-number-sequencing subtests by showing significant correlation ($p < .001$, $p < .05$) for almost all of their items. Only first item of digit span has non-significant coefficients again resulting from low variability of responses on initial items.

Considering the supplemental subtests of WISC-IV ^{PAK}, all word reasoning subtests items demonstrate significant internal consistency evidence, similarly except for few starting items, all items of picture completion, information, and arithmetic subtest also show significant coefficients of item-total correlations (see Appendix O for tabulated item-total correlations of WISC-IV ^{PAK} supplemental subtests).

Phase II: Validation of WISC-IV ^{PAK}

As the title implies, phase II was aimed at validation of WISC-IV ^{PAK}. The validation process included establishment of convergent and discriminant validity evidence; and the cross validation of WISC-IV factorial structure on Pakistan's normative sample. Correspondingly, this phase was carried out through two steps.

Step I: Establishment of Convergent and Discriminant Validity Evidence. The convergent and discriminant validity of WISC-IV ^{PAK} was established following a theoretical methodology presented by Campbell and Fiske (1959), which is known as the Multitrait-Multimethod Matrix method. In this method data support for a set of assumptions developed in a priori about the pattern of relationship of various variables has to be assessed in order to validate the construct or its measure (see Wechsler, 2003).

Presumed Correlational Pattern of WISC-IV ^{PAK} Subtest and Index Scores. For the construct (convergent and discriminant validity) validation various a priori assumptions were made regarding the intercorrelations of WISC-IV ^{PAK} subtests and index score on the bases of WISC-IV structure and previous research evidences.

First, it was proposed that all subtests would show low to moderate correlations with each other. This is based on the assumption that all subtests are measuring a general intelligence factor 'g' (as cited in Wechsler, 2003; Flanagan & Kaufman, 2004).

Second, based on the factor structural researches of WISC-IV (Wechsler, 2003) it was assumed that subtests that belongs to certain scale or index would have higher correlations with each other than with subtests constituting other scales or index.

Third, as previous research evidence shows that certain subtests are more related to general intelligence (e.g. Flanagan & Kaufman, 2004), so it was expected that regardless of their index membership, subtests with high g-loadings would display high correlations with each other than with other subtests.

Fourthly, based on the split loading patterns of certain subtests as indicated in the previous research, it was expected that few subtests like picture completion, picture concept, and Arithmetic would show moderate correlations with VCI and PRI subtests (also see Wechsler, 2003 for bases of these assumptions).

Sample. The same normative sample ($N = 800$) as was used in the previous phase of this study was used for validation of the WISC-IV ^{PAK} (for details see Table 36 and Description of Normative Sample section of this study).

Instrument and Materials. The Informed Consent Form and WISC-IV ^{PAK} (Urdu Standardization Edition) kits were used for test administrations (for detail see step I of phase I).

Procedure. The test administration, scoring, and data recording procedure is same as of Phase I of this study.

Results of the Step I. Step I of phase III was concerned with construct validation of WISC-IV PAK. The multitrait-multimethod matrix methodology (Campbell & Fiske, 1959) was implied to obtain evidence for convergent and discriminant validity of WISC-IV PAK as was implied and reported in WISC-IV Technical and Interpretive Manual (2003). This methodology predicts somewhat high correlations for some variables (convergent evidence) and comparatively low correlations for few other variables (discriminant evidence) when correlational patterns in the matrix are to be examined. Then these predictions in the form of a priori assumptions are tested with the supporting data to establish the construct validity.

Table 42

Intercorrelation Matrix of the Raw Scores for Subtests and Scales/Indexes on WISC-IV PAK (N=800)

Subtests	BD	SI	DS	PCn	CD	VC	LN	MR	CO	SS	PCm	CA	IN	AR	WR	VCI	PRI	WMI	PSI	FS
BD	-																			
SI	.71	-																		
DS	.61	.65	-																	
PCn	.61	.68	.52	-																
CD	.52	.56	.45	.46	-															
VC	.68	.84	.56	.63	.57	-														
LN	.64	.67	.62	.63	.55	.63	-													
MR	.67	.70	.53	.66	.46	.68	.61	-												
CO	.71	.81	.60	.68	.57	.80	.71	.69	-											
SS	.56	.58	.52	.47	.61	.52	.55	.49	.58	-										
PCm	.70	.71	.55	.69	.50	.67	.62	.72	.74	.53	-									
CA	.58	.59	.50	.52	.58	.57	.57	.54	.63	.55	.58	-								
IN	.74	.86	.63	.67	.58	.85	.69	.73	.81	.58	.73	.63	-							
AR	.69	.74	.63	.66	.49	.72	.72	.70	.76	.52	.72	.62	.81	-						
WR	.68	.80	.60	.63	.55	.77	.63	.67	.78	.55	.71	.61	.82	.73	-					
VCI	.74	.94	.64	.70	.61	.95	.71	.74	.92	.59	.75	.63	.90	.79	.84	-				
PRI	.94	.79	.64	.79	.56	.75	.71	.85	.78	.59	.79	.63	.81	.77	.75	.82	-			
WMI	.69	.73	.89	.64	.56	.66	.91	.64	.73	.59	.65	.59	.74	.75	.69	.75	.75	-		
PSI	.59	.63	.53	.52	.95	.61	.61	.52	.63	.83	.56	.63	.64	.56	.60	.67	.63	.63	-	
FS	.84	.89	.71	.75	.76	.79	.79	.78	.88	.74	.78	.70	.88	.81	.82	.94	.91	.84	.83	-
Mean	26.1	16.8	17.4	14.7	46.6	27.6	16.3	16.7	20.1	22.3	20.7	70.0	16.4	21.1	10.9	64.6	57.6	33.6	68.9	224.7
SD	12.6	8.6	4.5	4.5	14.4	11.3	4.9	5.9	22.3	8.5	6.6	22.4	5.9	5.4	4.2	25.9	20.6	8.5	20.5	67.21

Note. BD= Block Design; SI= Similarities; DS= Digit Span; PCn= Picture Concept; CD= Coding; VC= Vocabulary; LN= Letter-Number-Sequencing; MR= Matrix Reasoning; CO= Comprehension; SS= Symbol Search; PCm= Picture Completion; CA= Cancelation; IN= Information; AR= Arithmetic; WR= Word Reasoning; VCI= Verbal Comprehension Index; PRI= Perceptual Reasoning Index; WMI= Working Memory Index; PSI= Processing Sped Index; FS= Full Scale Score; $p < .001$ for all the correlation coefficients.

Table 42 shows a matrix of intercorrelations of all WISC-IV ^{PAK} subtests and scales with each other. Results indicate that all subtests and index scores show at least low to moderate correlation coefficients as the lowest correlation displayed in the matrix is of .46 (between CD & MR; CD & PCn). So the first a priori assumption stands true. It is also evident from the table that the subtests belonging to one scale or index have shown higher correlations with each other than with subtests belonging to other scales or index. The verbal comprehension subtests and index score are showing the highest intercorrelation coefficients ranging from .77 (between VC and WR) to .95 (between VC and VCI). This proves second a priori premise to be true and also gives evidence of convergent and discriminant validity for the WISC-IV ^{PAK}.

The third proposition assumes high correlations among subtests having high g-loadings. Results show that all VCI subtests and index scores have highest correlations (r = .94 between VCI & FS) with the full scale scores (FS) which represents 'g'. PRI subtests and index scores also show high correlations with full scale scores (r = .91 between PRI & FS). Whereas, the PSI subtests and index scores show lowest correlations with the full scale or g-score (r = .83 between PSI & FS). Accordingly, VCI and its subtests seem to have good correlations with PRI (r = .82 between VCI & PRI) and its subtests (r = in .60s or .70 among subtests) which are considered successive to VCI in terms of g-loading. Furthermore, VCI and its subtests seem to have lowest correlations with PSI (r = .67 between VCI & PSI) and its subtests. So these results not only prove third proposition to be true and but also add evidence for the convergent and discriminant validity of WISC-IV ^{PAK}.

The last assumption considers the split loading on the part of few WISC-IV subtests. Results indicate that arithmetic that put heavy demand on auditory comprehension and have relatively high g-loading (relative to other WMI subtests) show high correlations with VCI ($r = .79$) and its subtests (r ranges from .72 to .81). Similarly, picture concept and picture completion have shown almost as high correlation with VCI subtests as with PRI subtests may be due to children's use of verbal mediation to resolve problems. So, all the a priori propositions are found to be true giving strong evidence of WISC-IV ^{PAK}'s construct validity.

Step II: Cross Factorial Validation of WISC-IV ^{PAK}. A four factor structure has been proposed in WISC-IV Technical and Interpretive Manual (Wechsler, 2003). This factor structure is based on theory, research and extensive exploratory factor analyses findings. A replication or cross validation of this four factor structure using WISC-IV ^{PAK} can provide a strong evidence of its structural validity.

Sample. The same normative sample ($N = 800$) as was used in the previous phase of this study was used for cross factorial validation of the WISC-IV ^{PAK} (for details see Table 36 and Description of Normative Sample section of this study). For comparative analysis, sample was further divided into three age groups in this phase of the study including: 6-8 years ($n = 219$, 27.37%), 9-12 years ($n = 291$, 36.37%), and 13-16 years ($n = 290$, 36.25%).

Instrument and Materials. The Informed Consent Form and WISC-IV ^{PAK} (Urdu Standardization Edition) kits were used for test administrations (for detail see step I of phase I).

Procedure. The test administration, scoring, and data recording procedure is same as of Phase I of this study.

Result of Step II. This step is aimed for structural validation of the WISC-IV ^{PAK}. The WISC-IV factor structure is comprised of four factors, the Verbal Comprehension factor, the Perceptual Reasoning factor, the Working Memory factor, and the Processing Speed factor. All WISC-IV subtests are categorized into these four factors. The proposed factor structure is as following:

WISC-IV Factor Structure

Factors			
Verbal Comprehension	**Perceptual Reasoning**	**Working Memory**	**Processing Speed**
Similarities	Block Design	Digit Span	Coding
Vocabulary	Picture Concept	Letter-Number-Sequencing	Symbol Search
Comprehension	Matrix Reasoning	*Arithmetic*	*Cancellation*
Information	*Picture Completion*		
Word Reasoning			

Note. The names of the subtests in italics are the supplemental subtest of the respective factor.

Confirmatory Factor Analysis of WISC-IV ^{PAK}. A series of confirmatory factor analyses (CFA) were run for WISC-IV subtests with the overall sample (N = 800) and with the three age groups of 6-8 years (n = 219), 9-12 years (n = 291), and 13 to 16 years (n = 290). The software IBM AMOS version 19.0 was used to run CFA.

Confirmatory Factor Analysis of Ten Core WISC-IV Subtest. In this set of CFA the factor structure of core WISC-IV subtests was evaluated through different models. The three structural models to be tested for best fit to the data are as following:

1. Null Model: A model with no common latent factors.

2. Model 1: A model with one general latent factor. All the core subtests will be loaded on a general factor.
3. Model 2: A four factor model (as proposed by WISC-IV Technical and Interpretive Manual; Wechsler, 2003). Three Verbal Comprehension subtests on the first factor, 3 Perceptual Reasoning subtest on the second factor, two Working Memory subtests on the third factor, and two Processing Speed subtests on the fourth factor.

Data was analyzed for assessing the basic assumption of 'Structural Equation Modeling (SEM)' including the assumption of multivariate normal distribution. The CFA was conducted using the Maximum Likelihood Estimation (MLE) and the structural models were evaluated through various goodness-of-fit measures. As most of these measures are based on chi-square statistics and its value tends to be higher in large samples. So many goodness-of-fit measures can lead to underestimation of model fit (as cited in Wechsler, 2003). To overcome this problem, many goodness-of-fit measures with less dependency of sample size have been proposed. So the indices like the Goodness of Fit Index (GFI), the Adjusted Goodness-of-Fit Index (AGFI; Joreskog & Sorbom, 1993), the Root Mean Square Error of Approximation (RMSEA; Steiger, 1990), and the Tuker-Lewis Index (TLI; Tucker & Lewis, 1973) were used. These fit indices were used with chi-square and Comparative Fit Index (CFI) for evaluation of the three models.

Table 43

Goodness-of-Fit Statistics for Confirmatory Factor Analysis of Core Subtests for Full sample and for the Three Age Groups

Models	Goodness-of-fit Index			Improvement		Goodness-of-fit Index				
	X^2	*df*	X^2/df	ΔX^2	Δ *df*	GFI	AGFI	CFI	TLI	RMSEA
Age: 6-16										
Null Model	6146.66	45	136.59			.05	.18	.00	.00	.41
Model 1	337.39	35	9.64	5809.27	10	.92	.87	.95	.94	.10
Model 2	135.4	29	4.6	201.99	6	.97	.94	.98	.97	.06
Age: 6-8										
Null Model	668.44	45	14.85			.48	.37	.00	.00	.25
Model 1	128.46	35	3.67	539.98	10	.90	.84	.85	.81	.11
Model 2	73.87	29	2.54	54.6	6	.94	.89	.93	.89	.08
Age: 9-12										
Null Model	1304.18	45	28.98			.34	.19	.00	.00	.31
Model 1	114.35	35	3.27	1189.83	10	.92	.88	.94	.92	.08
Model 2	56.66	29	1.95	57.7	6	.96	.93	.98	.97	.06
Age:13-16										
Null Model	1163.91	45	25.86			.39	.25	.00	.00	.29
Model 1	161.29	35	4.61	1002.62	10	.87	.82	.89	.85	.11
Model 2	49.12	29	1.7	112.17	6	.97	.94	.98	.97	.05

Note. n = 800 for 6 – 16 years; n = 219 for 6-8 years; n = 291 for 9-12 years; n = 290 for 13-16 years; X^2 = Chi-Square value; *df*= Degrees of Freedom; ΔX^2 = Change in Chi-Square Value; GFI= Goodness-of-Fit Index; AGFI= Adjusted Goodness-of-Fit Index; CFI= Comparative Fit Index; TLI= Tucker-Lewis index; RMSEA= Root Mean Square Error of Approximation.

Table 43 shows the CFA Goodness-of-Fit Indices for WISC-IV core subtests. Three models have been tested through MLE for goodness of fit. Considering the overall sample it is quite evident from the results that the Model 2 (four-factor model) fits the data best as compared to Model 1 (one factor model) and the null model. Null model have shown a total mis-fit while, model 1 has shown some improvement considering the chi-square values and other goodness-of –fit indices. But model 2 has shown substantive improvement over model 1 on all indices of goodness-of-fit. Chi-square value has lowered substantively; the GFI and AGFI values are also close to 1.00 indicating a good fit (see Byrne, 2010). Similarly, CFI and TLI values for four-factor model are greater than .95 while for these indices a value close to .95 represents good fit (Hu & Bentler as cited in Byrne, 2010). Furthermore, considering RMSEA any value ≤ .05 shows good model fit, values as high as .08 are also considered as representing reasonable or adequate fit (see Byrne, 2010; Wechsler, 2004), so RMSEA value also indicate adequate fit of the Model 2 to the data.

This result is consistent for the three age groups as all goodness-of –fit indices are indicating best fit for the Model 2 (the four-factor model). Chi-square values for Model 2 has shown substantive improvement and the ratio of chi-square and degrees of freedom are also less than 2.5 signifying a good model fit for all the three age groups. Similarly, the absolute fit indices (GFI & AGFI) and comparative fit indices (CFI) also indicate good model fit for four-factor structure, and same holds true for TLI and RMSEA. The best fitted four-factor structure for the overall sample is presented in the Figure 3 below and for the figural presentation of the four-factor structure for the three age groups see Appendix P1 – P3.

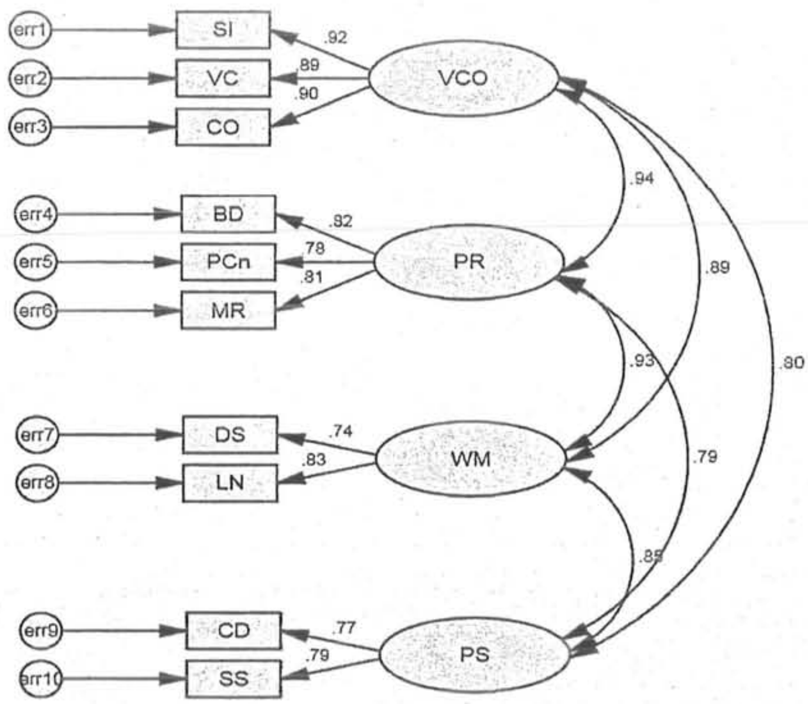


Figure 3: Four-Factor Structure of WISC-IV PAK Core Subtests

The oval ellipse represent latent factor; the rectangles represents the subtests; the small circles represent the error variance

The figure 3 shows the best fitted Four-Factor Structure of WISC-IV PAK core subtests. The similarities, vocabulary, and comprehension subtests are loaded on first factor (Verbal Comprehension- VCO); the block design, picture concept, and matrix reasoning subtests got loaded on second factor (Perceptual Reasoning- PR); the digit span and letter-number-sequencing subtests are loaded on third factor (Working Memory- WM), whereas the coding and symbol search subtests got loaded on fourth factor (Processing Speed- PS). Figure also shows that all the four factors are highly correlating which is well expected as all are measuring one big construct of general intelligence 'g'.

Table 44

Sources of Variance for each Core Subtest in the Four-Factor Model of WISC-IV ^{PAK} (N = 800)

Subtests	VC		PR		WM		PS	
	b	var	*b*	var	*b*	var	*b*	var
SI	.92	85	-	-	-	-	-	-
VC	.89	90	-	-	-	-	-	-
CO	.90	90	-	-	-	-	-	-
BD	-	-	.82	68	-	-	-	-
PCn	-	-	.78	61	-	-	-	-
MR	-	-	.81	66	-	-	-	-
DS	-	-	-	-	.74	55	-	-
LN	-	-	-	-	.83	70	-	-
CD	-	-	-	-	-	-	.77	59
SS	-	-	-	-	-	-	.79	62

Note. *b* = Standardized loading of Subtests on factor; var = Percent variance explained in the Subtest; VC= Verbal Comprehension; PR= Perceptual Reasoning; WM= Working Memory; PS= Processing Speed; BD= Block Design; SI= Similarities; DS= Digit Span; PCn= Picture Concept; CD= Coding; VC= Vocabulary; LN= Letter-Number-Sequencing; MR= Matrix Reasoning; CO= Comprehension; SS= Symbol Search.

Table 44 shows the standardized factor loadings of all the core subtests of WISC-IV ^{PAK} on their respective latent factor along with the variance explained by that factor. Factor loadings are quite high and convincing for all the subtests ranging from *b* = .74 (digit span loading on WM) at the lowest to .92 (similarities loading on VC) at the highest. This is also evident that the latent factor VC explained highest variance in its subtests (from 85% to 90%); whereas PR is accounted for more than 60% of variance in its subtest (ranging from 61% to 68 %). Similarly, WM and PS factors are also explaining significant account of variance in their subtests.

These results of having strong loadings on the subsequent latent factors are indicated to be consistent for the three age groups. Evidently, for the age group of 6-8 years $b = .50$ (coding on PS) to .83 (comprehension of VC); for 9 – 12 years $b = .57$ (digit span on WM) to .86 (similarities on VC); and for 13 – 16 years $b = .49$ (coding on PS) to .90 (similarities on VC). All these results strongly confirm a four-factor structure for WISC-IV ^{PAK} and give a good evidence of its structural validity.

Confirmatory Factor Analysis of All the WISC-IV Subtest. In this set of CFA the factor structure of all WISC-IV ^{Urdu} subtests (core and supplemental) was evaluated through different models. These models were evaluated for the overall sample (N= 100), and for the three age groups of 6 -8 years ($n = 219$), 9 – 12 years ($n = 291$), and 13 – 16 years ($n = 290$). The three structural models to be tested for fitting best to the data are again as following:

1. Null Model: A model with no common latent factors.
2. Model 1: A model with one general latent factor. All the subtests will be loaded on a general factor.
3. Model 2: A four factor model (as proposed by WISC-IV Technical and Interpretive Manual, 2003). Five Verbal Comprehension subtests on the first factor, four Perceptual Reasoning subtest on the second factor, three Working Memory subtests on the third factor, and three Processing Speed subtests on the fourth factor.

Data was analyzed for assessing the basic assumption of 'Structural Equation Modeling (SEM)' including the assumption of multivariate normal distribution. For the overall sample, the distribution seemed to violate the assumption of multivariate normality as the critical ratio (c.r.) for multivariate kurtosis exceeded the critical value (> 5.00) indicating non-normality in distribution (Bentler, 2005 as cited in Byrne, 2010). So for the overall data CFA was conducted using the Asymptotic Distribution Free (ADF) estimation (see Byrne, 2010). This assumption of multivariate normality was met for the sample divided in the three age groups so at the groups level CFA was conducted using the Maximum Likelihood Estimation (MLE). All three structural models were evaluated through various goodness-of-fit measures discussed previously including chi-square related indices, absolute fit indices, and comparative fit indices.

Table 45

Goodness-of-Fit Statistics for Confirmatory Factor Analysis of All Subtests for Full sample and for the Three Age Groups

Models	Goodness-of-fit Index			Improvement		Goodness-of-fit Index				
	X^2	*df*	X^2/df	ΔX^2	Δdf	GFI	AGFI	CFI	TLI	RMSEA
Age: 6-16										
Null Model	1366.03	105	13.01			.79	.76	.00	.00	.12
Model 1	599.98	90	6.66	766.05	15	.91	.88	.60	.53	.08
Model 2	368.48	84	4.38	231.5	6	.94	.92	.77	.72	.06
Age: 6-8										
Null Model	1317.06	105	12.54			.35	.26	.00	.00	.23
Model 1	306.82	90	3.41	1010.24	15	.84	.79	.82	.79	.10
Model 2	198.4	84	2.36	108.42	6	.90	.85	.91	.88	.08
Age: 9-12										
Null Model	2345.93	105	22.34			.25	.14	.00	.00	.27
Model 1	307.26	90	3.41	2038.67	15	.87	.83	.90	.89	.09
Model 2	200.73	84	2.4	106.53	6	.92	.88	.95	.93	.07
Age:13-16										
Null Model	2252.84	105	21.46			.27	.16	.00	.00	.27
Model 1	306.63	90	3.40	1946.21	15	.86	.81	.90	.88	.09
Model 2	170.44	84	2.02	136.19	6	.93	.90	.96	.95	.06

Note. $n = 800$ for 6 – 16 years; $n = 219$ for 6-8 years; $n = 291$ for 9-12 years; $n = 290$ for 13-16 years; X^2 = Chi-Square value; *df*= Degrees of Freedom; ΔX^2 = Change in Chi-Square Value; GFI= Goodness-of-Fit Index; AGFI= Adjusted Goodness-of-Fit Index; CFI= Comparative Fit Index; TLI= Tucker-Lewis index; RMSEA= Root Mean Square Error of Approximation.

Table 45 shows the CFA Goodness-of-Fit Indices for WISC-IV all the subtests. For the overall sample (N= 800) due to multivariate non-normality, the three models have been tested through ADF (Asymptotic Distribution Free) estimation for goodness of fit. Considering the overall sample it is quite evident from the results that the Model 2 (four-factor model) fits the data best as compared to Model 1 (one factor model) and the null model. Null model have shown a total misfit while, model 1 has shown some improvement considering the chi-square values and other goodness-of –fit indices. But model 2 has shown substantive improvement over model 1 on most indices of goodness-of-fit. Chi-square value has lowered substantively; the GFI and AGFI values are also close to 1.00 indicating a good fit (see Byrne, 2010). Noticeably although CFI and TLI values for four-factor model do not show good fit of the model may be due to sensitivity with multivariate normality. But, considering RMSEA for which values as high as .08 are considered as representing reasonable or adequate fit (see Byrne, 2010; Wechsler, 2004), Model 2 again seems to have adequate fit to the data.

This result is consistent for the three age groups as all goodness-of –fit indices are indicating best fit for the Model 2 (the four-factor model). For the three age groups the three models were tested through MLE for best fit. Chi-square values for Model 2 has shown substantive improvement and the ratio of chi-square and degrees of freedom are also less than 2.5 signifying a good model fit for all the three age groups. Similarly, the absolute fit indices (GFI & AGFI) and comparative fit indices (CFI) also indicate good to adequate model fit for four-factor structure, and same holds true for TLI and RMSEA. The best fitted four-factor structure for the overall sample is presented in the Figure 4 below and for the figural presentation of the four-factor structure for the three age groups for all subtests see Appendix Q1 – Q3.

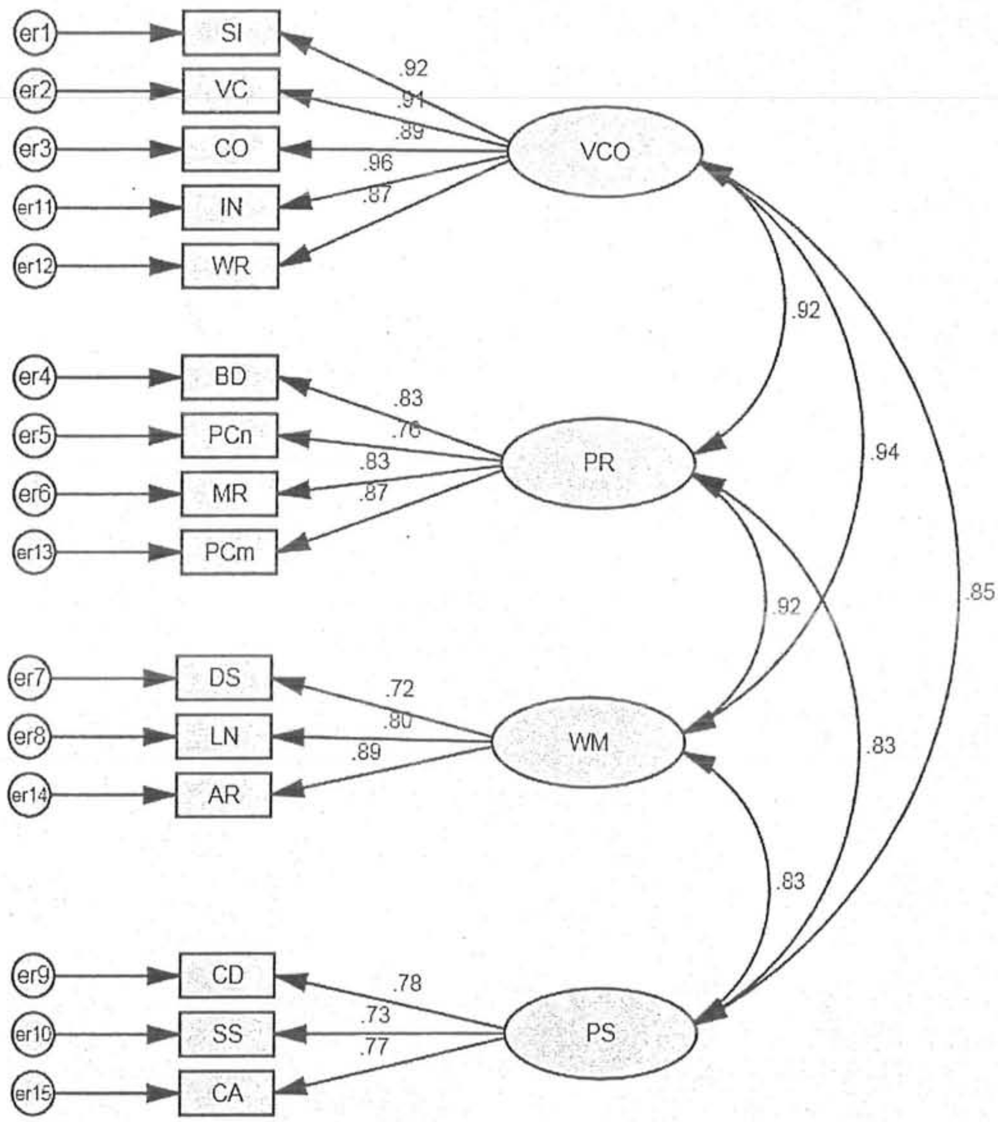


Figure 4: Four-Factor Structure of WISC-IV PAK Subtests

The oval ellipse represent latent factor; the rectangles represents the subtests; the small circles represent the error variance

The figure 4 shows the best fitted Four-Factor Structure for all subtests of WISC-IV PAK. The similarities, vocabulary, comprehension, information, and word reasoning subtests are loaded on first factor (Verbal Comprehension- VCO); the block design, picture concept, matrix reasoning, and picture completion subtests got loaded on second factor (Perceptual Reasoning- PR); the digit. span, letter-number-sequencing, and arithmetic subtests are loaded on third factor (Working Memory- WM), whereas the coding, symbol search, and cancelation subtests got loaded on fourth factor (Processing Speed- PS). Figure also shows that all the four factors are highly correlating which is well expected as all are measuring one umbrella construct of general intelligence 'g'.

Table 46

Sources of Variance for each Subtest in the Four-Factor Model of WISC-IV ^{PAK} (N = 800)

Subtests	VC		PR		WM		PS	
	b	var	*b*	var	*b*	var	*b*	var
SI	.92	84	-	-	-	-	-	-
VC	.91	83	-	-	-	-	-	-
CO	.89	80	-	-	-	-	-	-
IN	.96	92	-	-	-	-	-	-
WR	.87	76	-	-	-	-	-	-
BD	-	-	.83	68	-	-	-	-
PCn	-	-	.76	58	-	-	-	-
MR	-	-	.83	69	-	-	-	-
PCm	-	-	.87	75	-	-	-	-
DS	-	-	-	-	.72	51	-	-
LN	-	-	-	-	.80	64	-	-
AR	-	-	-	-	.89	79	-	-
CD	-	-	-	-	-	-	.78	62
SS	-	-	-	-	-	-	.73	53
CA	-	-	-	-	-	-	.77	59

Note. b = Standardized loading of Subtests on factor; var = Percent variance explained in the Subtest; VC= Verbal Comprehension; PR= Perceptual Reasoning; WM= Working Memory; PS= Processing Speed; BD=

Block Design; SI= Similarities; DS= Digit Span; PCn= Picture Concept; CD= Coding; VC= Vocabulary; LN= Letter-Number-Sequencing; MR= Matrix Reasoning; CO= Comprehension; SS= Symbol Search.

Table 46 shows the standardized factor loadings of all the subtests of WISC-IV PAK on their respective latent factor along with the variance explained by that factor. Factor loadings are quite high and convincing for all the subtests ranging from b = .72 (digit span loading on WM) at the lowest to .96 (information loading on VC) at the highest. This is also evident that the latent factor VC explained highest variance in its subtests (from 76% to 92%); whereas PR is accounted for more than 67% of variance in its subtest (ranging from 58% to 75%). Similarly, WM and PS factors are also explaining significant account of variance in their subtests.

These results of having strong loadings on the subsequent latent factors are indicated to be consistent for the three age groups. Accordingly, for the age group of 6-8 years b = .50 (cancelation on PS) to .83 (comprehension on VC); for 9 – 12 years b = .57 (digit span on WM) to .86 (similarities on VC); and for 13 – 16 years b = .46 (cancelation on PS) to .88 (information on VC). All these results strongly confirm a four-factor structure for WISC-IV PAK and again give a good evidence of its structural validity.

Higher-Order Confirmatory Factor Analysis of WISC-IV PAK. WISC-IV Technical and Interpretive Manual (Wechsler, 2003) has explained and confirmed a four-factor structure model but no description is given about the higher-order factor structure of WISC-IV. Although discussion about g-loadings of various subtests were done in few sections of the manual, but a higher-order factor structure with 'g' as second factor was not proposed or confirmed. Several researches have emphasized the role of 'g' in structure of WISC-IV and other Wechsler series tests (see for example Bodin et al., 2009;

Watkins, 2006; Canivez & Watkins, 2010a, 2010b; Canivez, 2008, 2011; Nelson & Canivez, 2012 as cited in Watkins et al., in press). So higher-order factor structure of WISC-IV ^{PAK} was also confirmed.

Initially higher-order factor structure of only core subtests was confirmed but later second-order factor structure of all subtests was confirmed. A model with four-first order factors and one second order factor was tested for fitness against a null model (with no common factor) for both CFAs (with core subtests only and with all subtests). Maximum Likelihood Estimation (MLE) was used for CFA of core subtests, but to compensate for multivariate kurtosis Asymptotic Distribution Free (ADF) estimation was used for confirming higher-order structure of all subtests of WISC-IV ^{PAK}. Models were evaluated through various goodness-of-fit indices as discussed before.

Table 47

Goodness-of-Fit Statistics for Higher-Order Confirmatory Factor Analysis of Core Subtests and of All Subtests for Full sample (N = 800)

Models	Goodness-of-fit Index			Improvement		Goodness-of-fit Index				
	X^2	*df*	X^2/df	ΔX^2	Δdf	GFI	AGFI	CFI	TLI	RMSEA
For Core Subtests										
Null Model	6146.6	45	136.59			.22	.05	.00	.00	.412
Model 1	151.49	31	4.88	5995.11	14	.96	.96	.98	.97	.07
For All Subtests										
Null Model	1366.03	105	103.01			.79	.76	.00	.00	.36
Model 1	369.26	86	4.29	996.77	19	.94	.92	.77	.73	.06

Note. X^2 = Chi-Square value; *df*= Degrees of Freedom; ΔX^2 = Change in Chi-Square Value; GFI= Goodness-of-Fit Index; AGFI= Adjusted Goodness-of-Fit Index; CFI= Comparative Fit Index; TLI= Tucker-Lewis index; RMSEA= Root Mean Square Error of Approximation.

Table 47 shows the Higher-order CFA Goodness-of-Fit Indices for WISC-IV core and all subtests. It is quite evident from the results that the Model 1 (four-factor model with one first factor) fits the data better than the null model both for the core subtests and all fifteen subtests. Null model have shown a total mis-fit whereas model 1 has shown substantive improvement over null model on all indices of goodness-of-fit. For the core subtests, Chi-square value has lowered substantively; the GFI and AGFI values are also close to 1.00 indicating a good fit (see Byrne, 2010). Similarly, CFI and TLI values for model 1 are greater than .95 while for these indices a value close to .95 represents good fit (Hu & Bentler as cited in Byrne, 2010). Furthermore, considering RMSEA any value $\leq .05$ shows good model fit, but values as high as .08 are also considered as representing reasonable or adequate fit (see Byrne, 2010; Wechsler, 2004), so RMSEA value also indicate adequate fit of the Model 1 to the data. Similarly, the substantive improvement in chi-square statistics; acceptable indices for GFI and AGFI ($> .90$); and adequate RMSEA value ($< .08$) all are signifying model 1 as fitting the data well.

Figure 5 and figure 6 presents the Higher-order Factor Structure Model for WISC-IV ^{PAK} core subtests and for all subtests of WISC-IV ^{PAK}, respectively.

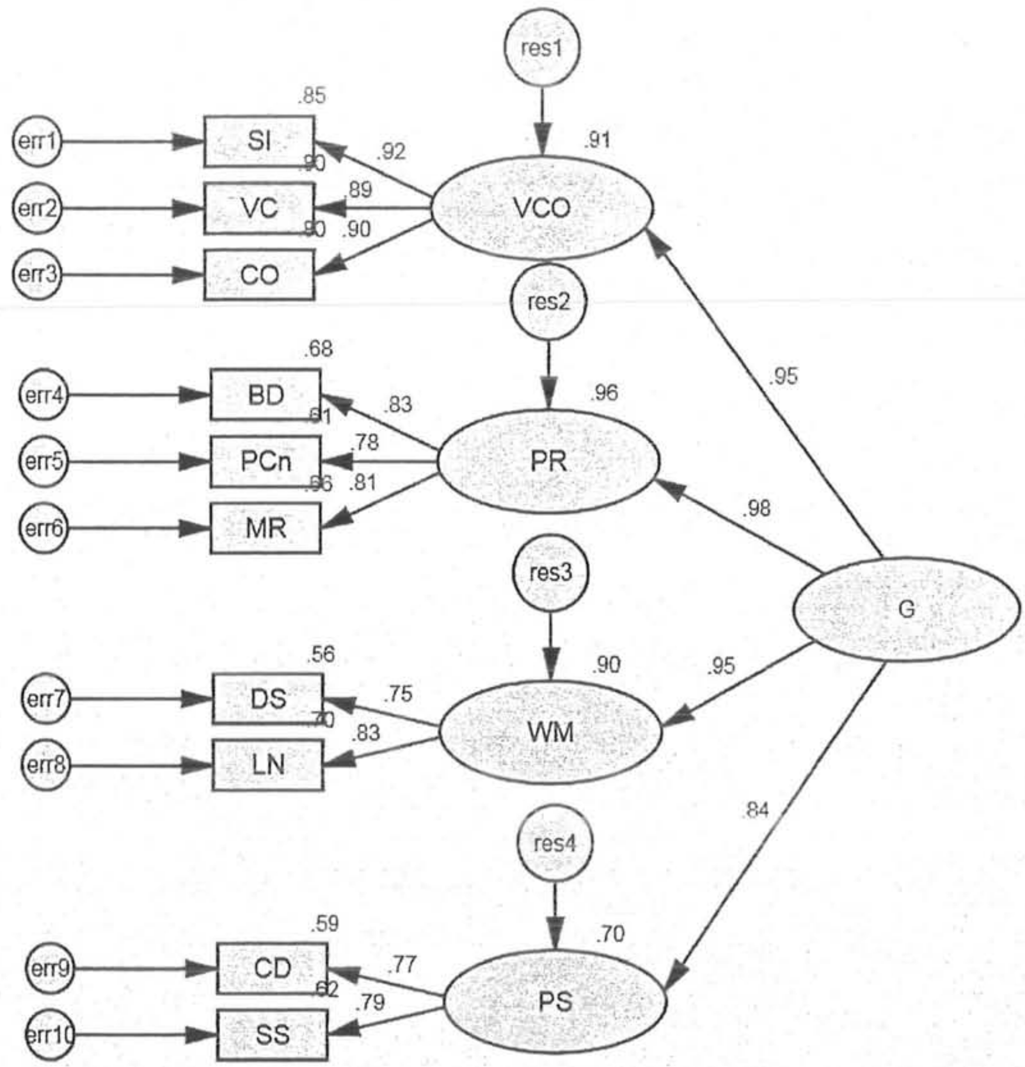


Figure 5: Higher-Order Four-Factor Structure of WISC-IV PAK Core Subtests

The single oval ellipse represent second order latent factor; the four oval ellipses represent first order factors; the rectangles represent the subtests; the small circles represent the error variance

Figure 5 reveals that among the four first-order factors, first factor has three subtests, second has three subtests, while third and fourth first-order factors have two subtests each as specified in the best fitted model of WISC-IV PAK core subtests. Moreover, it is also evident that all the first-order factors subsequently have heavy loadings on the second-order factor of 'G'.

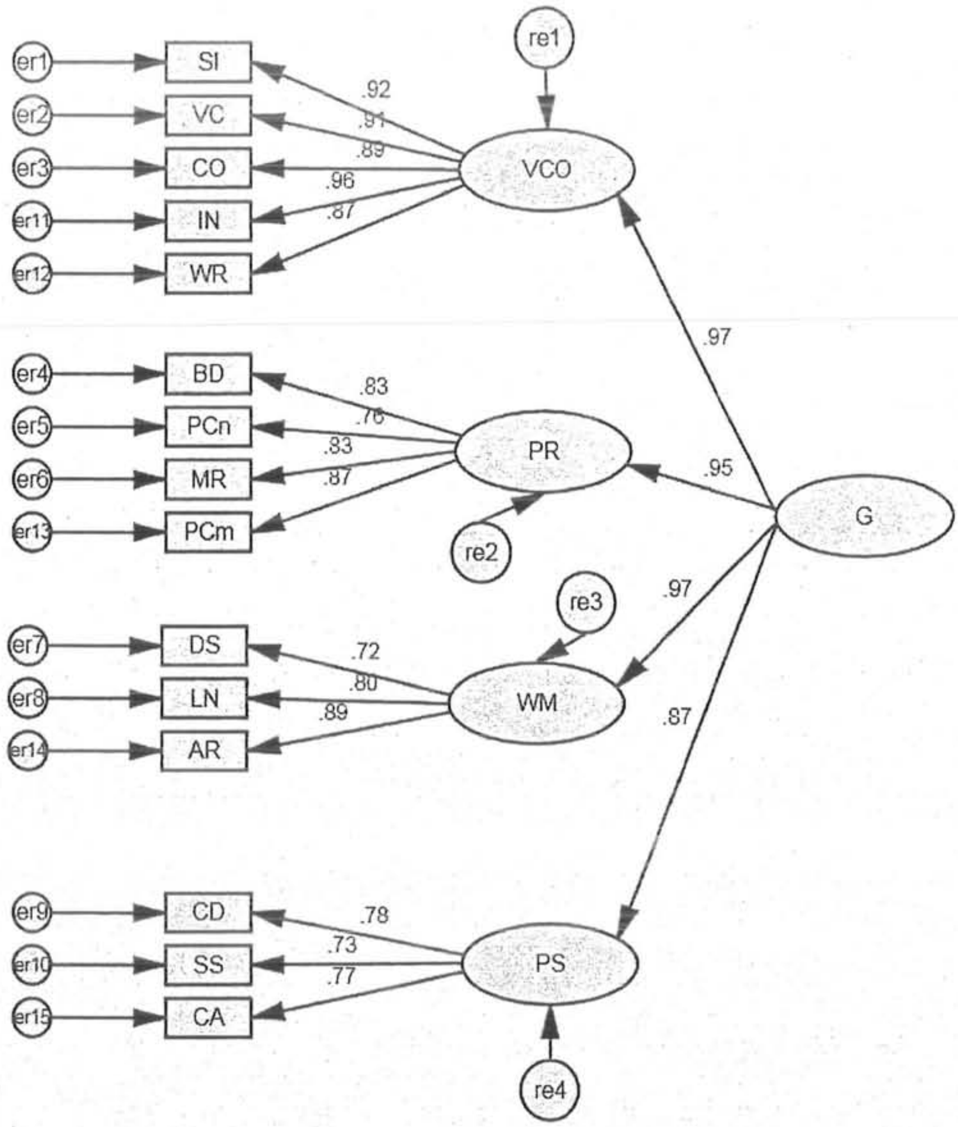


Figure 6: Higher order Four-Factor Structure of WISC-IV PAK Subtests

The single oval ellipse represent second order latent factor; the four oval ellipses represent first order factors; the rectangles represent the subtests; the small circles represent the error variance

Figure 6 reveals that among the four first-order factors, first factor has five subtests, second has four subtests, while third and fourth first-order factors have three subtests each as specified in the best fitted model for all subtests of WISC-IV ^{PAK}. Moreover, it is also evident that all the first-order factors subsequently have heavy loadings on the second-order factor of 'G'.

Table 48

Sources of Variance for each Subtest in the Four-Factor Model of WISC-IV ^{PAK} Core Subtests and All Subtests (N = 800)

Subtests	G		VCO		PR		WM		PS	
	b	var	*b*	var	*b*	var	*b*	var	*b*	Var
Core Subtests										
SI	-	-	.92	85	-	-	-	-	-	-
VC	-	-	.89	80	-	-	-	-	-	-
CO	-	-	.90	80	-	-	-	-	-	-
BD	-	-	-	-	.83	68	-	-	-	-
PCn	-	-	-	-	.78	61	-	-	-	-
MR	-	-	-	-	.81	66	-	-	-	-
DS	-	-	-	-	-	-	.75	56	-	-
LN	-	-	-	-	-	-	.83	70	-	-
CD	-	-	-	-	-	-	-	-	.77	59
SS	-	-	-	-	-	-	-	-	.79	62
VCO	.95	91	-	-	-	-	-	-	-	-
PR	.98	96	-	-	-	-	-	-	-	-
WM	.95	90	-	-	-	-	-	-	-	-
PS	.84	70	-	-	-	-	-	-	-	-
All Subtests										
SI	-	-	.92	84	-	-	-	-	-	-
VC	-	-	.91	83	-	-	-	-	-	-
CO	-	-	.89	80	-	-	-	-	-	-
IN	-	-	.96	92	-	-	-	-	-	-
WR	-	-	.87	76	-	-	-	-	-	-
BD	-	-	-	-	.83	68	-	-	-	-
PCn	-	-	-	-	.76	58	-	-	-	-
MR	-	-	-	-	.83	69	-	-	-	-
PCm	-	-	-	-	.87	75	-	-	-	-
DS	-	-	-	-	-	-	.72	52	-	-
LN	-	-	-	-	-	-	.80	64	-	-
AR	-	-	-	-	-	-	.89	79	-	-
CD	-	-	-	-	-	-	-	-	.78	61
SS	-	-	-	-	-	-	-	-	.73	53
CA	-	-	-	-	-	-	-	-	.77	60
VCO	.97	94	-	-	-	-	-	-	-	-
PR	.95	90	-	-	-	-	-	-	-	-
WM	.97	94	-	-	-	-	-	-	-	-
PS	.87	76	-	-	-	-	-	-	-	-

Note. *b* = Standardized loading of Subtests on factor; var = Percent variance explained in the Subtest; G= General Intelligence; VC= Verbal Comprehension; PR= Perceptual Reasoning; WM= Working Memory; PS= Processing Speed; BD= Block Design; SI= Similarities; DS= Digit Span; PCn= Picture Concept; CD= Coding; VC= Vocabulary; LN= Letter-Number-Sequencing; MR= Matrix Reasoning; CO= Comprehension; SS= Symbol Search.

Table 48 shows the standardized factor loadings '*b*' of the subtests on the first-order factors and the same of first-order factors on the second order factor of 'G'. The amount of variance in the subtests as accounted by its subsequent factor is also indicated. Considering the core subtests factor loadings on 1st factors are quite satisfactory for all the subtests ranging from $b = .77$ (coding loading on PS) at the lowest to .92 (similarities loading on VCO) at the highest. Similarly the loadings of 1st factor on the 2nd factor (G) are also quite high (from .84 for PS to .98 for PR). This is also evident that the latent factor VC explained highest variance in its subtests (from 80% to 85%); whereas PR is accounted for 61% to 68% variance in its subtests. Similarly, WM is explaining 56 to 70% variance in its subtests, and PS is explaining 59 to 62% variance in its subtests. The G factor also seems to account for very high amount of variance in the 1st order factors ranging from 70% for PS to 96% for PR.

Considering the all subtests factor loadings on 1st factors are quite satisfactory for all the subtests ranging from $b = .72$ (digit span loading on WM) at the lowest to .96 (information loading on VCO) at the highest. Similarly the loadings of 1st factor on the 2nd factor (G) are also quite high (from .87 for PS to .97 for VCO). The latent factor VC seems to explained highest variance in its subtests (from 76% to 92%); whereas PR is accounted for 58% to 69% variance in its subtests. Similarly, WM is explaining 52 to 79% variance in its subtests, and PS is explaining 53 to 61% variance in its subtests. The G factor also seems to account for very high amount of variance in the 1st order factors ranging from 76% for PS to 94% for VCO.

Phase III: Development of Norms for the WISC-IV ^{PAK} in Pakistan

Phase III aimed at derivation of Pakistani norms for the WISC-IV ^{PAK} subtests. As two types of norms (Standard Scores and Test-Age-Equivalent norms) are to be established so norms derivation followed various steps.

Sample. The same normative sample (N = 800) that was used in Phase I and II was used to derive norms for the adapted subtests. As the norms are to be developed on age group of one year, so sample was divided into 11 age groups of one year each including: 6:00 - 6:11, 7:00 – 7:11, 8:00 – 8:11, 9:00 – 9:11, 10:00 – 10:11, 11:00 – 11:11, 12:00 – 12:11, 13:00 – 13:11, 14:00 – 14:11, 15:00 – 15:11, and 16:00 – 16:11. Each age group comprised of 72 or 73 students (for detailed Sampling Design see Table 36).

Instrument and Materials. The Informed Consent Form and WISC-IV ^{PAK} (Urdu Standardization Edition) kits were used for test administrations. Each kit includes WISC-IV (Urdu Standardization Edition) Administration and Scoring Manual; Record Form (Urdu); Response Booklet 1 and 2; and a Stimulus Booklet along with block design box and scoring keys (for details see Phase I).

Procedure. The test administration, scoring, and data recording procedure is same as of Phase I of this study.

Results of Phase III. Phase III aimed at the development of normative information for proper use of the WISC-IV ^{PAK} subtests in Pakistan. For that purpose scaled scores and composite scores (standard scores); and age equivalent norms are developed. The method and procedures used for norms development of WISC-IV, as mentioned in the *WISC-IV Technical and Interpretive Manual*, 2003 has been followed for development of Pakistani norms.

Step I: Development of Standard Score Norms. The conversion of raw scores into standard scores makes it possible to compare scores within the WISC-IV ^{PAK} subtests, and between WISC-IV and other related measures. Two types of age based standard scores are derived for adapted subtests: scaled scores and composite scores.

Derivation of Scaled Scores for WISC-IV ^{PAK} Subtests. For each of the subtests, the distribution of each age group's total raw score is converted to a scale with a mean of 10 and a standard deviation of 3. This conversion is accomplished by formation of frequency distribution of raw scores for each age group of one year, normalizing these distributions (using z-scores), and calculating the appropriate scaled score for each total raw score. The progression of scaled score with in an age group and from age group to age group is carefully examined, and minor irregularities are eliminated by smoothing. The derived scaled scores range from 1 to 19, providing a range of 3 standard deviations on either side of the mean. The scaled score equivalents are developed for all the eleven age groups (starting from 6 years to 16 years and 11 months) and are reported in the appendices section (Appendix R1 – R11).

Means and Standard Deviations of the Scaled Scores for the WISC-IV ^{PAK} subtests. For having a comparative picture of Pakistani scaled score norms, means and standard deviations of the scaled scores for all the WISC-IV subtests are computed for the whole population, and for the 11 age groups.

Table 49

Means and Standard Deviations of the Raw Scores and Scaled Scores on WISC-IV ^{PAK} Subtests for the Normative Sample (N = 800)

Subtests	WISC-IV ^{PAK} Raw Scores		WISC-IV ^{PAK} Scaled Scores	
	M	*SD*	*M*	*SD*
Block Design	26.15	12.62	10.00	2.98
Similarities	16.83	8.59	10.00	2.98
Digit Span	17.36	4.55	10.00	2.98
Picture Concept	14.73	4.53	10.00	2.98
Coding	46.59	14.41	10.00	2.98
Vocabulary	27.62	11.35	10.00	2.98
Letter-N-Sequence	16.27	4.94	10.00	2.98
Matrix Reasoning	16.74	5.97	10.00	2.98
Comprehension	20.13	7.71	10.00	2.98
Symbol Search	22.35	8.22	10.00	2.98
PictureCompletion	20.66	6.60	10.00	2.98
Cancellation	70.06	22.38	10.00	2.98
Information	16.39	5.96	10.00	2.98
Arithmetic	21.06	5.38	10.00	2.98
Word Reasoning	10.93	4.16	10.00	2.98

Table 49 shows the means and standard deviations of the raw scores and scaled scores for the adapted WISC-IV ^{PAK} subtests for the overall sample. Means and standard deviations of scaled scores for all subtests are approximating the presumed mean (10) and standard deviation (3) of subtests for Wechsler tests series (see Wechsler, 2003).

Means and standard deviations of the scaled scores on WISC-IV ^{PAK} subtests are also computed for the 11 age groups of one year.

Table 50

Means and Standard Deviations of the Scaled Scores on WISC-IV ^{PAK} Subtests for All the Eleven Age Groups (N = 800)

Subtests	6 years (*n*=73)		7 years (*n*=73)		8 years (*n*=73)		9 years (*n*=73)		10 years (*n*=72)		11 years (*n*=73)		12 years (*n*=73)		13 years (*n*=73)		14 years (*n*=73)		15 years (*n*=72)		16 years (*n*=72)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
BD	10.0	3.0	10.0	2.9	10.0	3.0	10.0	3.0	10.0	2.9	10.0	2.9	10.0	3.0	10.0	3.0	10.0	3.0	10.0	2.9	10.0	3.0
SI	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	2.9	10.0	3.0	10.0	2.9	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0
DS	10.0	3.0	10.0	3.0	10.0	2.9	10.0	3.0	10.0	3.0	10.0	2.9	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0
PCn	10.0	3.0	10.0	2.9	10.0	2.9	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	2.9	10.0	3.0	10.0	2.9	10.0	3.0
CD	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	2.9	10.0	2.9	10.0	3.0	10.0	3.0	10.0	3.0
VC	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	2.9	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	2.9	10.0	3.0
LNS	10.0	2.9	10.0	3.0	10.0	3.0	10.0	2.9	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0
MR	10.0	2.9	10.0	3.0	10.0	3.0	10.0	3.0	10.0	2.9	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0
CO	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	2.9	10.0	3.0	10.0	3.0	10.0	2.9
SS	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0
PCm	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	2.9	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	2.9
CA	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	2.9	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	2.9
IN	10.0	3.0	10.0	3.0	10.0	3.0	10.0	2.9	10.0	3.0	10.0	3.0	10.0	3.0	10.0	2.9	10.0	3.0	10.0	3.0	10.0	3.0
AR	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0
WR	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0	10.0	3.0

Note. BD=Block Design; SI=Similarities; DS=Digit Span; PCn=Picture Concept; CD=Coding; VC=Vocabulary; LN=Letter-Number-Sequencing; MR=Matrix Reasoning; CO= Comprehension; SS= Symbol Search; PCm= Picture Completion; CA= Cancelation; IN= Information; AR= Arithmetic; WR= Word Reasoning.

Table 50 shows the means and standard deviations of scaled scores of all the adapted subtests for the 11 age groups of 6 years to 16 years. Scaled score means computed for all WISC-IV ^{PAK} subtests are exactly the same as presumed mean ($M = 10$) of the Wechsler subtests, whereas the standard deviations are also the same or closely approximating the presumed standard deviations for WISC-IV ($SD = 3$). This result is consistent for all the age groups.

Derivation of Composite Scores for WISC-IV ^{PAK} Indices. For interpretive purpose five kind of composite scores were also developed. Four are based on four indices of WISC-IV (as VCI, PRI, WMI, and PSI), while the fifth composite score is Full Scale IQ. The composite scores are based on sums of scaled scores. VCI equivalent or composite is based on sum of scaled scores of core subtest (Similarities, Vocabulary, and Comprehension). PRI equivalent or composite is based on sum of scaled scores of core subtest (Block Design, Picture Concept, and Matrix Reasoning). WMI equivalent or composite is based on sum of scaled scores of its core subtest (Digit Span and Letter-number-Sequence). PSI equivalent or composite is based on sum of scaled scores of its core subtest (Coding and Symbol Search). To construct the composite score table sum of scaled scores of the core subtests has been computed for their subsequent scale/index for each student in the standardization sample. To derive composite scores for the scales, the distribution of the sum of scaled scores has been normalized and given a mean of 100 and standard deviation of 15. The appropriate composite scores are then assigned to each sum of scaled score. The resulting composite score distributions are then smoothed visually to remove any irregularities, while attempting to keep the mean and standard deviation of the scale close to 100 and 15.

The composite norm tables depict index based composites and FSIQs around at least ± 2 *SD* from the presumed mean of 100 for WISC-IV. Any score lying below or above ± 2 *SD* depicts borderline or very superior intelligence, respectively. All the five composite scores norms The Four Index based Composite Scores and FSIQ derived from sum of scaled scores for WISC-IV ^{PAK} along with the confidence interval are attached in the Appendix S1 – S5.

Table 51

Means and Standard Deviations of Composite Scores for WISC-IV ^{PAK} Indices and Full Scale IQ (N=800)

Composites	M	SD	Maximum	Minimum
VCI	100.00	14.06	139	72
PRI	99.99	13.12	133	66
WMI	100.26	13.49	142	60
PSI	100.26	13.47	140	61
FSIQ	100.01	12.12	130	69

Note. VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index; FSIQ = Full Scale IQ

Table 51 shows mean and standard deviations of the five newly derived composite scores along with minimum and maximum values. All means for composite scales and their standard deviations are approximating the presumed means and standard deviation of children on Wechsler tests. Same is true for the FSIQ. The lowest sum of scaled score calculated on the present study's sample was 10. So the computed VCI equivalents for adapted version range from 72 to 139, for PRI it ranged from 66 to 133, for WMI it ranged from 60 to 142, while for PSI the equivalents ranged from 61 to 140.

Computation of Confidence Interval. Confidence intervals for the composite scores are also computed. A child's true score is more accurately represented by establishing a confidence interval of true scores. It is like a band of true scores in which the true score is likely to lie. Confidence intervals can be computed by using Standard Error of Measurement (SEM) or by using Standard Error of Estimate (SEE) and estimated true scores (Wechsler, 2003). For the present study, confidence interval is computed by using the SEM.

Table 52

Standard Error of Measurement and Confidence Interval for Composite Scales on Adapted WISC-IV ^{Urdu} Subtests (N = 800)

Composite Scales	SEM	95% Confidence Interval
VCI	2.6	Observed score ± 5.1
PRI	4.5	Observed score ± 8.8
WMI	5.2	Observed score ± 10.2
PSI	8.35	Observed score ± 16.4
FSIQ	4.0	Observed score ± 7.8

Note. VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index; FSIQ = Full Scale IQ; SEM = Standard Error of Measurement.

Table 52 shows the standard error of measurement and confidence interval for the five composite scales of adapted versions. After computation of SEM, for 95% confidence level observed score has to be added or subtracted by 1.96 (SEM). This procedure provided a band/range of confidence interval for every individual score (composite score) of the VCI, PRI, WMI, PSI, and FSIQ.

Step II: Comparison of Pakistani and UK Norms on WISC-IV ^{PAK} Scales. To explore the effect of regional norms on any adapted instrument WISC-IV ^{PAK} scores were compared using both Pakistani and UK norms (Table A.1 – A.6 in WISC-IV ^{UK} Administration and Scoring Manual; Wechsler, 2004).

Sample. It comprised of 44 children taken at random from the normative sample. From each of the 11 normative age groups four children were randomly selected (using SPSS select cases option).

Instruments. For Pakistani norms, age based Scaled scores norms and Composite Scores norms (attached in Appendix R1- R11 & S1- S5) were used, whereas, for UK norms Tables A.1 – A.6 of WISC-IV ^{UK} Administration and Scoring Manual (Wechsler, 2004) were consulted.

Procedure. Children raw scores on WISC-IV ^{PAK} core subtests were converted into two sets of scaled scores, first with Pakistani scaled score norms and second with UK norms. Then after summing up the scale scores, the five respective composite scores were derived for the two sets of data. These two sets of data were then analyzed for mean differences by using paired sample *t*-test.

Results. For both sets of data, means and standard divinations were calculated and then compared for all five composite scores including VCI, PRI, WMI, PSI, and FSIQ composites using paired sample *t*-test.

Table 53

Differences of Means and Standard Deviations of Four Index based Composite Scores and FSIQ when Pakistani and UK Norms are Used ($N = 44$)

Comp. scores	Pakistani norms		UK norms		t	p	95% Confidence Interval		d
	M	SD	M	SD			UL	LL	
VCI	102.45	12.02	90.27	13.94	10.34	.000	14.63	9.73	2.35
PRI	100.95	10.50	86.95	10.80	20.76	.000	15.40	12.60	4.43
WMI	103.59	12.44	105.14	17.46	1.14	.264	1.26	-4.35	-0.40
PSI	102.15	14.49	95.05	16.21	5.91	.000	9.59	4.59	1.32
FSIQ	101.86	10.01	91.59	14.87	8.32	.000	12.84	7.71	3.21

Note. VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index; FSIQ = Full Scale IQ; $df = 43$.

Table 53 shows the differences in means and standard deviations of five composite scores when computed from Pakistani and UK norms. For VCI, PRI, PSI, and FSIQ the means derived from Pakistani norms are indicated to be significantly ($p < .001$) higher then means derived from UK norms. Contrary to other composites, in case of WMI composite though the mean difference was non significant but it has a larger value for the second set of data (i.e. mean composite scores based on UK norms). Mean values for VCI, PRI, and PSI indicate a 7 to 14 points gain when Pakistani norms are used. Whereas, for FSIQ this gain is almost 10 points.

This result strongly proves the utility of regional/national norms in interpreting children's scores on even an indigenized test (a test that has been originally developed in a foreign country and was cultural/linguistically adapted for use in another region).

Step III: Development of Test-Age Equivalent Norms. Test-age equivalents represent the average age at which a given total raw score is typically obtained by children of a specific age. Age equivalents for the subtest total raw scores are derived by identifying the total raw score corresponding to a scaled score of 10 in each of the normative age groups for each subtest. Minor irregularities are removed by visual smoothing.

Table 54

Test- Age Equivalents of Total Raw Scores for Adapted WISC-IV ^{PAK} Core Subtests (N = 800)

Subtests	Age Equivalents for One Year Age groups										
	6	7	8	9	10	11	12	13	14	15	16
BD	11-10	14	16-18	21	27-29	26-27	30-33	34-35	34-35	35-38	38-40
SI	8	9	10	12	14-15	16-17	19-20	22-23	21-22	24-25	26-27
DS	13	14	14	16	17	18	19	20	19	20	20-21
PCn	9	11	12	12-13	15	15-16	17	17	17	18	18
CD	39-42	41-44	31-32	33-34	41-42	40-42	52-55	52-55	55-59	56-59	60-62
VC	16	17	17-18	21	23-24	26	29-31	34-35	34-35	38-40	43
LNS	10	12-13	13	14	16-17	17	19	19	18-19	20	20
MR	10-11	12	12	14	16-17	17-18	19	20-21	19-20	21-22	22
CO	10-11	12-13	13-14	17	19-20	19-20	23	25-26	25-26	26-27	28-29
SS	17-18	19-20	16-17	16-17	20	21-22	24-25	26-27	26-28	26	28-30

Note. BD= Block Design; SI = Similarities; DS = Digit Span; PCn = Picture Concept; CD = Coding; VC = Vocabulary; LNS = letter Number Sequencing; MR = Matrix Reasoning; CO = Comprehension; SS = Symbol Search.

Table 54 shows the test-age equivalents of total raw scores for core adapted subtests of WISC-IV. Table indicates that there is a gradual increase in age-typical total raw scores for all the subtest across the increasing age groups.

Phase IV: Exploring the Relationship of Few Demographic Variables with Children Scores

Beside development of norms for WISC-IV ^{PAK}, another objective of the Phase II was to assess the relationship of age, gender groups, geographical area based groups, and parental education level based groups with the performance of students on WISC-IV ^{PAK}. For achievement of this objective data has been further analyzed.

The sample, instrument, and the procedure for this phase are exactly the same as of phase I of this study so it is not re-described here.

Means and Standard Deviations of the WISC-IV ^{PAK} Composite Scores. Means and standard deviations of all four index based composite scores and FSIQ are also computed for the 11 age groups, two gender groups, five geographical region based groups, and three parental educational level groups. Analysis of differences in means and standard deviation can help in understanding the importance of factors like age, gender, and parental education level in attaining verbal conceptualization and reasoning among children.

Means and Standard Deviations of the WISC-IV ^{PAK} Composite Scores on Age Groups. To explore the influence of increase in age on the FSIQ and other composite scores mean differences are analyzed through ANOVA.

Table 55

Differences in Mean and Standard Deviation of FSIQ of the 11 Age Groups (N = 800)

Age Grps.	*n*	*M*	*SD*	*F*	*df*	*p*
6 years	73	84.88	6.00	152.328	10, 789	.000
7 years	73	89.05	6.04			
8 years	73	88.56	5.81			
9 years	73	91.85	6.02			
10 years	72	97.92	7.33			
11 years	73	99.59	6.60			
12 years	73	105.84	7.44			
13 years	73	108.70	6.58			
14 years	73	108.53	10.40			
15 years	72	111.35	7.22			
16 years	72	114.21	7.70			

Table 55 shows the mean raw scores, standard deviation and F-ratios of FSIQ for each of the 11 age groups. The age group of 6 years has the lowest mean and standard deviation for all the subtests, while the age group of 16 years have highest mean and standard deviation for all the subtests. This indicates that all the items of the subtests are working well across the age groups as there is a gradual increase in the mean raw scores with increasing age. F-ratio show that means and standard deviations for all 11 age groups differ significantly ($p < .001$) on FSIQ.

Post-hoc multiple comparison through Gabriel procedure (as sizes of the groups are little different) indicates that each group has significant mean differences from all other age groups except from the immediately preceding or proceeding age group. For example, mean FSIQ score of 13 years old children differed significantly from means of all other age group except for 12 (preceding) and 14 (proceeding) years age groups mean FSIQ scores.

Results also indicated that the significant mean differences on age groups are also true for the four indices based composite scales. For VCI, mean differed significantly for the age group as F (10, 789) = 149.943 and p = .000, whereas for PRI F (10, 789) = 109.178 and p = .000 again indicating significant differences on age groups. Similarly, WMI (F (10, 789) = 65.743 & p = .000) and PSI (F (10, 789) = 63.319 & p = .000) also have significant mean differences on the age groups with the mean increasing across the increasing age groups.

Comparison of Means and Standard Deviations of Composite Scores on gender. For assessing the influence of gender on composite scores means and standard deviations of VCI, PRI, WMI, PSI, and FSIQ are compared for girls (n = 400) and boys (n = 400) through application of t-test. Results indicated non-significant gender based mean differences on all five composite scores. Girls have shown slight advantage in VCI and PSI; while boys have shown relatively higher scores on PRI. These mean differences on WMI and FSIQ seem to be almost negligible.

Comparison of Means and Standard Deviations of Composite Scores on Geographical Area. Beside age and gender the normative sample has also been stratified on the five geographical regions of Punjab (n = 450), Sindh (n = 190), KPK (n = 110); Balochistan (n = 40); and Capital territory (n = 10). Mean differences on the composite scores were also calculated on groups based on belonging to these geographical areas.

Results indicated that there are non-significant geographical area based mean differences on FSIQ, VCI, and PRI. But on WMI groups have shown significant difference with F (4, 795) = 2.78 at p = .026 where Balochistan (M = 107.08; SD = 13.24) has shown the highest mean, while Sindh (M = 99.61; SD = 13.81) has shown the lowest mean. Similarly, mean differences are also significant on PSI with F (4, 795) = 2.74 at p = .028. For PSI Capital territory (M = 102.20; SD = 13.52) has shown the highest mean while Sindh (M = 98.16; SD = 13.01) again has the lowest mean PSI score.

Comparison of Means and Standard Deviations of Composite Scores on Parental Education Level. Normative sample also has a division on the bases of children's average parental education level. The three age groups include: P. E. I (n = 284) with less than 10 years of average parental education level; P. E. II (n = 323) with less than 14 years of average parental education level; and P. E. III (n = 193) with average parental education level of 14 or above 14 years. Mean differences of composite scores on parental education level are also computed.

Results indicated significant mean differences on FSIQ (F (2, 797) = 10.31 & p = .000), VCI (F (2, 797) = 12.28 & p = .000), PRI (F (2, 797) = 11.96 & p = .000), and WMI (F (2, 797) = 7.02 & p = .001), while for PSI the mean differences are non-significant. The mean scores of P. E. III (with average parental education level of 14 or above 14 years) were the highest, while the mean scores of P. E. I (with less than 10 years of average parental education level) were the lowest for all the five composite scores. This suggests that examinees from the P.E. I showed relatively poor performance on WISC-IV than the other two groups.

Discussion on the Study III

The unavailability of psychometrically strong indigenous test; and the cultural and contextual dissimilarities in the west and Pakistan necessitates indigenization of tests developed in western cultures (Raza & Sheikh, 1991). This indigenization not only involves cultural adaptation of the test but also its standardization in that culture. So study III was aimed at standardization of WISC-IV ^{PAK}. This standardization process was completed through four phases. To accomplish this task a normative sample of 800 children was selected following the stratified random sampling design.

As with any assessment tool, WISC-IV ^{PAK} needs to have sufficient reliability and validity evidence for having utility in practical fields. So, phase I of this study was concerned with the establishment of reliability for the adapted WISC-IV subtests. Reliability involved establishment of both temporal stability and internal consistency evidence. To establish the test-retest reliability, after the first administration the sample was retested with a gap of 21 to 48 days (average interval of 35 days). For original WISC-IV test-retest reliability evidence, sample was retested with a mean interval of 32 days (Wechsler, 2003). The stability coefficients for all WISC- IV ^{PAK} subtests were significant and ranged from moderate or satisfactory (in .70s) to excellent (above .90) except for Symbol Search with a relatively low coefficient of .63. All coefficients are found to be comparable with the stability coefficients of original version (ranging from .76 of picture concept to .92 of vocabulary). Overall, VCI and PRI subtests seemed to have more temporal stability than the WMI and PSI subtests, and this is consistent with most findings on WISC-IV temporal stability (see for example Wechsler, 2003). The stability coefficient for vocabulary, being a measure of crystallized intelligence, is highest among all subtests for both adapted and original versions indicating it involves long term memory processes. Considering the test-retest gains, all mean scores were higher for retest sessions but with small effect sizes (ranging from .03 to .37). Most pronounced gain was in PSI subtests (symbol search, coding, and cancellation). These retest gains were likely as practice effect and task familiarity can effects performance on any subtest but these effects are more prominent for tasks involving speed.

Considering internal consistency, alpha coefficients for all adapted subtests were significant ranging from good (above .80) to excellent (above .90) indicating its high internal consistency. Verbal comprehension subtests were found to be the most internal consistent one and the same was found out in many other psychometric studies of WISC-IV For example, Dang et al. (2012) also found that the subtests in verbal comprehension index subtests displayed highest internal consistency in Vietnam's adaptation of WISC-IV. The same was reported in WISC-IV technical and interpretive manual (Wechsler, 2003). Considering the age groups, the internal consistency coefficients for the 6-8 years group was relatively lower than the higher age groups. For WISC-IV administration of each subtest begins at the age specific start points (see Wechsler, 2004). Due to these start points, increasing difficulty level of the items, and discontinuation rule younger children respond to a lesser number of items than the older children. This low number of items along with low variability in responses might have resulted in low correlation coefficients for this age group. Item total correlations were also explored and these were satisfactory for most items of the WISC-IV ^{PAK} subtests except for basal items. According to administration format, the age specific starting or basal items are to be responded correctly by the children to proceed further, so there is a lack of variability in responses to these items resulting in low coefficients.

Validation of WISC-IV ^{PAK} involved construct validation and cross validation of factorial structure and was carried out in phase II of this study. The convergent and discriminant validity of WISC-IV ^{PAK} was established following Multitrait-Multimethod Matrix method. This is a theoretical methodology presented by Campbell and Fiske (1959) in which several theory and research based hypotheses were tested in order to establish validity. Accordingly, all subtests showed low to moderate correlations with each other proving that all of them are measures of general intelligence 'g'. The fact that the subtests having high g-loading (VCI and PRI subtests) also have higher correlations with each other than with those subtests that have relatively low g-loadings (PSI subtests) establishes convergent and discriminant validity of the WISC-IV ^{PAK} subtests (see Flanagan & Kaufman, 2004, Wechsler, 2003). Moreover, stronger correlations among subtests of the same index than with the subtests of other indices also indicates convergent and discriminant validity of the adapted subtests. Observing the split-loading hypothesis has further established the convergent validity. According to that subtests like Arithmetic (for being a high g-loading subtest involving auditory comprehension), picture concept, and picture completion (may be involving use of verbal mediation for responding) displayed substantial loading on VCI as well as on their own index. Arithmetic subtest have also shown significant split loading on VCI and WMI in many other studies on WISC-IV structural validity (see for example Chen et al., 2009).

Cross validation of WISC-IV ^{PAK} structure as purposed in WISC-IV *Technical and Interpretive Manual* (Wechsler, 2003) was also successfully carried out. This was accomplished through conduction of multiple confirmatory factor analyses. Results indicated strong support for the four-factor model proposed for WISC-IV in comparison with a single factor model and null model (no factor, all subtests are independent). This finding of having best fit for a four-factor structure was consistent when run for core subtests or for all subtest; and for the whole sample or for the three age groups of 6-8 years, 9-12 years, and 13-16 years. Many other CFA based studies on WISC-IV structure has also confirmed this four-factor structure as best fitting the data (see for example Watkins et. al, in press; Chen et al., 2009). Although WISC-IV manual has not proposed higher order factor structure but many studies have indicated strong evidence of a second-order factor (G) for WISC-IV and other Wechsler series tests (see for example Watkins, 2006; Canivez & Watkins, 2010a,2010b; Canivez, 2008, 201; Nelson & Canivez as cited in Watkins et al., in press). So higher-order CFA was also conducted for the core subtests and for all the subtests of WISC-IV ^{PAK}. The results were consistent with the previous research findings (see for example Bodin et al., 2009) as a four-factor first order structure with the second-order factor 'G' accounting for most of the variance (76% to 94%) in the first-order factors (VCI, PRI, WMI, and PSI) was strongly supported. These multiple CFAs have strongly established the factorial cross validity of WISC-IV in Pakistan. The psychometric evidences discussed above indicates that the use of WISC-IV ^{PAK} seems promising in our culture. But as for any other test, further research and updating of this psychometric evidence is needed periodically in future.

Scores on test do not have any meaning without an interpretive frame work, which is provided by norms. So the aim of phase III was development of norms for the WISC-IV ^{PAK} in Pakistan. Two types of norms were developed, standard score norms and test-age equivalent norms. For the development of Pakistani norms same method was followed that was used for norms development of WISC-IV, as mentioned in the *WISC-IV Technical and Interpretive Manual*, 2003.

To facilitate comparisons within adapted subtests, and between adapted subtests and other related measures, the conversion of raw scores into standard scores was done. Two types of age based standard scores were derived including scaled scores and composite scores. The scaled scores are derived scored with a mean of 10 and standard deviation of 3 and range from 1 to 19, providing a range of 3 standard deviations on either side of the mean. The scaled score equivalents are developed for all 11 age groups separately from their respective population means and standard deviations. Means and standard deviations of scaled scores for whole population as well as for the 11 age groups were generally in close approximation to the presumed mean (10) and standard deviation (3) of subtests for Wechsler tests series. These scaled scores also seem to have good comparability with the UK means and standard deviations of scaled scores (see for example *WISC-IV Administration and Scoring Manual,* 2004). This is indicative of appropriateness of Pakistani scaled score norms.

The four index based composite scores are derived from sum of scaled scores of each index core subtests and they have a mean of 100 and *SD* of 15. Whereas, the composite score of FSIQ is derived from sum of scaled scores of all 10 core subtests. All the five composite scores (VCI, PRI, WMI, PSI, & FSIQ) for the Pakistani and UK samples seem quite comparable (see UK norms in WISC-IV Administration and Scoring Manual, 2004). A child's true score is more accurately represented by establishing a confidence interval of true scores. It is like a band of true scores in which the true score is likely to lie. The confidence interval for the FSIQ was found to be ± 7.8 and it was computed from Standard error of measurement.

The test-age equivalents represent the average age at which a given total raw score is typically obtained by children of a specific age and are considered very easy to understand. Beside their utility, test-age equivalent norms should be interpreted with caution and should not be recommended as primary interpretive scores (Wechsler, 2003). This is may be due to the fact that they are not comparable across subtests and provide little information about a child's standing relative to his or her same age peers.

Development of local/national norms are important as many studies have suggested that use of foreign norms may result in over or under estimation of one's IQ. An Indian study stated that Indian children scored significantly (16, 21, and 20 points respectively) lower on WISC-III Verbal IQ, Performance, and Full scale IQ when UK norms were used as compared to the scores when Indian norms were used (Paniker et al., 2006). Current study also found significant drop in VCI, PRI, PSI, and FSIQ composites when UK norms were used instead of Pakistani newly developed norms. On using regional norms FSIQ increased by 10 points, while for index based composites the maximum gains were found to be in VCI and PRI. Consistently, in another Pakistani study when scores of children were compared by using Pakistani scaled and VCI composites instead of UK norms, a 15 to 20 points increase in the VCI equivalent scores was observed. The gains were also quite significant for all the subtests on newly developed scaled scores with the highest gain for vocabulary (3 to 7 points) subtest (Ambreen, 2008).

Intelligence has strong heredity influences, but still it is not something that remains uninfluenced by an individual's surroundings or circumstances. Importantly, development of cognitive ability, performance, and skill acquisition can be influenced by several factors (Weiss et al., 2006). So the aim for phase IV of the study was to explore the relationship of age, gender, geographical area, and parental education level with the scores of children on WISC-IV ^{PAK} using Pakistani norms. Considering age, the performance of students showed significant differences on the 11 age groups. For composite scores and FSIQ, gradual improvement of performance with increasing age may be is indicative of increase in intelligence level with increasing age for the said age. This finding is again consistent with other researches on WISC-IV or its adaptations (see for example Dang et al., 2012). Considering the association of gender, results showed that gender has non significant effect on performance of boys and girls. This finding was consistent with findings of some other researches. Like, Rojhan and Naglieri (2005) found no meaningful gender differences in IQs of children from age group of 6-17 years. There were no gender differences between 6 and 9 years; females scored slightly higher between 10 and 13 years; and males were ahead of females between the ages of 15 and 16 but these differences were non significant. For the present study the sample comprising of school going children of similar schools and socio-economic level might have resulted in smaller gender differences on FSIQ, where girls and boys are having almost same cognitive opportunities, especially educational opportunities.

Living in or belonging to a particular geographical area with its culture, language, and level of advancement may influence children's performance on the IQ tests. But this was not what came out of current study as on FSIQ non-significant differences were found in groups based on geographical area. Again the types of school and the uniformity of curriculum might have resulted in small differences in performance of children from different geographical areas. Considering the relationship of parental education with student's performance, results indicated that the students from the group of parents having higher level of education performed significantly better than the students having parents with lower level of education. Highly educated parents can give more cognitive stimulation to their child resulting in better cognitive development. This finding was again consistent with many research findings including researches on WISC-IV, it was found that children's IQ test scores vary sharply and systematically with the level of education achieved by their parents and these two variables were found to be positively ($r = .43$) correlated (Weiss et al., 2006). In another study, Pearson (1969) found significant positive correlations between children's IQ and the mean parental education of the respective parents. So overall, age and parental education seems to influence student's performance on WISC-IV ^{PAK}.

The study III concludes the standardization of adapted WISC-IV subtests in Pakistan. Availability of a reliable and valid measure along with a strong interpretive framework would facilitate the assessment of children's intellectual abilities in the clinical and psycho-educational settings.

GENERAL DISCUSSION

General Discussion

The present research is aimed at adaptation and standardization of Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV) in Pakistan. The version of WISC-IV used in this research has been developed for use in South Asia, and this research has been conducted in collaboration with NCS Pearson, India with an aim of standardizing it for South Asia. The main objectives of the research include adaptation and/or translation of WISC-IV subtests, establishment of its reliability and validity evidence, and development of Pakistani norms for the adapted subtests. Besides that influence of age, gender, geographical area, and parental education was also to be explored on WISC-IV scores. These objectives were achieved through conduction of three studies with various phases of their own.

Intelligence testing has always been an area of interest in psychological measurement due to its importance in clinical as well as in psycho-educational assessments. Despite of the heredity influences on it, intelligence is considered as a culturally dependent construct (Cattell as cited in Marnat; Okagaki & Sternberg as cited in Cianciolo & Sternberg, 2004). Like in many other developing countries, intelligence testing is gaining attention in Pakistan with the progress in the fields of psychological research, mental health, and psycho-educational assessment. Though some of the world wide popular intelligence tests are being used (such as RPM; Raven, 1936) in these fields, but with the realization of concepts like test bias it has been realized that these foreign tests are not serving their full function in Pakistan. Then many intelligence tests have been locally developed for testing purposes in our country. Among these, most of the tests are nonverbal in nature (see Hashmi, 2000; Gardezi, 2001) while in psycho-educational assessment; verbal intelligence is of much importance. Some verbal intelligence tests were also developed but mostly for a very specific targeted population (see for example Hussain, 2001, Hussain et al., 2012). Many of these locally developed tests are lacking in their psychometric strength or are not properly validated and standardized. Furthermore, they do not follow the recent intelligence testing trends.

This necessitated the need of an indigenous or indigenized intelligence assessment tool. Indigenization of a tool for a certain culture involves both adaptation and standardization in that culture. Adaptation of tests have many advantages over test development as it limits duplication of efforts in test construction, save test developmental cost, help in achieving fairness in assessment, and facilitate comparative studies across cultures.

Intelligence testing in children has wide educational and clinical applications. WISC-IV is considered as one of the most widely used intelligence test for children throughout the world. Besides its strong psychometric strength and clinical utility, it has few limitations. Being a norm-referenced test, differences in performance on the test may be attributed to various cultural factors such as familiarity with the test stimuli; use of culturally-appropriate stimuli; inputs and training at home. WISC's limitations are being compensated through its various adaptations and revisions. With successive revisions, the Wechsler tests are changing from being considered measures of intelligence only to becoming more comprehensive test batteries of cognitive functions (Heaton, Taylor, & Manly as cited in Egeland et al., 2009). So it was decided to adapt WISC-IV in Pakistan.

Adaptation of WISC-IV

Pre-testing. Adaptation process was completed in first two studies of the research. It started with a pre-test of original version, its findings gave the empirical evidence to adapt and translate WISC-IV for Pakistani children. Pre-testing not only involved analysis of children's responses to the items but their feedback was also taken. Despite of its length, children took keen interest in test administration mainly because of the variety of tasks involved. Considering verbal tests, children found word reasoning and Arithmetic subtests as most interesting, and vocabulary subtest as most difficult. In vocabulary subtest, they commented few words as being totally unfamiliar, they also pointed out that the order of administration of few items was not appropriate. On similarities subtest, children initially showed anxiety and confusion but with increased understanding of what is required to answer they started showing interest and found most items appropriate. In comprehension subtest, length of few items became matter of concern for some students but mostly items were responded as appropriate. Comprehension subtest assesses a child's social judgment and common sense (Sattler, as cited in Beebe, McBurnett, & Pfiffner, 2000), on many of its items Pakistani children gave social judgments based on religious values as this is prevailing practice in our culture. In information subtest again few items were commented to be totally unfamiliar for the children of all ages. In Arithmetic subtests they showed concerns on unfamiliar names used in the items but shown their approval for the type of the items. Whereas, they took much interest in word reasoning items. Language problems were reported on almost all verbal subtests suggesting a strong need to translate the subtests. These findings were consistent with the results of children's responses to the items, poor response rate was also observed on items of the subtests that children pointed out to be less familiar, culturally inappropriate, or very difficult in relation to their order of administration. On performance subtests most items were found appropriate but children responded quite poorly on tasks involving speed.

Qualitative/Judgmental Steps. The evidence from pre-testing led to an iterative process of test adaptation involving cycles of translations, adaptations, field testing, and modifications in adaptation. As Van de Vijver and Tanzer stated that an appropriate translation needs a balanced treatment of psychological, linguistic, and cultural concerns (as cited in Sireci et al., 2006) so the adaptation committees consulted in present research were comprised of educationists, psychologists, and linguistic experts. A qualified 12 membered translation committee translated the examinee directed instructions of all subtests and items of only verbal subtest. Then a committee of four psychologists decided for best translations and adaptive changes. Keeping in mind the finding of pre-testing along with other cultural and psychological considerations committee decided to replace five information subtest items and one comprehension item. Sampled response options for few items was also replaced or modified and less familiar names in arithmetic subtest were replaced by common names. Importantly, for vocabulary subtest all new Urdu words along with their sampled responses were selected and compiles. Only one picture (in picture completion subtest) modification was suggested in performance subtests. As a quality check, all the adapted and translated subtests were then reviewed by a committee of linguistic experts (from National Language Authority of Pakistan) and psychologists having vast experience of test construction and administration. Minor modifications were made before finalizing the adapted subtests for field testing.

Quantitative Steps. As both a priori (judgmental) and posteriori (statistical) procedures were implied for adaptation in this research, so study II involved multiple tryouts to field test and modify the adapted subtests of WISC-IV. The requirement that both statistical and qualitative analyses be performed to validate adapted tests is asserted in both the ITC test adaptation guidelines (2010, 2012) and the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association & National Council on Measurement in Education as cited in Sireci et al., 2006). A rigorous psychometric evaluation of all adapted subtests was carried out implying both CTT and IRT techniques. CTT based techniques aim for improved reliability of the tests, whereas IRT analyses emphasize more on item level statistics. So, all the items were analyzed for item difficulty, item discrimination, and subtest reliability as well as for IRT based item fit indices. These tryouts resulted in more changes or modifications in WISC-IV content including reordering of items in nine subtests, minor modifications in translation of similarities subtest item, replacement of items in information, vocabulary, and arithmetic subtests, and replacement or modification in sampled response options of comprehension, vocabulary, and information subtest.

Overall, beside construction of new items for vocabulary subtest, translations of all items, and addition in sampled responses or reordering of items, only eight items from information, comprehension, and arithmetic subtest went through replacement with a new item. This small number of adaptive changes are due to the fact that this version of WISC-IV has already been adapted for use in South Asia by Pearson Assessment. Moreover, it is suggested that the test constructor should attempt to keep the number of adapted items as small as possible in order to avoid the risk of changing the construct, or not being able to establish its equivalence (Georgas, Van de Vijver, Weiss, & Saklofske as cited in Khaleefa, 2006). In current study, among the adaptive changes vocabulary subtest had the maximum changes, followed by information and comprehension subtest. This was also consistent with a large study aiming at comparison of WISC-III adaptation process of 16 countries. Researcher also found that vocabulary subtest has the highest number of adaptive changes followed by information and comprehension subtest (Georgas et al. as cited in Prifitera, Saklofske, and Weiss, 2008) So the first two studies finalized the adapted subtests and WISC-IV ^{PAK} was now considered to be ready for standardization.

Standardization of WISC-IV ^{PAK}

Standardization of any test involved establishment of test's psychometric evaluation and regional/national norms development. The standardization of WISC-IV ^{PAK} was accomplished in the third study of this research which was completed through four phases. For standardization a normative sample of 800 children from the age group of 6 to 16 years and 11 months was selected following a stratified random sampling design. The sample was stratified on 11 age groups of one year, two gender groups, five geographical region based groups, and three parental education based groups. The same stratification variables were selected for standardization sample of WISC-IV (Wechsler, 2003) with an addition of ethnic groups. Keeping in mind the issues related to lack of uniformity in school curriculum and differential influence of socioeconomic status, normative sample was selected from govt., semi govt., and private schools having similar status and instructional medium. A trained team of nine test administrators accompanied researcher in collecting data from all four provinces and capital territory of Pakistan. Some of the areas like Gilgit-Bultistan was avoided in data collection due to instable law and order situation in these areas. The proportion of sample belonging to various geographical regions was in close approximation to the population of that region according to 1998 census reports.

Establishment of Reliability and Validity Evidence of the WISC-IV ^{PAK} Subtests. For application in any practical fields, evaluation of a test's psychometric properties is a must. Psychometric strength of any test lies in its reliability and validity. Reliability of the test is concerned with stability of test scores and its internal consistency. Reliability of WISC-IV ^{PAK} subtest was established following all those measures employed in reliability establishment of original WISC-IV (Wechsler, 2003). Stability coefficients for all adapted subtests were quite significant and comparable with the stability coefficients of WISC-IV UK (Wechsler, 2004). The test-retest mean interval was of 35 days and in measurement practice a test-retest interval of fortnight to six months is considered appropriate (Anastasi, 1982). Being sensitive to the practice effect stability coefficients for WMI and PSI subtests were lower than VCI and PRI subtests. Considering internal consistency, both alpha coefficients and item-total correlations evidenced the reliability of all adapted subtests. These coefficients were though lower for 6-8year old children. Test administration rules like age based starting points and increasing level of item difficulty have resulted in poor item endorsement rate by the younger children as compared to the older ones. This poor response rate might have resulted in low reliability coefficients.

Considering validation, construct and factorial validity of WISC-IV PAK was established. For evidence of convergent and discriminant validity a Multitrait-Multimethod matrix method (Campbell & Fiske as cited in Wechsler, 2003) was used. This method implies verification of a set of literature and research based priori hypotheses about correlation of various variables/subtests with each other. Accordingly, all WISC-IV PAK subtests belonging to same index showed higher correlations (convergent evidence) than the correlations with subtest of other indices (discriminant evidence). Similarly, correlation matrix indicates high correlations among test having high g-loadings, this again establishes convergent validity of the subtests. Correlational patterns further prove split loadings of few subtests (such as arithmetic) as they have high correlations with both their own index and verbal comprehension index subtests. This split loading results from influence of verbal content on presentation or responding to the subtest item. Cross validation of WISC-IV factor structure was also done in this study. A four-factor structure was validated as was consistently found in many other researches on Wechsler recent scales (see for example, Watkins et al., in press; Chen et al., 2009). Considering the well researched importance of 'G', the general intelligence, a higher-order confirmatory factor analysis was also conducted. Findings were again consistent with many other researches (see for example, Nelson & Canivez as cited in Watkins et al., in press), a first order 'G' factor with four 2nd order factors provided the best fit. Presence of this general intelligence factor emphasizes the role of FSIQ in interpretation of WISC-IV. Over all, WISC-IV PAK seems to have acceptable psychometric strength to be used in further studies.

Norms Development for WISC-IV PAK in Pakistan. Norms provide an interpretive framework to raw scores. They are the yard sticks through which individual's performance can be interpreted and compared with reference to his/her respective population. An individual's performance on any task is influenced not only by demands of the task itself but also by the history and characteristics, the individual brings into the task and by the factors building in which the testing is carried out. Research comparing IQ scores of the American standardization sample when applied to children from Canadian standardization sample showed a significant increase in scores in the later group (Reynolds & Wilson as cited in Paniker et al., 2006). For this reason development of local norms for making interpretations becomes quite important.

For WISC-IV PAK, two types of norms were developed standard score (scaled score and composite score norms) and test-age equivalent norms following the same method that was used for the development of norms for the original version, as mentioned in the *WISC-IV Technical and Interpretive Manual*, 2003. With the application of appropriate scaling and normalization procedures the norms developed for Pakistani population proved much comparable to the norms of the original version. In both versions means and standard deviations for the scaled score and five composites (VCI, PRI, WMI, PSI, and FSIQ) closely approximated the pre-assumed means and standard deviations for scaled scores (M = 10, SD = 3) and composite scores (M = 100, SD = 15) of WISC-IV.

To assess the influence of Pakistani norms development on interpretations, students composite scores derived from Pakistani norms were compared with their composite score derived from the UK norms. Results indicated substantial gains in FSIQ (of 10 points) and in three of the index based composites VCI, PRI, and PSI (of 7 to 14 points). This finding was also consistent with the findings of an Indian study, which stated that Indian children scored significantly lower (16, 15, and 20 points on Verbal, performance, and Full scale IQ, respectively) when UK norms were used as compared to the scores when Indian norms were used (Paniker et al., 2006). This further supports the logic of developing local norms. These performance differences do not suggest the supremacy of one population over other. These findings should be viewed in the context of culture and previous experiences, which influences test performance in different countries. For interpretive considerations among the developed norms, standard scores are suggested to provide the most accurate descriptions of the test scores but for individuals who are unfamiliar with the test interpretation, standard scores would be difficult to understand. So other information such as test-age equivalents should be used in conjunction with standard scores (as cited in Wechsler, 2003)

Relationship of Age, Gender, Geographical region, and Parental Education with WISC-IV ^{PAK} Scores. An individual's level of intelligence is never independent of different personality and contextual variables. Development of cognitive abilities, performance, and skill acquisition is influenced by several variables or factors. So, demographic variables like age, gender, geographical residence, and parent's education can influence ones intellectual abilities. So the relationship of these variables with WISC-IV ^{PAK} scores of the children was also explored as the secondary aim of the present study. Considering age, the performance of students showed significant differences on the 11 age groups. The gradual improvement of performance on various subtests with increasing age is may be indicative of increase in intelligence level with increasing age. This progressive increase in scores with increasing age is in agreement with the validational theory of Binet and Simon put forward in their early work on the measurement of ability (as cited in Ahmad & Aziz, 1993). Jenson (1998) stated that due to largely genetic factors, people enter in school with 'wide variability in cognitive ability', and 'readiness to learn'. This initial variability then heavily influences (along with environmental factors) the amount a given student will pick-up as he or she matures. So as the student gains knowledge, his / her initial ability and newly acquired knowledge interact and result in improved performance (as cited in Watkins et al., 2006).

Considering the relationship of gender, belonging to a specific gender group do not seem to have any substantial influence on children's FSIQ. This is consistent with the research findings on WISC-IV, as it was found that there were non significant differences in performance of individuals among male and female children (Weiss et al., 2006). In another study, Rojahn and Naglieri (2005) found no meaningful developmental gender differences on children IQ, but they suggested that these differences may exist and identified when intelligence is defined as operationalized by the Cognitive Assessment System instead of verbal and performance aspects. Moreover, this negligible gender influence is may be due to the reason that, differences between the genders are generally quite small as compared to the individual differences among individuals of the same gender. Findings further indicate non significant influence of geographical area on children's FSIQ. This absence of meaningful differences in gender groups and geographical area in current study is may be due to the reason that sample is selected from a region where girls and boys are having almost same cognitive opportunities, especially educational opportunities. They are studying in schools of similar standards and having same syllabus, so their skill acquisition, performance and cognitive abilities are getting influenced by same environmental challenges and opportunities.

Considering the influence of parental education, results are providing evidence that parental education level has significant impact on student's FSIQ. A research finding on WISC-IV showed that mean FSIQ scores of children generally increase substantially with each subsequent level of education obtained by their parents (Weiss et al., 2006). In another study, it was found out that parental education exerts a modest shared environment effect on verbal intelligence of children. Both biological and adaptive influence of parental education on children verbal IQ was found to be statistically significant (Neiss & Rowe, 2000). The reason for significant impact of parent's education is two fold, first influence comes from the inherited level of cognitive ability passed from parents to the child and second influence is due to the cognitive stimulation educated parents provide to their children through the expectations, and home environment. Parental education itself is related to many other factors like economic status and quality of educational opportunities they give to their children. So parent's education indirectly represents a lot of variables related to the level of cognitive enrichment parents can provide to their child (Weiss et al., 2006).

This exploration concludes the efforts to adapt and standardize WISC-IV in Pakistan and that resulted in provision of a psychometrically strong indigenized intellectual assessment tool for children. This study is part of indigenization process of WISC-IV, it should not be considered as an end point. So, periodic review of it cultural appropriateness and psychometric strength is strongly recommended.

Implications of the Present Study

Outcomes of the present study in the form of well adapted and standardized WISC-IV ^{PAK} have implication in many fields. Intelligence has been shown repeatedly to be predictive of a wide variety of important life out comes (Gottfredson as cited in Weiss et al., 2006), so any measure of intelligence would have multifold importance. Due to its psychometric strength, comprehensive interpretation, and clinical utility, WISC-IV is a widely used intelligence test for children throughout the world along with its many adapted and translated versions. In Pakistan such highly standardized instrument was much needed for children with proper adaptation in clinical, educational, and research fields.

WISC-IV with its strong normative framework is widely used as an effective tool for clinical diagnosis and for making therapeutic suggestions. It can be used reliably for the diagnosis of mental retardation and learning disabilities. Moreover, by identifying a child's intellectual strengths and weaknesses, therapist can suggest management plans and treatments for various clinical disorders.

In the field of education, assessment of intellectual abilities especially 'general ability' is of great importance and researches have proved that measures of general ability can help in predicting the academic achievement of the students. School ability is itself viewed as g-demanding (Jenson as cited in Watkins et al., 2006). So, presence of a well adapted and standardized measure will facilitate the educational assessments. Different WISC-IV subtests measure different areas of intellectual functioning like verbal reasoning, verbal conceptualization, and verbal memory, perceptual reasoning, abstract reasoning, memory, and scanning speed etc, even only the subtest of Vocabulary is indicative of word recognition, semantic understanding, memory and retrieval, as well as expressive language skills. So psycho-educational assessments through WISC-IV ^{PAK} will help in identification of those intellectual abilities in which children are lacking. Then the fact that many of the WISC-IV subtests are measures of crystallized (cultural specific and learned) intelligence again has much clinical and educational implication, it implies that it is not an immutable trait and there is possibility of modification or improvement in the lacking intellectual abilities through different measures. For example, during the data collection and scoring, researcher herself observed that the sampled student showed much poor performance on vocabulary as compared to other subtests, knowing that, methods can be employed in the school settings to improve the vocabulary level of the students which in turn will result in enhancement of different areas of intellectual ability.

Considering the field of psychological and educational research, presence of a well adapted and standardized intelligence measure will facilitate researchers a lot. Especially, in case of cross cultural and comparative researches WISC-IV ^{PAK} can be used as a reliable and valid tool for information gathering. Adapted WISC-IV itself calls for further researches on its structure, validation, factorial make up, differential item functioning, and evidence of clinical utility. Similarly, different findings of the present study with reference to the influence of factors like age, gender, geographic region, and parental education on intelligence need to be explored further. Such finding offer many areas of future research.

Limitations and Recommendations

Besides much effort of the researcher for strong psychometric results from the adaptation and standardization of WISC-IV, present study has few limitations. As with all researches, results of the present study must be considered within the limitations of its sample and design. First of all sampling issue is of great importance. Despite of using a stratified random sampling plan, at school levels sampling became relatively less representative. Data was collected from similar kind of govt. and private school from all over the country. The schools can be considered similar as their medium of instruction, curriculum, fee structure, and other learning opportunities have a certain level of uniformity. Furthermore, due to certain situational constraints such as poor law and order condition, few areas of the country (like Gilgit-Baltistan) could not be accessed. So in future revisions, selection of a more representative sample from whole population of the country will help in improvement of standardization procedures.

Secondly, WISC-IV is a very long test, and working on its adaptation and standardization in a time and financially bound research activity of Ph.D. might have resulted in some kind of quality compromises. So exclusion of such time based and monetary constraints in future studies will enhance the work quality. Thirdly, due to time constraints certain areas like factors effecting FSIQ scores could not be explored deeply, deep exploration of such areas may lead to new findings, so future research is strongly needed. Similarly, establishment of its statistical linguistic equivalence is also lacking and is recommended for future research. Moreover, periodic review of WISC-IV ^{PAK} test content and psychometric properties is also strongly recommended.

The clinical and educational utility WISC-IV ^{PAK} can be enhanced by developing shorter/brief versions of the test. After careful validations in specific clinical and educational settings, these brief versions could be used as quick screening tools for learning disabilities and could also help in identification of student's educational strengths and weaknesses.

From the discussion above, it can be concluded that WISC-IV ^{PAK} has sufficient psychometric strengths to be used in the practical fields of education and research. But limitations of the study should also be kept in mind while making interpretations and generalizations.

REFERENCES

References

Abbas, I., & Israr, N. (1990). *Test of Intellectual Development for Pakistani pre-school children: Development and validation* (Technical Report). Islamabad: National Institute of Psychology.

Ahmed, I., Haque, A., & Anila. (1985). *Development of Urdu version of California Psychological Inventory in Pakistan* (Technical Report). Islamabad: National Institute of Psychology.

Ahmad, I., & Aziz, S. (1993). Adaptation of Columbia Mental Maturity Scale in Pakistan. *Pakistan Journal of Psychological Research, 8*(1-2), 31-41.

Ahmed, I., & Khan, S. (1984). *Adaptation of Columbia Mental Maturity Scale in Pakistan* (Technical Report). Islamabad: National Institute of Psychology.

Ahmed, M. (2000). *Emic perspective of intelligence in Punjabi culture* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Ain, Q. (1985). A validation study of Cattell's Culture Fair Intelligence Test on Pakistani children. In Z. A. Ansari, P. N. Tariq, R. M. Zaman, S. Chowdhri, & S. Watanabe (Eds.), *Proceedings of the Fifth Conference of Pakistan Psychological Association*, Islamabad: Pakistan Psychological Association.

Akhter, N., Hanif, R., Tariq, N., & Atta, M. (2011). Parenting Style as predictor of externalizing and internalizing behavior problems among children. *Pakistan Journal of Psychological Research, 26*(1), 23-41.

Akkok, F., & Askar, P. (1989). The adaptation and standardization of the teacher version of the Child Behavior Profile: Turkish boys aged 7-12. *International Journal of Psychology,24*, 129-136.

Alla-ud-Din, S. (2003). Relationship between ambivalent sexism, gender stereotyping, and sexual harassment among professionals. *Pakistan Psychological Abstract, 5.*

Allen, D. N., Thaler, N. S., Donohue, B., & Mayfield, J. (2010). WISC-IV Profiles in children with Traumatic Brain Injury: Similarities to and differences from the WISC-III. *Psychological Assessment, 22*(1), 57-64.

Ambreen, S. (2008). *WISC-IV UK: Adaptation and norms development in Pakistan* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Ambreen, S., & Kamal, A. (2011). Adaptation and validation of the Verbal Comprehension Index (VCI) Subtests of WISC-IV in Pakistan. *Journal of Behavioural Sciences, 21*(2). 18-34.

Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: McMillan Publishing.

Anastasi, A. (1990). *Psychological testing* (6th ed.). New York: McMillan Publishing.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). New Jersey: Prentice Hall.

Anila., Khan, S., & Pervez, S. (1991). Adaptation of Home Inventory (Infant version) for Pakistani children. *Pakistan Journal of Psychological Research, 6*(3-4), 103-115.

Ansari, Z. A. (1976). Creativity, intelligence and academic achievement. *Pakistan Journal of Psychology, 9*(3-4), 17-26, 44.

Ansari, Z. A., & Iftikhar, M. (1984). Validity of Raven's Standard Progressive Matrices for urban and rural school children in Pakistan. Part 1: Basic facts. *Psychology Quarterly, 19*, 14-27.

Ansari, Z. A., Tariq, P. N., & Iftikhar, M. (1990). *Educational Ability Test Level 5: Development and validation* (Technical Report). Islamabad: National Institute of Psychology.

Ashraf, S. (2010). *Women experience of husband's controlling behavior and their self-esteem* (Unpublished M.Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Austin, D. W., Reid, D. S., & Scott, W. C. (2007). Investigation of the WISC-III and WASI in clinical child population. *Canadian Journal of School Psychology, 22*(2), 249-254. Retrieved from http://scholar.google.com/scholar?q=%22author%3AW.c.+author%3Ascott%22

Axelrod, B. N., & Donders, J. (2002). Two-subtest estimation of WAIS-III factor index score. *Psychological Assessment, 14*(3), 360-364.

Axelrod, B. N., Ryan, J. J., & Ward, L. C. (2000). Confirmatory factor analysis of the WAIS-III standardization data. *Psychological Assessment, 12*(3), 341-345.

Aziz, S. (1997). *Adaptation of Columbia Mental Maturity Scale in Pakistan* (Technical Report). Islamabad: National Institute of Psychology.

Beebe, D. W., McBurnett, K., & Pfiffner, L. J. (2000). Evaluation of the validity of the WISC-III Comprehension and Picture Completion subtests as measures of social intelligence. *Psychological Assessment*, *12*(1), 97-101.

Binet, A., & Simmon, T. (1916). *The development of intelligence in children*. Baltimore: William and William.

Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychology, 15*(5), 417 – 424.doi: 10.1080/09297040802603661

Brooks, B. C. (2010). Seeing the forest for the trees: Prevalence of low scores on the Wechsler Intelligence Scale for Children, 4th edition (WISC-IV). *Psychological Assessment, 22*(3), 650-656.

Byrne, B. M. (2010). Structural Equation Modeling in AMOS, Basic concepts, applications, and programming (2nd Edition). New York: Routledge-Taylor & Francis.

Canivez, G. L., & Watkins, M. V. (2004). Temporal stability of WISC-III subtests composite: Strengths and weaknesses. *Psychological Assessment*, *16*(2), 133-138.

Canivez, G. L., & Watkins, M. W. (2010). Investigation of the Factor Structure of WAIS-IV: Exploratory and Higher Order Factor Analysis. *Psychological Assessment, 22*(4), 827-836. doi: 10.1037/a002042a

Chen, H. Y., Keith, T. Z., Chen, Y., & Chang, B. (2009). What does the WISC-IV measures? Validation of the scoring and CHC-based interpretative approach. *Journal of Research in education Sciences, 54*(3), 85-108.

Chishti, M. A. (2002). *Translation and adaptation of Revised NEO-PI* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Chowdhury, A. N. (1989). Bengali adaptation of Spielberger's STAI, Form X. *Journal of Personality and Clinical Studies*, *5*(2), 257-260.

Cianciolo, A. T. & Sternberg, R. J. (2004). *Intelligence: A brief history.* MA: Blackwell publishing.

Cohen, R. J., & Swerdlik, M. E. (2005). *Psychological testing and assessment: An introduction to tests and measurements* (6th ed.). New York: McGraw Hill.

Cowan, N., Baddeley, A. D., Elliott, E. M., & Norris, J. (2003). List composition and the word length effect in immediate recall: A comparison of localist and globalist assumptions. *Psychonomic Bulletin & Review, 10*, 74-79.

Coyne, I. (2003). *ITC Test adaptation guidelines*. Retrieved from http://www.intestcom.org/test_adaptation.htm

Crawford, J. R., Allan, K. M., Besson, J. A., Cochrane, R. H., & Stewart, L. E. (1990). A comparison of WAIS and WAIS-R in matched UK samples. *British Journal of Clinical Psychology*, *29*, 105-109.

Dang, H., Weiss, B., Pollack, A., & Nguyen, M. C. (2011). Adaptation of the Wechsler Intelligence Scale for Children (WISC-IV) for Vietnam. *Psychology Stud, 56*(4), 387-392. doi: 10.1007/s12646-011-0099-5

Domino, G., & Domino, M. L. (2006). *Psychological testing: An introduction* (2nd edition). New York: Cambridge University Press.

Egberink, I. J. L. (2010). Application of Item Response Theory to non-cognitive data (Unpublished Ph.D. Dissertation). University of Groningen, the Netherlands

Egeland, J., Bosnes, O., & Johansen, H. (2009). Structure of the Norwegian version of the WAIS-III in a Clinical Sample: The Arithmetic Problem. *Assessment;16*; 292. Retrieved from http://asm.sagepub.com/cgi/content/abstract/16/3/292

Erdodi, L. A., Richard, D. C. S., & Hopwood, C. (2009). The Importance of Relying on the Manual: Scoring Error Variance in the WISC-IV Vocabulary Subtest. *Journal of Psychoeducational Assessment*, 27; 374. Retrieved from http://jpa.sagepub.com/cgi/content/abstract/27/5/374

Flanagan, P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment*. Retrieved from http://books.google.com/books?id=kieth+2004 +essentials+of+%22WISC+IV%22&

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd Edition). London, California: SAGE.

Gardezi, A. H. (1994). *Development of a Non-verbal Test for adolescents* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Gardezi, A. H. (2001). *Development and standardization of an indigenous Non-verbal Test for adolescents* (Unpublished Ph.D. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Glaub, V. E., & Kamphaus, R. W. (1991). Construction of a non-verbal adaptation of the Stanford-Binet Fourth Edition. *Educational and Psychological Measurement, 51* (1), 231-241.

Gorsuch, R. L., Hildebrand, D. K., & Saklofske, D. H. (2000). Replication of the factor structure of the WAIS-III with a Canadian sample. *Psychological Assessment, 12*(4), 436-439.

Gregoire, J. (2001). Factor structure of the French adaptation of the WISC-III: Three or four factors? *.International Journal of Testing, 1*(3), 271-281. Retrieved from http://www.informaworld.com/smpp/content~content=a785034479~db=all~order=page

Grob, A., Petermann, F., Lipsius, M., Costan, J., Petermann, U., & Deseking, M. (2008). Difference in Swiss and Germen children's intelligence as measured by HAWIK-IV. *Swiss Journal of Psychology, 67*(2), 113-118.

Guido, J. J. (2009). Guido's Guide to PROC UNIVARIATE: A tutorial for SAS® udders, NESUG. 1-18.

Guilmette, T. J., Kennedy, M. L., & Queally, P. T. (2001). A comparison of the WISC-III and the Otis-Lennon School Ability Test with students referred for learning disabilities. *Journal of Psychoeducational Assessment, 19*, 239-244.

Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of the adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology.* Retrieved from http://www.testpublishers.org/Documents/journal0114.pdf

Hanif, R., & Pervez, S. (2003). Translation and adaptation of Teacher Stress Inventory. *Pakistan Journal of Psychological Research, 18*(1-2), 45-58.

Hashmi, M. A. (2000). *Standardization of an intelligence test for the middle level students* (Unpublished Ph.D. dissertation). Department of Education, Bahauddin Zakaria University,Multan, Pakistan. Retrieved from prr.hec.gov.pk/thesis/1202.pdf

Hassan, K. E. (2006). *Issues in test adaptation: The Lebanon Experience*. Retrieved from http://www.intestcom.org/Downloads//ITC2006Brussels/Session3-3.5EI.Hassan.pdf

Hopwood, C. J., & Richard, D. C. S. (2005). Graduate Student WAIS-III Scoring Accuracy Is a Function of Full Scale IQ and Complexity of Examiner Tasks. *Assessment,* 12; 445. Retrieved fromhttp://asm.sagepub.com/cgi/content/abstract/12/4/445

Hussan, R. (1981). A comparative study of the Pakistani bilingual and monolingual school children's performance in verbal and non-verbal test. *Pakistan Journal of Psychology, 12*(3-4), 3-12.

Hussain. L., Jamil, A., Siraji, M. J., & Maroof, K. (2012). Development and standardization of Intelligence test for children. *International Journal of Learning and Development, 2*(5), 190-202. doi: 10.5296/ijld.v2i5.2451

Hussain, S. S. (1993). *Development of Group Verbal Intelligence Test in Urdu for high school students* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Hussain, S. S. (2001). *Development, validation and standardization of a Group Verbal Intelligence Test in Urdu for adolescents* (Unpublished Ph.D. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Ijaz, M. (2007). *Urdu 5000 most frequently used-words*. Lahore: Center for Research in Urdu Language Processing, National University of Computer and Emerging Sciences.

Ikram, S. I. (2009). *Relationship between work-life balance and personality factors among male bank employees* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Imam, A., & Munaf, S. (1988). Birth order and intelligence. *Pakistan Journal of Psychology, 19*(1-4), 27-35.

International Test Commission (2010). *International Test Commission Guidelines for translating and adapting tests.* Retrieved from http://www.intestcom.org

International Test Commission (2012). *International Test Commission Guidelines for translating and adapting tests.* Retrieved from http://www.intestcom.org

Ismail, Z., & Mehmood, L. (1986). Social class and sex differences in Pakistan on Raven's Standard Progressive Matrices. *Pakistan Journal of Psychology, 17*(3-4), 23-35.

Israr, M. (1985). Cognitive development of Pakistani primary school children: Salient findings of a research project. In Z. A. Ansari; P. N. Tariq; R. M. Zaman; S. Chowdhri; and S. Watanabe (Eds.), *New directions in Pakistani psychology: proceedinds of the fifth Conference of Pakistan Psychological Association* (pp. 67-75) Islamabad: Pakistan Psychological Association.

Israr, M. (1988). *A psychological interpretation of mathematics learning problems among secondary school students* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Jamal, S. (1964). *The validity of Raven's Colored Progressive Matrices Test and Otis Quick Scoring Mental Ability Test for Pakistani children* (Unpublished Masters dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Kaplan, R. M., Saccuzzo, D. P. (2001). *Psychological testing: principles, applications and issues* (5th ed.). Belmont: Wadsworth/ Thomson Learning.

Kausar, R. (1998). *Validity study of Raven's Colored Progressive Matrices in Pakistan* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Kausar, R. (2007). *A comparison of public and private school children of 6th and 7th class on Information and Arithmetic subtests of WISC: A normative approach* (Unpublished M. Phil. dissertation). Unit of Clinical Psychology, Govt. College University, Lahore, Pakistan.

Kaushik, S. S., & Sheikh, K. (1990). Hindi adaptation of Social Avoidance and Distress Scale. *Journal of Personality and Clinical Studies, 6*(2), 237-241.

Keung, H., & Leung, M. C. (1990). The adaptation of the Family Environment Scale to Chinese children and adolescents in Hong Kong. *International Journal of Psychology, 25*(4), 545-555.

Khaleefa, O. (2006). Adaptation of the WISC-111 in Sudan and Japan: A cross-cultural study. *Arab psynet e. Journal, 12* (Electronic version).

Khan, M. J. (2008). *Relationship of parental mental health with children behavioral problems and role of moderating factors* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Khan, R. A. (2008). *Development and validation of Self-Report Measure of Emotional Intelligence (SRMEI) for heart patients and healthy individuals* (Unpublished Ph.D. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Khan, S. (1992). Translation and adaptation of a Vocational Interest Inventory: App Ki Dilchaspian. *Pakistan Journal of Psychological Research, 7(3-4),* 75-86.

Kiff, J. (2010). *Categories of psychometric measures.* Retrieved from psychology.wikia.com/wiki/Psych-Assess:Types_of_test

Kline, P. (1991). Intelligence. *The psychometric view.* London: Routledge.

Leo, S. A., Kadlubek, R. M., & Marks, W. J. (2007). Administration and Scoring Errors on the WISC-IV among Graduate Student Examiners. *Journal of Psychoeducational Assessment*; 25. Retrieved from http://jpa.sagepub.com/cgi/content/abstract/25/3/237

Linacre, J. M. (2001). *A User's Guide to WINSTEPS - MINISTEPS Rasch-Model Computer Programs (Program Manual (3.71.0).* Winsteps.com

Longman, R. S., & Saklofske, D. H. (2007). WAIS-III Percentile Scores by education and sex for U.S. and Canadian populations. *Assessment,* 14; 426. Retrieved from http://asm.sagepub.com/cgi/content/abstract/14/4/426

Looona, M. I. (2002). Academic performance and school social behavior of Attention Deficit Hyperactivity Disorder (ADHD) and non-ADHD children. *Pakistan Psychological Abstract, 5.*

Loona, M. I., & Kamal, A. (2011). Translation and adaptation of Disruptive Behavior Disorder Rating Scale. *Pakistan Journal of Psychological Research, 26*(2). 149-165

Mahmood, Z. (1991). Intelligence, IQ, and the third world. *Pakistan Journal of Psychological Research*, *6*(1-2), 31-53.

Mahmud, N. (1990). *Development and validation of Educational Ability Test for Pakistani pre-school children* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Malda, M., Van de Vijver, F. J. R., Sirinivasan, K., Transler, C., Sukumar, P., & Rao, K. (2008). Adapting a cognitive test for a different culture: An illustration of qualitative procedures. *Psychology Science Quarterly, 50*(4), 451-468.

Manzoor, A. (2000). Relationship between early recollections and personality characteristics. *Pakistan Psychological Abstract, 5*.

Marnat, G. G. (1990). *Hand book of psychological assessment* (2nd ed.), New York: John Wiley and Sons.

Mayes, S. D., & Calhoun, S. L. (2006). WISC-IV and WISC-III Profiles in Children with ADHD. *Journal of Attention disorder, 9*; 486. Retrieved from http://jad.sagepub.com/cgi/content/abstract/9/3/486

McGrew, K. S., Taub, G. E., & Witta, E. L. (2004). A confirmatory analysis of the factor structure and cross age invariance of the Wechsler Adult Intelligence Scale- Third Edition. *Psychological Assessment, 16*(1), 85-89.

Montes, L. E. S. M., Puente, A. E., Allen, D. N. & Neblina, C. (2010). Validity of the WISC-IV Spanish for a clinically referred sample of Hispanic children. *Psychological Assessment, 22*(2), 465-469.

Naheed, G. (1993). *Development of a Verbal Test of Intelligence for Pakistani urban primary school children* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Naqvi, I. (2007). *Patterns of delinquency and personality traits of adolescents in child labour* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Neiss, M., & Rowe, D. C. (2000). Parental education and child's verbal IQ in adoptive and biological families in the national longitudinal study of adolescent health. *Behavior Genetics, 30*(6), 487-495. Retrieved from http://www.ingentaconnect.com/content/klu/bege/2000/00000030/00000006/00303266

Noor, A., & Najam, N. (2009). Comparative Study of memory deficits in Younger and Older Adults. *Pakistan Journal of Psychological Research, 24*(3-4), 145-158.

Oakland, T. & Harris, J. G. (2009). Impact of test-taking behaviors on Full-Scale IQ scores from the Wechsler Intelligence Scale for Children IV Spanish Edition. *Journal of Psychoeducational Assessment,* 27; 366. Retrieved from http://jpa.sagepub.com/cgi/content/abstract/27/5/366

Population Census Organization (n.d.). Pakistani Population by Province/Region since 1951. Retrieved from http://www.census.gov.pk

Panicker, A. S., Hirisave, V., & Subbakrishna, D. K. (2006). WISC-III UK: Comparison of Indian and UK norms. *Journal of Indian Associates for Child Adolescents Mental Health*, *2*(4), 108-111. Retrieved from http://www.jiacam.org/0204/WISC_III.pdf

Park, H. M. (2008). *Univariate analysis and normality test using SAS, STATA and SPSS* (Working Paper). The University Information Technology Services (UITS) Centre for Statistical and Mathematical Computing, Indiana University. Retrieved from http://www.indiana.edu/~statmath/stat/all/normality/index.html

Price, L. R., & Tulsky, P. S. (2003). The joint WAIS-III and WMS-III factor structure: Development and cross-validation of a six factor model of cognitive functioning. *Psychological Assessment*, *15*(2), 149-162.

Prifitera, A., Saklofske, D. H., & Weiss, L. G. (Eds.). (2008). *WISC-IV Clinical Assessment and Intervention (2nd edition)*. New York: Elsevier

Rahman, N. K., & Saleem, A. (1992). Urdu adaptation and standardization of Irrational Beliefs Test on Pakistani population and development of norms for psychiatric, drug addict and normal populations. *Pakistan Journal of Psychology*, *23*(3-4), 17-27.

Rana, I. A. (1995). *A cognitive investigation of learning disabilities in Pakistan* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Raza, A., & Sheikh, M. H. (1991). Translation and adaptation of Children's Personality Questionnare (CPQ) of the IPAT series into Urdu and development of local norms for Pakistani population. *Pakistan Journal of Psychology,22*(3-4), 25-37.

Reiter, B. A. (2004). The application of a WISC-III short form for screening gifted elementary students in Canada. *Canadian Journal of School Psychology, 19*(1-2), 191-203. doi: 10.1177/082957350401900110

Riaz, M. N. (1979). A study of intelligence-creativity distinction and their relationship with academic achievement. *Pakistan Psychological Studies*, *3*, 38-70.

Rizvi, S. S. (2009). *Father's Masculinity Ideology and their adolescent's perception of Fatherhood* (Unpublished M. Phil .dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Saklofske, D. H., Tulsky, D. S., Weiss, L. G., & Wilkins, C. (2001). Development of a General Ability Index for the WAIS-III. *Psychological Assessment, 13*(4), 566-571.

Saklofske, D. H., Zhu, J., Coalson, D. L., Raiford, S. E., & Weiss, L. G. (2010). Cognitive Proficiency Index for the Canadian edition of the Wechsler Intelligence Scale for Children, Fourth Edition. *Journal of School Psychology, 25*(3). 277-286. doi:10.1177/0829573510380539

Sanchez-Escobedo, P., Hollingworth, L., & Fina, A. D. (2011). A cross-cultural, comparative study of the American, Spanish, and Mexican versions of the WISC-IV. *Tesol quarterly, 45*(4), 781-792. doi: 10.5054/tq.2011.268057

Sans, M. C. (1984). *Adaptation of Wechsler Intelligence Scales for Children (WISC) to Argentinian subjects.* Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6535384

Scott, W. C., Austin, D. W., & Reid, D. S. (2007). Investigation of the WISC-III and WASI in clinical child population. *Canadian Journal of School Psychology, 22*(2), 249-254. doi: 10.1177/0829573507308162

Scott, R. L., Lucio, G. M., & Reyes-Lagunes, E. (1994). MMPI-2 Mexico: Translation and adaptation. *Journal of Personality Assessment, 63*(1), 105-116.

Shahid, H. (2010). *Perception of social support, marital satisfaction and coping strategies among infertile couples* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Shamama-Tus-Sabah, S. (2010). *Chaos at home and the relationship with cognitive ability and socio-emotional adjustment of school* (Unpublished Ph.D. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Shamama-Tus-Sabah, S., Gilani, N., Iftikhar, R. (2012). Raven Progressive Matrices: Psychometric evidence, gender and social class differences in middle childhood. *Journal of Behavioral Sciences, 22*(3).

Sheikh, M. H. (1982). *Zahanat Paimana: translation and adaptation of Otis Quick Scoring Mental Ability Test (Form C&D)* (Unpublished M. Phil. Dissertation) Department of Applied Psychology, University of Punjab, Lahore.

Sireci, S. G., Yang, Y., Harter, J., & Ehrlich, E. J. (2006). Translation Quality Evaluating Guidelines For Test Adaptations: A Methodological Analysis of Translation's Quality. *Journal of Cross-Cultural Psychology*; 37; 557. Retrieved from http://jcc.sagepub.com/cgi/content/abstract/37/5/557

Sternberg, R. J., Lautrey, J., & Lubart, T. I. (Eds.). (2003). *Models of Intelligence: International perspective*. Washington: American Psychological Association.

Storfer, M. D. (1990). *Intelligence and Giftedness: The contribution of heredity and early environment*. New York: Jossy-Bass.

Syed, N. (1993). *Development of a Non-Verbal Test of Intelligence for Pakistani urban primary school* (Unpublished M. Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Tareen, K. (1982). *WISC-R (Adapted version), Bachoon Ke Zahanat Ka Paimana.* Lahore: National Institute of Mentally and Emotionally Disabled Children.

Tempest, P. (1998). Local Navajo norms for the Wechsler Intelligence Scale for Children-Third Edition. *Journal of American Indian Education, 37*(3). 140-151

Watkins, M. W. (2010). Structure of WISC-IV among a national sample of referred students. *Psychological Assessment, 22*(4), 782-787.

Watkins, M. W., Canivez, G. L., James, T., James, K., & Good, R. (In press). Construct validity of WISC-IV UK with a large referred Irish Sample. *International Journal of School and Education Psychology*.

Watkins, M. W., Lei, P. W., & Canivez, G. L. (2006). Psychometric intelligence and achievement: A cross-lagged panel analysis (Electronic version). *Intelligence*, 35(1), 59- 68.

Watkins, M. V., Wilson, S. M., Kotz, K. M., Carbone, M. C., & Babula, T. (2006). *Educational & Psychological Measurement*, *66*(6), 975-983. Retrieved from http://epm.sagepub.com/cgi/reprint/66/6/975

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence- WASI.* Psychological Corporation. Retrieved from http://cjs.sagepub.com/cgi/content/abstract/22/2/249

Wechsler, D. (2003). *WISC-IV Technical and Interpretive Manual,* SA: Harcourt Assessment.

Wechsler, D. (2004). *WISC-IV UK Administration and Scoring Manual,* London: Harcourt Assessment.

Wechsler, D. (2005). *WISC-IV Spanish.* Retrieved from http://harcourtassessment.com/HAIWEB/cultures/en-u/Productdetail.htm?pid+015-8978-846&Mode=summary

Wechsler, D. (2007). HAWIK-IV. Retrieved from www.pearsonassessment.de/wisc-iv-ehem-HAWIK-IV.html

Wechsler, D. (2012). *Wechsler Intelligence Scale for Children, Fourth edition (India) (WISC-IV ^{INDIA})*. Retrieved from http//pearsonclinical.in/WISC-IV-India.php

Weinberg, R. A. (1989). Intelligence and IQ: Landmark issues and great debates. *American Psychologist*, *44*, 98-104.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WAIS-IV and clinical validation of four and five-factor interpretative approach. *Journal of Psychoeducational Assessment, 31*(2), 94-113. doi: 10.1177/0734282913478030

Weiss, L. G., Prifitera, A., Saklofske, D. H., & Holdnack, J. A. (2006). *WISC-IV Advanced Clinical Interpretation.* Oxford: Elsevier.

WISC-IV. Retrieved from http://en.wikipedia.org/wiki/wechsler_intelligence_scalefor_children

WISC-IV Australian. (2003). Retrieved from http://www.harcourt-au.com/default.asp?action+article&ID+129

WISC-IV Canadian. (2004). *Tool Review.* Retrieved from http://www.cup.ualberta.ca/component/option,com_docman/task,doc_view/gid,77/itemid,204/-

WISC-IV Spanish. (2005). *News Letter.* Retrieved from http://www.harcourt.com/about/news/articles/2005/072605_assessment_wechsler_spanish.pdf

WISC-IV Urdu Standardization Edition (n.d.). *Administration and Scoring Manual* (Unpublished).

Yopp, J. H., & Brown, C. R. (1999). *Reducing cultural boundaries: Initiatives in quality assurance, curricular reform and test adaptation.* Retrieved from http://www.ineer.org/Events/ICEE2000/Proceedings/papers/WA6-3.pdf

Yu, H. C. (2012). *A Simple Guide to the Item Response Theory (IRT)*. Retrieved from http://www.creative-wisdom.com

Yusuf, S. (2009). Psychological stress, Type-I personality traits and stigmatization among Hepatitis C patients (Unpublished M.Phil. dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

Zahid, G. & Pervez, S. (2009). Translation, Adaptation, and Validation of Children's Action Tendency Scale. *Pakistan Journal of Psychological Research, 24*(3-4), 127-143.

Zander, E., & Dahlgren, S. O. (2010). WISC-III Index score Profiles of 520 Swedish children with Pervasive Developmental Disorders. *Psychological Assessment, 22*(2), 213-222.

Zimmaro, D. M. (2003). *Test Item Analysis & Decision Making*. Austin: Center for Teaching and Learning. Retrieved fromnwww.utexas.edu/academic/ctl

Zoofashan, T. (1982). *Computer model for statistical analysis of Progressive Matrices*(Unpublished Masters dissertation). National Institute of Psychology, Quaid-e-Azam University, Islamabad, Pakistan.

APPENDICIES

Appendix A

INFORMED CONSENT FORM

National Institute of Psychology (NIP), Quaid-e-Azam University, Islamabad has taken up a research project regarding Translation and Standardization of Wechsler Intelligence Scale for Children (WISC-IV South Asia) in collaboration with Pearson Assessment (NCS Pearson India private limited, Bangalore Office). This is part of a larger project aiming at establishing Urdu norms for WISC-IV South Asia that can be used for Urdu speaking children of 6 to 16 years 11 month of age throughout the world. WISC-IV is considered as one of the most widely used test for the clinical and psycho-educational assessment of children in the world.

I am (Saima Ambreen, Ph.D. Scholar, NIP) handling this project as part of my Ph.D. research work titled as **Wechsler Intelligence Scale for Children (WISC-IV): Adaptation, Translation and Standardization in Pakistan**. For field-testing of adapted test and for establishing norms we need to administer the above mentioned test on a nationwide representative sample of school going children from the age group of 6 to 16 years 11 month. For the above mentioned purpose I require data on a large scale (more than 1000 students). I will administer the above mentioned subtests on all the participating students individually with the help of my trained team and each individual administration will take 90 to 120 minutes.

I request you to support my purpose and participate in this research project. I assure you that the information taken from you will be kept confidential and will be used only for the research purposes. You have the right to withdraw your provided information at any stage of the research.

Your help support and participation will be highly appreciated.

Thank You!

INFORMED CONSENT

Name (Optional): ______________________

Institute: ______________________________

Contact no. (Optional): ____________________

I am willing to participate in this research.

____

Signature

Please contact if any query:

Saima Ambreen,

M.Phil – IV,

National Institute of Psychology (NIP),

Quaid-i-Azam University, Islamabad.

e.mail address: saima_ambreen_awan@hotmail.com

Phon: no. 051-90644111

Appendix B1

WISC-IV Record Form (English)

India Standardization Edition

Examiner's Name:________________ Site ID Number:______ Region:__________

Examinee's Name:____________________ Examinee's ID Number:__________

Teacher's Signature____________________

	Year	Month	Day
Today's Date			
Date of Birth			
Age at testing			

Gender	Female	Male	
Handedness	Left	Right	Both
Colorblind?	No	Yes	

Does the examinee have any disabling conditions? No Yes*

If yes, please explain how the examinee meets inclusion criteria in the box below No Yes

Does the examinee wear prescription lenses, glasses, or a hearing aid? No Yes*

*If yes, was he/she wearing them during testing? No Yes

1. Block Design (Time Limit: See Item)

Start Time___:___ Hr. Min.

Start
Ages 6–7: Item 1
Ages 8–16: Item 3

Reverse
Ages 8–16: Score of 0 or 1 on *either* of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 3 consecutive scores of 0

Score
Items 1–3: Score 0, 1, or 2 points
Items 4–8: Score 0 or 4 points
Items 9–14: Score 0 or the appropriate time bonus score
BDN
Items 1–3: Score 0, 1, or 2 points
Items 4–14: Score 0 or 4 points

Design	Presentation Method	Time Limit (Seconds)	Completion Time	Correct Design	Constructed Design	Score
6–7 → 1. Child / Examiner	Model	30		Y N	Trial 1 Trial 2	Trial 2 Trial 1 0 1 2
2.	Model	45		Y N	Trial 1 Trial 2	Trial 2 Trial 1 0 1 2
8–16 → 3.	Model and Picture	45		Y N	Trial 1 Trial 2	Trial 2 Trial 1 0 1 2
4.	Picture	45		Y N		0 4
5.	Picture	45		Y N		0 4
6.	Picture	75		Y N		0 4
7.	Picture	75		Y N		0 4
8.	Picture	75		Y N		0 4
9.	Picture	75		Y N		31-75 21-30 11-20 1-10 0 4 5 6 7
10.	Picture	75		Y N		31-75 21-30 11-20 1-10 0 4 5 6 7
11.	Picture	120		Y N		71-120 51-70 31-50 1-30 0 4 5 6 7
12.	Picture	120		Y N		71-120 51-70 31-50 1-30 0 4 5 6 7
13.	Picture	120		Y N		71-120 51-70 31-50 1-30 0 4 5 6 7
14.	Picture	120		Y N		71-120 51-70 31-50 1-30 0 4 5 6 7

Stop Time___:___ Hr. Min.

Total Raw Score (Maximum = 68)

Block Design No Time Bonus (BDN) Total Raw Score (Maximum = 50)

2. Similarities

Start Time___:___
Hr. Min.

Start
Ages 6–8: Sample, then Item 1
Ages 9–11: Sample, then Item 3
Ages 12–16: Sample, then Item 5

Reverse
Ages 9–16: Score of 0 or 1 on *either* of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 5 consecutive scores of 0

Score
Items 1–2: Score 0 or 1 point
Items 3–23: Score 0, 1, or 2 points
See *Administration and Scoring Manual* for sample responses.

Start	Item	Response	Score
6–16	S. Red–Blue		
6–8	†1. Milk–Water		0 1
	†2. Pen–Pencil		0 1
9–11	3. Apple–Banana		0 1 2
	4. Shirt–Shoe		0 1 2
12–16	5. Cat–Mouse		0 1 2
	6. Butterfly–Bee		0 1 2
	7. Winter–Summer		0 1 2
	8. Anger–Joy		0 1 2
	9. Elbow–Knee		0 1 2
	10. Timber–Bricks		0 1 2
	11. Painting–Statue		0 1 2
	12. Frown–Smile		0 1 2
	13. Ice–Steam		0 1 2
	14. Poet–Painter		0 1 2
	15. Flood–Drought		0 1 2
	16. Mountain–Lake		0 1 2
	17. First–Last		0 1 2
	18. Revenge–Forgiveness		0 1 2

† If the child does not give a 1-point response, provide the response indicated in the *Administration and Scoring Manual.*

Continue

2. Similarities *(Continued)*

Discontinue after 5 consecutive scores of 0

Item	Response	Score
19. Rubber–Paper		0 1 2
20. Permission–Limitation		0 1 2
21. Salt–Water		0 1 2
22. Reality–Dream		0 1 2
23. Space–Time		0 1 2

Stop Time___:___ Hr. Min.

Total Raw Score (Maximum = 44)

Stop Time___:___ Hr. Min.

3. Digit Span

Start
Ages 6–16:
Forwards: Item 1
Backwards: Sample, then Item 1

Discontinue
Forwards: After scores of 0 on *both trials* of an item
Backwards: After scores of 0 on *both trials* of an item

Score
Score 0 or 1 point for each trial
DSF & DSB
Total Raw Score for DS Forwards and Backwards, respectively
LDSF & LDSB
Number of digits recalled on last trial scored 1 point for DS Forwards and Backwards, respectively

Forwards

Trial		Response	Trial Score	Item Score
6–16 1.	2 – 9		0 1	0 1 2
	4 – 6		0 1	
2.	3 – 8 – 6		0 1	0 1 2
	6 – 1 – 2		0 1	
3.	3 – 4 – 1 – 7		0 1	0 1 2
	6 – 1 – 5 – 8		0 1	
4.	8 – 4 – 2 – 3 – 9		0 1	0 1 2
	5 – 2 – 1 – 8 – 6		0 1	
5.	3 – 8 – 9 – 1 – 7 – 4		0 1	0 1 2
	7 – 9 – 6 – 4 – 8 – 3		0 1	
6.	5 – 1 – 7 – 4 – 2 – 3 – 8		0 1	0 1 2
	9 – 8 – 5 – 2 – 1 – 6 – 3		0 1	
7.	1 – 8 – 4 – 5 – 9 – 7 – 6 – 3		0 1	0 1 2
	2 – 9 – 7 – 6 – 3 – 1 – 5 – 4		0 1	
8.	5 – 3 – 8 – 7 – 1 – 2 – 4 – 6 – 9		0 1	0 1 2
	4 – 2 – 6 – 9 – 1 – 7 – 8 – 3 – 5		0 1	

Backwards

Trial		Response	Trial Score	Item Score
6–16	8 – 2			
	5 – 6			
1.	2 – 1		0 1	0 1 2
	1 – 3		0 1	
2.	3 – 5		0 1	0 1 2
	6 – 4		0 1	
3.	5 – 7 – 4		0 1	0 1 2
	2 – 5 – 9		0 1	
4.	7 – 2 – 9 – 6		0 1	0 1 2
	8 – 4 – 9 – 3		0 1	
5.	4 – 1 – 3 – 5 – 7		0 1	0 1 2
	9 – 7 – 8 – 5 – 2		0 1	
6.	1 – 6 – 5 – 2 – 9 – 8		0 1	0 1 2
	3 – 6 – 7 – 1 – 9 – 4		0 1	
7.	8 – 5 – 9 – 2 – 3 – 4 – 6		0 1	0 1 2
	4 – 5 – 7 – 9 – 2 – 8 – 1		0 1	
8.	6 – 9 – 1 – 7 – 3 – 2 – 5 – 8		0 1	0 1 2
	3 – 1 – 7 – 9 – 5 – 4 – 8 – 2		0 1	

LDSF Max = 9

Digit Span Forwards (DSF) Total Raw Score (Maximum = 16)

LDSB Max = 8

Digit Span Backwards (DSB) Total Raw Score (Maximum = 16)

Stop Time___:___ Hr. Min.

Total Raw Score (Maximum = 32)

4. Picture Concepts

Start Time ___:___ Hr. Min.

Start
Ages 6–8: Samples A & B, then Item 1
Ages 9–11: Samples A & B, then Item 5
Ages 12–16: Samples A & B, then Item 7

Reverse
Ages 9–16: Score of 0 on *a* of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 5 consecutive scores of 0

Score
Score 0 or 1 point
Correct responses are in colour

	Item	Response	Score
6–16	A.	1 2 | 3 4 DK	
	B.	1 2 | 3 4 DK	
6–8	1.	1 2 | 3 4 DK	0 1
	2.	1 2 | 3 4 DK	0 1
	3.	1 2 | 3 4 DK	0 1
	4.	1 2 | 3 4 DK	0 1
9–11	5.	1 2 | 3 4 DK	0 1
	6.	1 2 3 | 4 5 6 DK	0 1
12–16	7.	1 2 3 | 4 5 6 DK	0 1
	8.	1 2 3 | 4 5 6 DK	0 1
	9.	1 2 3 | 4 5 6 DK	0 1
	10.	1 2 3 | 4 5 6 DK	0 1
	11.	1 2 3 | 4 5 6 DK	0 1

Item	Response	Score
12.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
13.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
14.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
15.	1 2 3 | 4 5 6 DK	0 1
16.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
17.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
18.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
19.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
20.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
21.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
22.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
23.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
24.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
25.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
26.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
27.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
28.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1

Stop Time ___:___ Hr. Min.

Total Raw Score (Maximum = 28)

5. Coding

(Time Limit: 120 seconds)

Start Time ___:___ Hr. Min.

Start
Ages 6–7: Coding A Sample Items, then Test Items
Ages 8–16: Coding B Sample Items, then Test Items

Discontinue
After 120 seconds

Score
Use the Scoring Key to check the child's responses.
Score 1 point for each correct response

	Form	Time Limit (Seconds)	Completion Time	Total Raw Score
6–7	A.	120		Max = 65
8–16	B.	120		Max = 119

Coding A Time-Bonus Scores for Perfect Performance

Time in Seconds	116–120	111–115	106–110	101–105	96–100	86–95	≤85
Score	59	60	61	62	63	64	65

Stop Time ___:___ Hr. Min.

6. Vocabulary

Start Time ___:___ Hr. Min.

Start Ages 6–8: Item 5; Ages 9–11: Item 7; Ages 12–16: Item 9

Reverse Ages 6–16: Score of 0 or 1 on *either* of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue After 5 consecutive scores of 0

Score Items 1–4: Score 0 or 1 point. Items 5–36: Score 0, 1, or 2 points. See *Administration and Scoring Manual* for sample responses.

Item	Response	Score
Picture Items		
1. Car		0 1
2. Flower		0 1
3. Train		0 1
4. Bucket		0 1
Verbal Items		
6–8 †5. Hat		0 1 2
†6. Umbrella		0 1 2
9–11 7. Clock		0 1 2
8. Cow		0 1 2
12–16 9. Thief		0 1 2
10. Bicycle		0 1 2
11. Alphabet		0 1 2
*12. Leave		0 1 2
13. Ancient		0 1 2
*14. Bark		0 1 2
15. Brave		0 1 2
16. Obey		0 1 2
17. Island		0 1 2

† If the child does not give a 2-point response, provide the response indicated in the *Administration and Scoring Manual*.
* Responses requiring specific query are identified in the *Administration and Scoring Manual*.

Continue

6. Vocabulary *(Continued)*

Discontinue after 5 consecutive scores of 0

Item	Response	Score
18. Nonsense		0 1 2
19. Absorb		0 1 2
20. Transparent		0 1 2
*21. Precise		0 1 2
22. Mimic		0 1 2
23. Fable		0 1 2
24. Migrate		0 1 2
25. Rivalry		0 1 2
26. Foresight		0 1 2
27. Seldom		0 1 2
28. Strenuous		0 1 2
29. Unanimous		0 1 2
*30. Imminent		0 1 2
31. Amendment		0 1 2
32. Compel		0 1 2
*33. Affliction		0 1 2
34. Garrulous		0 1 2
35. Dilatory		0 1 2
36. Aberration		0 1 2

* Responses requiring specific query are identified in the *Administration and Scoring Manual.*

Total Raw Score (Maximum = 68) ____

Stop Time___:___
Hr. Min.

7. Letter–Number Sequencing

Start Time ___:___
Hr. Min.

Start
Ages 6–7: Qualifying Items, Sample Item, then Item 1
Ages 8–16: Sample Item, then Item 1

Discontinue
If child is unable to respond correctly to either Qualifying Item *or* after scores of 0 on *all three trials* of an item.

Score
Score 0 or 1 point for each trial

	Qualifying Item	Correct Response	Correct
6–7	Counting	Child counts to three.	Y N
	Alphabet	Child recites alphabet to the letter C.	Y N

Item		Trial	Correct Responses		Verbatim Response	Trial Score	Item Score
8–16 S	1.	A - 2	2 - A	A - 2			
	2.	B - 3	3 - B	B - 3			
1.	1.	A - 3	3 - A	A - 3		0 1	0 1 2 3
		If the child responds A- 3, immediately correct the child as instructed in the manual.					
	2.	B - 1	1 - B	B - 1		0 1	
	3.	2 - C	2 - C	C - 2		0 1	
2.	1.	C - 4	4 - C	C - 4		0 1	0 1 2 3
	2.	5 - E	5 - E	E - 5		0 1	
	3.	D - 3	3 - D	D - 3		0 1	
3.	1.	B - 1 - 2	1 - 2 - B	B - 1 - 2		0 1	0 1 2 3
	2.	1 - 3 - C	1 - 3 - C	C - 1 - 3		0 1	
	3.	2 - A - 3	2 - 3 - A	A - 2 - 3		0 1	
4.	1.	D - 2 - 9	2 - 9 - D	D - 2 - 9		0 1	0 1 2 3
	2.	R - 5 - B	5 - B - R	B - R - 5		0 1	
		If the child responds 5-R-B or R-B-5, say, **Remember to say the letters in order.**					
	3.	H - 9 - K	9 - H - K	H - K - 9		0 1	
5.	1.	3 - E - 2	2 - 3 - E	E - 2 - 3		0 1	0 1 2 3
		If the child responds 3-2-E or E-3-2, say, **Remember to say the numbers in order.**					
	2.	9 - J - 4	4 - 9 - J	J - 4 - 9		0 1	
	3.	B - 5 - F	5 - B - F	B - F - 5		0 1	
6.	1.	1 - C - 3 - J	1 - 3 - C - J	C - J - 1 - 3		0 1	0 1 2 3
	2.	5 - A - 2 - B	2 - 5 - A - B	A - B - 2 - 5		0 1	
	3.	D - 8 - M - 1	1 - 8 - D - M	D - M - 1 - 8		0 1	
7.	1.	1 - B - 3 - G - 7	1 - 3 - 7 - B - G	B - G - 1 - 3 - 7		0 1	0 1 2 3
	2.	9 - V - 1 - T - 7	1 - 7 - 9 - T - V	T - V - 1 - 7 - 9		0 1	
	3.	P - 3 - J - 1 - M	1 - 3 - J - M - P	J - M - P - 1 - 3		0 1	
8.	1.	1 - D - 4 - E - 9 - G	1 - 4 - 9 - D - E - G	D - E - G - 1 - 4 - 9		0 1	0 1 2 3
	2.	H - 3 - B - 4 - F - 8	3 - 4 - 8 - B - F - H	B - F - H - 3 - 4 - 8		0 1	
	3.	7 - Q - 6 - M - 3 - Z	3 - 6 - 7 - M - Q - Z	M - Q - Z - 3 - 6 - 7		0 1	
9.	1.	S - 3 - K - 4 - Y - 1 - G	1 - 3 - 4 - G - K - S - Y	G - K - S - Y - 1 - 3 - 4		0 1	0 1 2 3
	2.	7 - S - 9 - K - 1 - T - 6	1 - 6 - 7 - 9 - K - S - T	K - S - T - 1 - 6 - 7 - 9		0 1	
	3.	L - 2 - J - 6 - Q - 3 - G	2 - 3 - 6 - G - J - L - Q	G - J - L - Q - 2 - 3 - 6		0 1	
10.	1.	4 - B - 8 - R - 1 - M - 7 - H	1 - 4 - 7 - 8 - B - H - M - R	B - H - M - R - 1 - 4 - 7 - 8		0 1	0 1 2 3
	2.	J - 2 - U - 8 - A - 5 - C - 4	2 - 4 - 5 - 8 - A - C - J - U	A - C - J - U - 2 - 4 - 5 - 8		0 1	
	3.	6 - L - 1 - Z - 5 - H - 2 - W	1 - 2 - 5 - 6 - H - L - W - Z	H - L - W - Z - 1 - 2 - 5 - 6		0 1	

Stop Time ___:___
Hr. Min.

Total Raw Score (Maximum = 30) [ ]

8. Matrix Reasoning

Start Time ___:___ Hr. Min.

Start Ages 6–8: Samples A–C, then Item 4; Ages 9–11: Samples A–C, then Item 7; Ages 12–16: Samples A–C, then Item 11

Reverse Ages 6–16: Score of 0 on *either* of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue After 4 consecutive scores of 0 or 4 scores of 0 on *five* consecutive items

Score Score 0 or 1 point. Correct responses are in colour.

Item	Response	Score
6–16 A.	1 2 3 4 5 DK	
B.	1 2 3 4 5 DK	
C.	1 2 3 4 5 DK	
1.	1 2 3 4 5 DK	0 1
2.	1 2 3 4 5 DK	0 1
3.	1 2 3 4 5 DK	0 1
6–8 4.	1 2 3 4 5 DK	0 1
5.	1 2 3 4 5 DK	0 1
6.	1 2 3 4 5 DK	0 1
9–11 7.	1 2 3 4 5 DK	0 1
8.	1 2 3 4 5 DK	0 1
9.	1 2 3 4 5 DK	0 1
10.	1 2 3 4 5 DK	0 1
12–16 11.	1 2 3 4 5 DK	0 1

Item	Response	Score
12.	1 2 3 4 5 DK	0 1
13.	1 2 3 4 5 DK	0 1
14.	1 2 3 4 5 DK	0 1
15.	1 2 3 4 5 DK	0 1
16.	1 2 3 4 5 DK	0 1
17.	1 2 3 4 5 DK	0 1
18.	1 2 3 4 5 DK	0 1
19.	1 2 3 4 5 DK	0 1
20.	1 2 3 4 5 DK	0 1
21.	1 2 3 4 5 DK	0 1
22.	1 2 3 4 5 DK	0 1
23.	1 2 3 4 5 DK	0 1
24.	1 2 3 4 5 DK	0 1
25.	1 2 3 4 5 DK	0 1

Item	Response	Score
26.	1 2 3 4 5 DK	0 1
27.	1 2 3 4 5 DK	0 1
28.	1 2 3 4 5 DK	0 1
29.	1 2 3 4 5 DK	0 1
30.	1 2 3 4 5 DK	0 1
31.	1 2 3 4 5 DK	0 1
32.	1 2 3 4 5 DK	0 1
33.	1 2 3 4 5 DK	0 1
34.	1 2 3 4 5 DK	0 1
35.	1 2 3 4 5 DK	0 1

Total Raw Score (Maximum = 35)

Stop Time ___:___ Hr. Min.

9. Comprehension

Start Time ___:___ Hr. Min.

Start Ages 6–8: Item 1; Ages 9–11: Item 3; Ages 12–16: Item 5

Reverse Ages 9–16: Score of 0 or 1 on first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue After 4 consecutive scores of 0

Score Score 0, 1, or 2 points. See the *Administration and Scoring Manual* for sample responses.

Item	Response	Score
6–8 †1. Teeth		0 1 2
2. Vegetables		0 1 2
9–11 3. Wallet		0 1 2
4. Seatbelts		0 1 2
12–16 5. Police		0 1 2
*6. Smoke		0 1 2

† If the child does not give a 2-point response, provide the response indicated in the *Administration and Scoring Manual*.

* If the child replies with only one general idea, ask for a second response as indicated in the *Administration and Scoring Manual*.

Continue

9. Comprehension *(Continued)*

Discontinue after 4 consecutive scores of 0

Item	Response	Score
7. Fight		0 1 2
8. Exercise		0 1 2
9. Apologise		0 1 2
*10. Lights		0 1 2
11. Inspect		0 1 2
*12. Libraries		0 1 2
*13. Doctors		0 1 2
*14. Newspaper		0 1 2
15. Promise		0 1 2
16. Copyrights		0 1 2
17. Stamps		0 1 2
*18. Owning		0 1 2
*19. Democracy		0 1 2
*20. Technology		0 1 2
*21. Communication		0 1 2

* If the child replies with one general idea, ask for a second response as indicated in the *Administration and Scoring Manual.*

Total Raw Score (Maximum = 42) ______

Stop Time___:___
Hr. Min.

10. Symbol Search

(Time Limit: 120 seconds)

Start Time___:___
Hr. Min.

Start
Ages 6–7: Symbol Search A Sample Items, Practice Items, then Test Items
Ages 8–16: Symbol Search B Sample Items, Practice Items, then Test Items

Discontinue
After 120 seconds

Score
Use the Scoring Key to check the child's responses.
Subtract Number Incorrect from Number Correct.
If the Total Raw Score is <0, enter 0.

Completion Time

Stop Time___:___
Hr. Min.

Number Correct – Number Incorrect = Total Raw Score
(Ages 6–7: Max = 45)
(Ages 8–16: Max = 60)

11. Picture Completion

(Time Limit: 20 seconds)

Start Time___:___
Hr. Min.

Start
Ages 6–8: Sample, then Item 1
Ages 9–11: Sample, then Item 5
Ages 12–16: Sample, then Item 10

Reverse
Ages 9–16: Score of 0 on & the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 6 consecutive scores of 0

Score
Score 0 or 1 point
See the *Administration and Scoring Manual* for sample responses.

	Item	Response	Score	Item	Response	Score	Item	Response	Score
6–16	S. Pencil			14. Face		0 1	28. Thermometer		0 1
6–8	†1. Jacket		0 1	15. Belt		0 1	29. Trellis		0 1
	†2. Fox		0 1	16. Bulb		0 1	30. Supermarket		0 1
	3. Cat		0 1	17. Dice		0 1	31. Bicycle		0 1
	4. Bell		0 1	18. Football		0 1	32. Goldfish		0 1
9–11	5. Hand		0 1	19. Bike		0 1	33. Umbrella		0 1
	6. Mirror		0 1	20. Tree		0 1	34. Profile		0 1
	7. Skipping		0 1	21. Scissors		0 1	35. House		0 1
	8. Leaf		0 1	22. Band		0 1	36. Water		0 1
	9. Door		0 1	23. Bridge		0 1	37. Family		0 1
12–16	10. Man		0 1	24. Bath		0 1	38. Shoe		0 1
	11. Ladder		0 1	25. Pig		0 1			
	12. Chest of Drawers		0 1	26. Whistle		0 1			
	13. Clock		0 1	*27. Orange		0 1			

Total Raw Score (Maximum = 38)

Stop Time___:___
Hr. Min.

† If the child does not give a 1-point response, provide the response indicated in the *Administration and Scoring Manual*.
* Responses requiring specific query are identified in the *Administration and Scoring Manual*.

12. Cancellation

(Time Limit: 45 seconds)

Start Time___:___
Hr. Min.

Start
Ages 6–16: Sample, Practice, then Item 1

Discontinue
After 45 seconds for each item

Score
Use the Scoring Template to check the child's responses.
Subtract Number Incorrect from Number Correct. If the Total Raw Score is <0, enter 0.
CAR and CAS: Total Raw Score for Items 1 and 2, respectively.

	Item	Time Limit (Seconds)	Completion Time	Number Correct	Number Incorrect	Difference	Bonus Points	Total Raw Score
6–16	1. Random	45					Max = 4	CAR Max = 68
	2. Structured	45					Max = 4	CAS Max = 68

Time Bonus Points
If the child completes an item before 45 seconds and the difference is ≥ 60, award bonus points.

Time in Seconds	45	40–44	35–39	30–34	0–29
Bonus Points	0	1	2	3	4

Total Raw Score (Maximum = 136)

Stop Time___:___
Hr. Min.

13. Information

Start Time___:___ Hr. Min.

Start
Ages 6–8: Item 5
Ages 9–11: Item 10
Ages 12–16: Item 12

Reverse
Ages 6–16: Score of 0 on *either* of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 5 consecutive scores of 0.

Score
Score 0 or 1 point
See *Administration and Scoring Manual* for sample responses.

Item	Response	Score
†1. Foot		0 1
†2. Nose		0 1
3. Eat		0 1
*4. Ears		0 1
6–8 ▶ 5. Old		0 1
6. Legs		0 1
*7. Thursday		0 1
*8. Coins		0 1
9. March		0 1
9–11 ▶ *10. Week		0 1
11. Boil		0 1
12–16 ▶ *12. Seasons		0 1
*13. Stomach		0 1
14. Dozen		0 1
15. Month		0 1
16. Fossil		0 1
*17. Year		0 1

Item	Response	Score
18. Hieroglyphics		0 1
19. Oxygen		0 1
20. Leaves		0 1
21. Kalidasa		0 1
22. Ozone		0 1
23. Malaysia		0 1
*24. Population		0 1
*25. Rust		0 1
26. Barometer		0 1
*27. Darwin		0 1
28. Solstice		0 1
*29. Mumbai		0 1
30. Diamonds		0 1
31. Fission		0 1
32. Confucius		0 1
*33. Turpentine		0 1

Stop Time___:___ Hr. Min.

Total Raw Score (Maximum = 33)

† If the child does not give a 1-point response, provide the response indicated in the *Administration and Scoring Manual.*

* Responses requiring specific query are identified in the *Administration and Scoring Manua*

14. Arithmetic (Time Limit: 30 seconds)

Start Time___:___ Hr. Min.

Start
Ages 8–9: Item 9
Ages 10–16: Item 12

Reverse
two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 4 consecutive scores of 0

Score
Score 0 or 1 point

Item	Correct Response	Response	Score
†1. Birds	1, 2, 3		0 1
†2. Chicks	1, 2, 3, 4, 5		0 1
6–7 ▶ †3. Trees	1, 2,...10		0 1
4. Butterflies	9		0 1
5. Nuts	2		0 1
6. Crayons	5		0 1
7. Books	4		0 1
8. Biscuits	3		0 1
8–9 ▶ 9. Pieces	2		0 1
10. Rupees	6		0 1
11. Pencils	6		0 1
10–16 ▶ 12. Chocolate	7		0 1

Item	Correct Response	Response	Score
13. Cars	15		0 1
14. Pens	14		0 1
15. Marbles	25		0 1
16. Cows	5		0 1
17. Apples	9		0 1
18. Balloons	7		0 1
19. Watching	6		0 1
20. Points	32		0 1
21. Toffees	20 paise		0 1
22. Magazines	3		0 1
23. Prizes	24		0 1
24. Students	19		0 1

Item	Correct Response	Response	Score
25. Change	7		0 1
26. Money	Rs.8.50		0 1
27. Classrooms	20		0 1
28. Book	30		0 1
29. Drive	60		0 1
30. Temperature	3		0 1
31. Toy	34		0 1
32. Car wash	48		0 1
33. Flight	2:00		0 1
34. Work	40		0 1

Total Raw Score (Maximum = 34)

† If the child does not give a 1-point response, provide the response indicated in the *Administration and Scoring Manual.*

Stop Time___:___ Hr. Min.

15. Word Reasoning

Start Time ___:___
Hr. Min.

Start
Ages 6–9: Samples A & B, then Item 1
Ages 10–16: Samples A & B, then Item 5

Reverse
Ages 10–16: Score of 0 on a first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 5 consecutive scores of 0

Score
Score 0 or 1 point
See *Administration and Scoring Manual* for sample responses.

	Item	Clue	Response	Correct	Score
6–16	A	I. This is an animal that goes "bow-wow."		Y N	
	B	I. This has a long handle...		Y N	
		II. and is used to clean the floor.		Y N	
6–9	1.	I. This is used to dry yourself after a bath.		Y N	0 1
	2.	I. This is a room where people sleep.		Y N	0 1
	3.	I. This is an animal with a long trunk and big ears.		Y N	0 1
	4.	I. This is the part of your head used to smell things.		Y N	0 1
10–16	5.	I. This has a handle and people can open it and walk through it.		Y N	0 1
	6.	I. This is something you wear to cover your head.		Y N	0 1
	7.	I. These are up in the sky...		Y N	0 1
		II. and you can only see these at night.		Y N	
	8.	I. This comes in many colours...		Y N	0 1
		II. and it is used on walls.		Y N	
	9.	I. This is a place of learning...		Y N	0 1
		II. and it can have many things from the past in it.		Y N	
	10.	I. This is on the ground after it rains...		Y N	0 1
		II. and it can make your clothes dirty.		Y N	
	11.	I. This is what people do to make things "like new"...		Y N	0 1
		II. and people do this to things that are broken.		Y N	
	12.	I. This can make food taste better...		Y N	0 1
		II. and it can be found in the ocean.		Y N	
	13.	I. These are protected by skin and bones...		Y N	0 1
		II. and they can sometimes be replaced.		Y N	
	14.	I. This can be broken...		Y N	0 1
		II. and it is written down to protect people.		Y N	
	15.	I. This is found in nature...		Y N	0 1
		II. and it has water flowing from its mouth.		Y N	

Continue

15. Word Reasoning *(Continued)*

Discontinue after 5 consecutive scores of 0

Item	Clue	Response	Correct	Score
16.	I. This leads to new discoveries...		Y N	0 1
	II. and it involves a step-by-step process...		Y N	
	III. and it may include experiments.		Y N	
17.	I. This allows people to do or own something...		Y N	0 1
	II. and it is given by the government...		Y N	
	III. and people may have to take a test to get it.		Y N	
18.	I. This is something everything has...		Y N	0 1
	II. and it changes every year...		Y N	
	III. and it can never decrease.		Y N	
19.	I. This can't be touched...		Y N	0 1
	II. and it can't go backwards...		Y N	
	III. and it can't stop.		Y N	
20.	I. This is something most people want...		Y N	0 1
	II. and conflict prevents it...		Y N	
	III. and governments try to keep it.		Y N	
21.	I. This can be a river...		Y N	0 1
	II. and wars can change it...		Y N	
	III. and two countries can share it.		Y N	
22.	I. This is made by people...		Y N	0 1
	II. and it can *not* be touched...		Y N	
	III. and it has already happened.		Y N	
23.	I. This is a place...		Y N	0 1
	II. and it is protected from the weather...		Y N	
	III. and it is found within something else.		Y N	
24.	I. This has never been seen or done before...		Y N	0 1
	II. and it can make our lives better or easier...		Y N	
	III. and it is a product of the mind.		Y N	

Stop Time___:___
Hr. Min.

Total Raw Score (Maximum = 24)

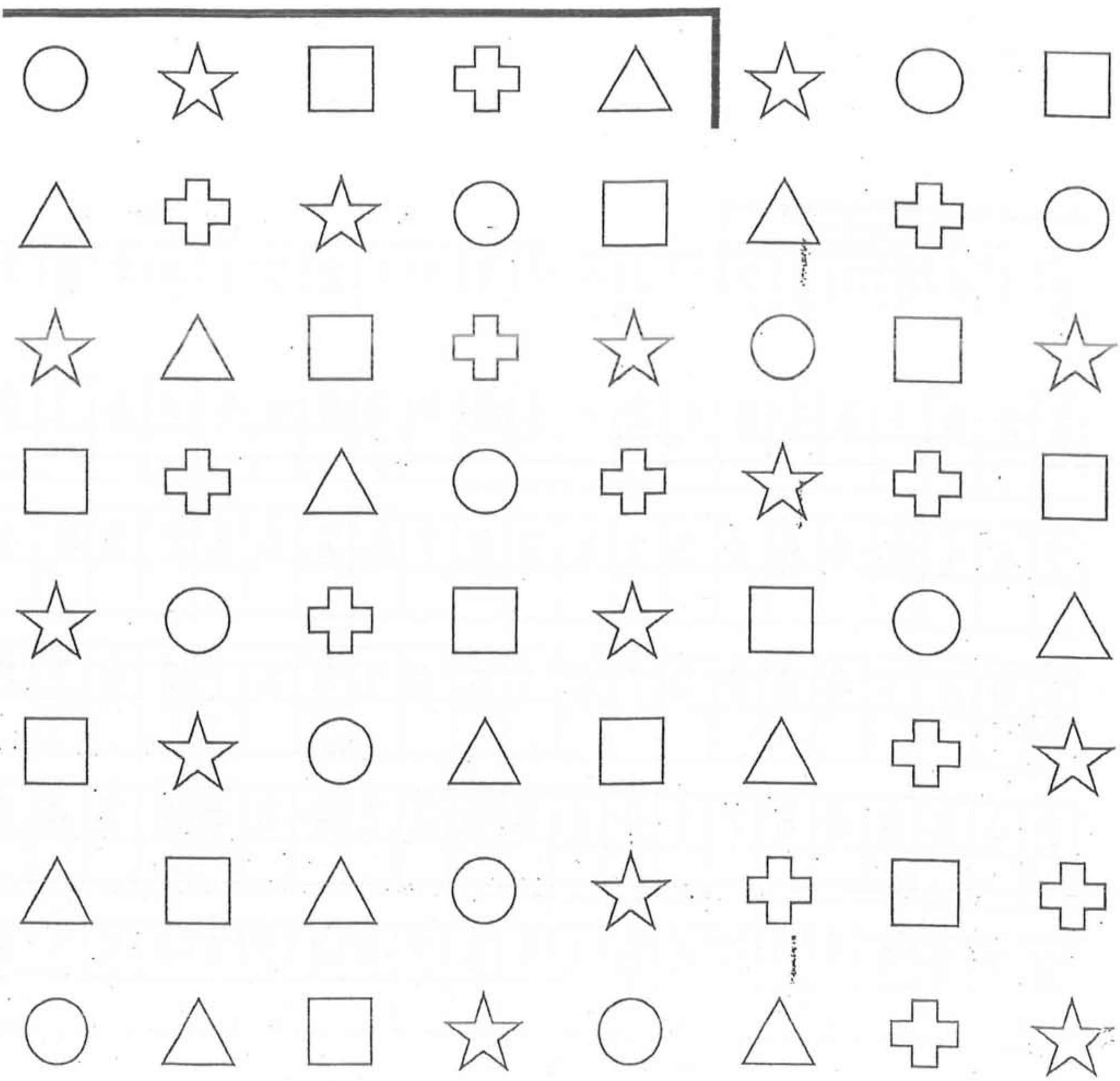
Notes / Behavioral Observations :

Appendix B2.

Coding A

Ages 6–7

SAMPLE ITEMS



Symbol Search A

Ages 8–16, turn to page

Ages 6–7

SAMPLE ITEMS

<	⊕	L	<	جی ہاں	نہیں
ǂ	∪	↝	⊗	جی ہاں	نہیں

PRACTICE ITEMS

⊞	$\bar{\cap}$	⊞	┘	جی ہاں	نہیں
⋚	L	∼	∩	جی ہاں	نہیں

Proceed to page

A *(Continued)*

⇋	⇋	↝	⥲	جی ہاں	نہیں
⩳	≉	⥲	⊄	جی ہاں	نہیں
⊪	⊪	♉	♈	جی ہاں	نہیں
♈	♉	♈	⟦	جی ہاں	نہیں
⤙	♈	≉	⤙	جی ہاں	نہیں
⭁	⇌	⥬	≉	جی ہاں	نہیں
⌝	∟	⊣	⊩	جی ہاں	نہیں
⟈	⊚	⩠	≉	جی ہاں	نہیں
⫨	≉	⊪	⫨	جی ہاں	نہیں
‡	⊧	∓	⊩	جی ہاں	نہیں
∟	⊢	⌝	⊪	جی ہاں	نہیں
⥲	⥲	≈	⇌	جی ہاں	نہیں
⤙	⋚	⊅	⤙	جی ہاں	نہیں
∼	⊙	∼	⩡	جی ہاں	نہیں
⊩	⟦	⊪	≉	جی ہاں	نہیں

C I

Symbol Search B

Ages 8–16

SAMPLE ITEMS

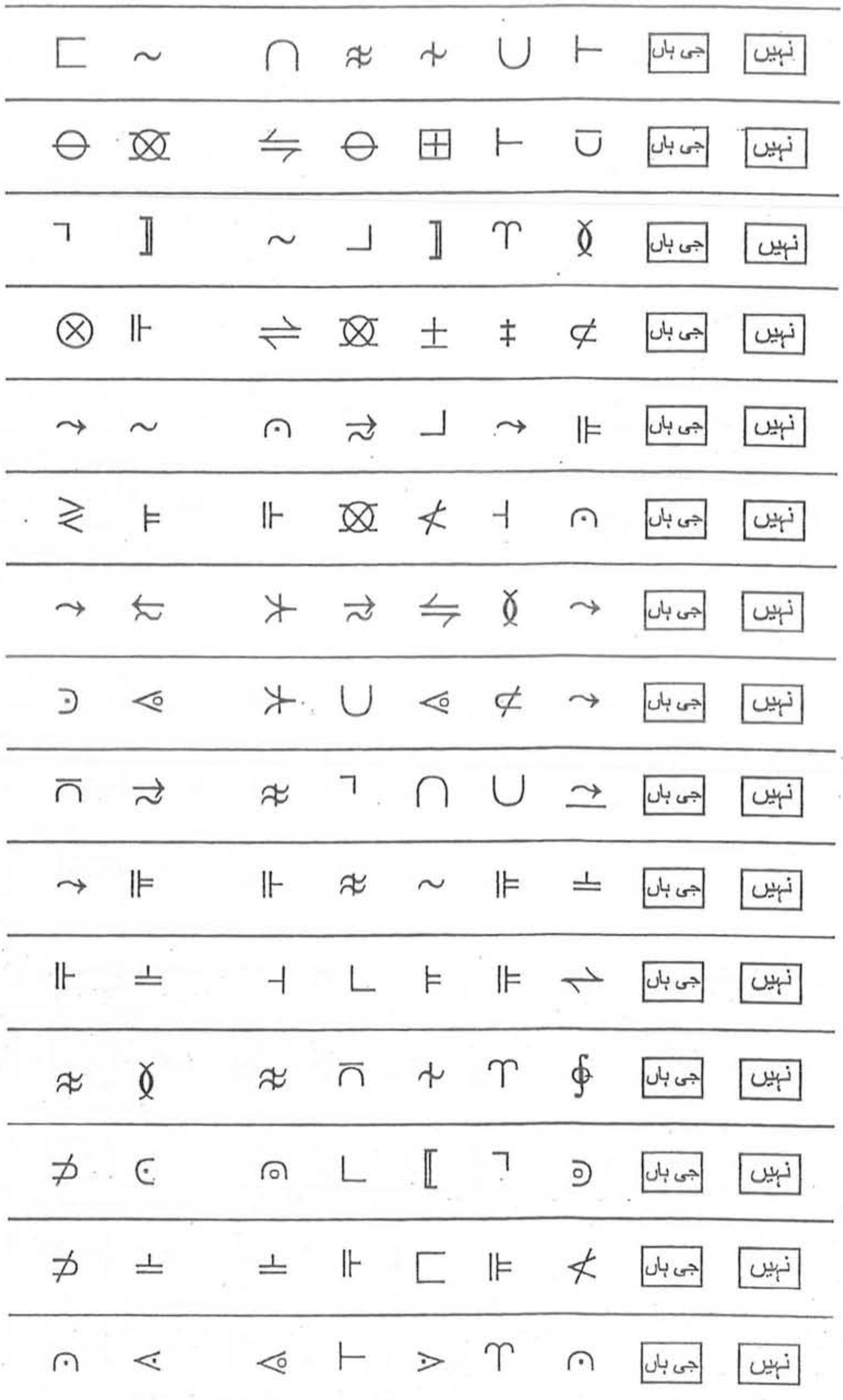
⊕	⊖	⊕	∟	<	⊢	~	جی ہاں	نہیں
↝	∟	ǂ	∩	ϒ	⋚	⊞	جی ہاں	نہیں

PRACTICE ITEMS

⊫	<	↝	⊫	±	∢	⊖	جی ہاں	نہیں
≈	⊖	∩̄	±	┘	ǂ	ϒ	جی ہاں	نہیں

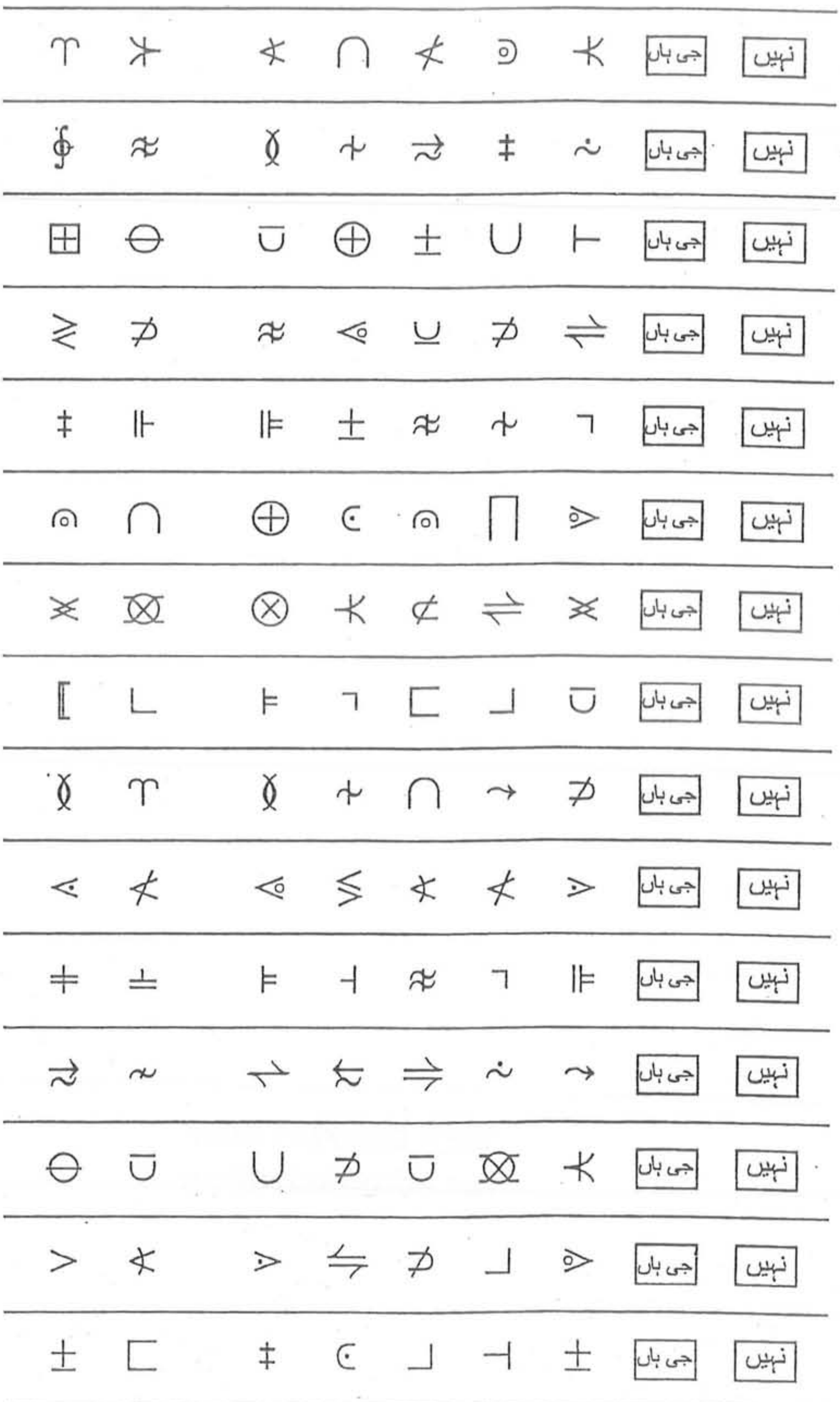
Proceed to page

B *(Continued)*



C I

B *(Continued)*



C ____ I ____

Appendix B3

WISC-IV UK Cancellation

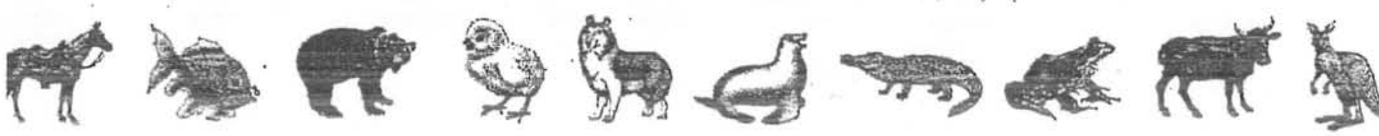
Response Booklet 2

Child's Name: ______________________________

Examiner's Name: ______________________________

Today's Date: ____________________ Age: ____________

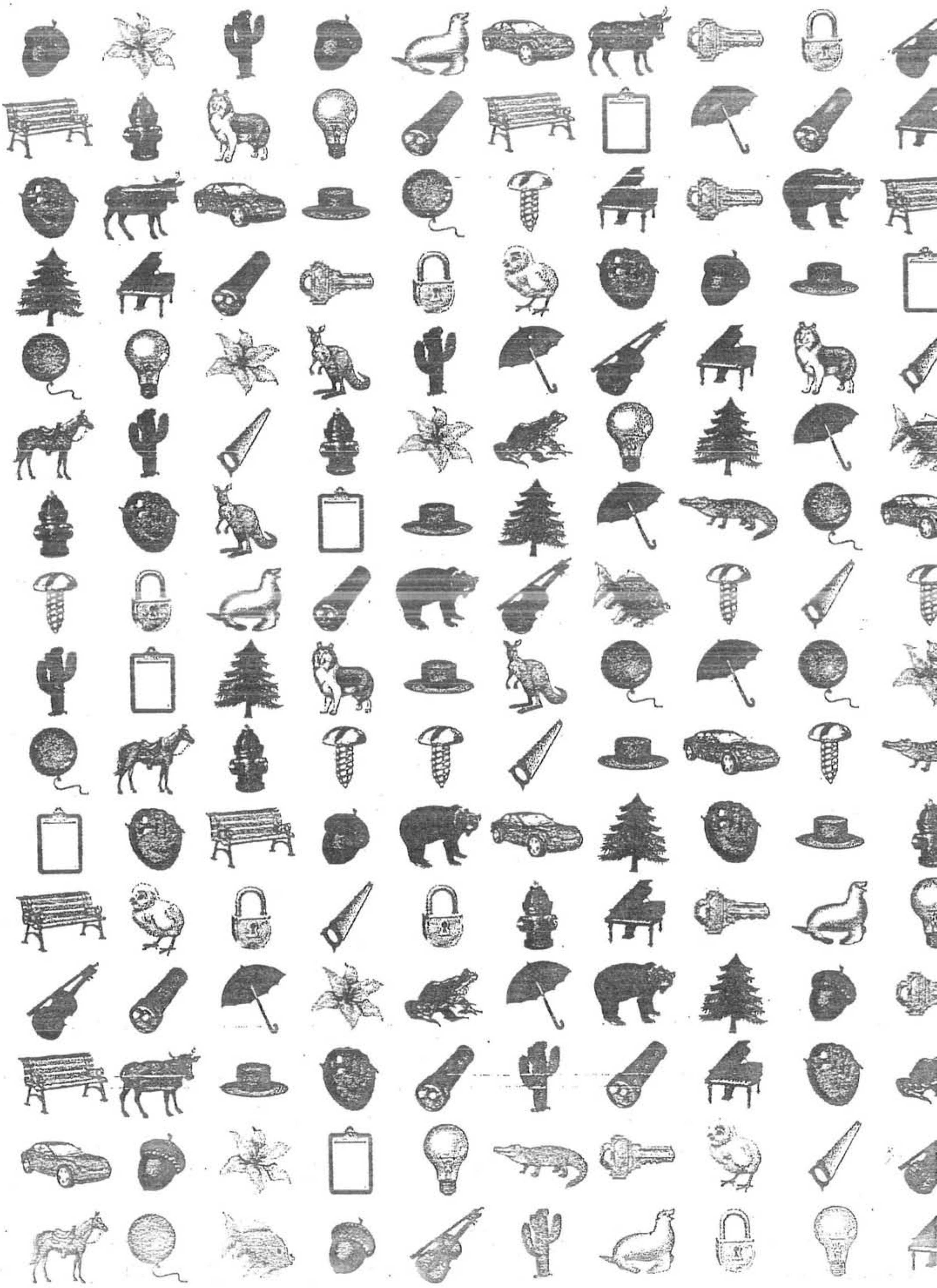
Animals



Sample:

Practice:

2.



Appendix C1

INITIAL LIST OF WORDS TO BE REPLACED FOR VOCABULARY SUBTEST

	English Word List for Vocabulary Subtest	Urdu Replacement Word List for Vocabulary Subtest
1.	Hat	ٹوپی
2.	Umbrella	چھتری
3.	Clock	گھڑیال
4.	Cow	گائے
5.	Thief	چور
6.	Bicycle	سائیکل
7.	Alphabet	حروف تہجی
8.	Leave	رخصت
9.	Ancient	قدیم
10.	Bark	چھال
11.	Brave	بہادر
12.	Obey	فرمابرداری کرنا
13.	Island	جزیرہ
14	Nonsense	احمقانہ
15.	Absorb	جذب کرنا
16.	Transparent	شفاف
17.	Precise	جامع
18	Mimic	نقالی کرنا
19.	Fable	حکایت
20	Migrate	ہجرت کرنا

21.	Rivalry	رقابت
22.	Foresight	دوراندیشی
23.	Seldom	شاذونادر
24.	Strenuous	دقت طلب
25.	Unanimous	متفقہ
26.	Imminent	ناگزیر
27.	Amendment	ترمیم
28.	Compel	زبردستی کرنا
29.	Affliction	ایذا
30.	Garrulous	بڑبولا
31.	Dilatory	دیرباز
32.	Aberration	انحراف / خلل

Appendix C2

FINAL LIST OF WORDS TO BE REPLACED FOR VOCABULARY SUBTEST

	English Word List for Vocabulary Subtest	Urdu Replacement Word List for Vocabulary Subtest
1.	Hat	ٹوپی
2.	Umbrella	چھتری
3.	Clock	گھڑی
4.	Cow	گائے
5.	Thief	چور
6.	Bicycle	سائیکل
7.	Alphabet	حروف تہجی
8.	Leave	رخصت
9.	Ancient	قدیم
10.	Bark	چھال
11.	Brave	بہادر
12.	Obey	فرمابرداری کرنا
13.	Island	جزیرہ
14	Nonsense	احمقانہ
15.	Absorb	جذب کرنا
16.	Transparent	شفاف
17.	Precise	جامع
18	Mimic	نقالی کرنا
19.	Fable	حکایت
20	Migrate	ہجرت کرنا

21.	Rivalry	رقابت
22.	Foresight	دوراندیشی
23.	Seldom	شاذ و نادر
24.	Strenuous	دقت طلب
25.	Unanimous	متفقہ
26.	Imminent	ناگزیر
27.	Amendment	ترمیم
28.	Compel	جبر کرنا
29.	Affliction	ایذا
30.	Garrulous	بڑبولا
31.	Dilatory	دیرباز
32.	Aberration	انحراف

Appendix D1

List of Items Suggested for Adaptation Along With the Replaced Items

Similarities Subtest

ماتھے پہ بل پڑنے اور مسکرانے میں کیا مشابہت ہے؟
ماتھے پہ شکن پڑنے اور مسکرانے میں کیا مشابہت ہے؟

Vocabulary Subtest

گھڑیال کسے کہتے ہیں؟
گھڑی کسے کہتے ہیں؟
حروف کسے کہتے ہیں؟
حروفِ تہجی کسے کہتے ہیں؟
خلل سے کیا مُراد ہے؟
انحراف سے کیا مراد ہے؟

Comprehension

گورنمنٹ کے لئے یہ کیوں ضروری ہے کہ وہ بچوں کو اسکول میں دوپہر کا کھانا دینے سے پہلے اُس کا معائنہ کرے؟
گورنمنٹ کے لئے یہ کیوں اہم ہے کہ وہ گوشت کے بکنے سے پہلے اس کا معائنہ کرے؟

Information Subtest

کالی داس کون تھا؟

کرسٹوفر کولمبس کون تھا؟

ہائڈروگرافیکس سے کیا مراد ہے؟

خطاطی (Calligraphy) کیا ہے؟

Winter Solicitice سے کیا مراد ہے؟

سال کا سب سے چھوٹا دن کونسا ہوتا ہے؟

کنفیوشس کون تھا؟

سقراط (Socrates) کون تھا؟

ممبئی سے دبئی کا کتنا فاصلہ ہے؟

کراچی سے دبئی کا کتنا فاصلہ ہے؟

Arithmetic Subtest

اگر آپ 40 پیسہ فی ٹافی کے حساب سے 2 ٹافیاں خریدتے ہیں تو ایک روپے میں سے آپ کو کتنا بقایا واپس ملے گا

اگر آپ 20 سیکنڈ فی سوال کے حساب سے 2 سوالات حل کرتے ہیں تو آپ کے پاس 1 منٹ میں سے کتنا وقت باقی بچے گا؟

Note: The were many changes in sampled responses of these subtests. in Arithmetic subtest name of persons in many items have also been changed.

Changes in Content of WISC-IV Subtests Items or Sampled Responses based on Pre-testing/Expert Opinion

Subtest	Item number	Change in item Content	Change in Sampled Response	Basis for Change
Comprehension	11	Item replaced with the item no. 11 (WISC-IV UK, 2004)		Pre-testing & expert opinion
Information	8		1 point Response Added:	Pre-testing & expert opinion
	18	Replaced with a new item	New 1 point & 0 point responses	Pre-testing & expert opinion
	21	Replaced with the item no. 21 (WISC-IV UK, 2004)	New 1 point & 0 point responses	Pre-testing & expert opinion
	28	Replaced with a new item	New 1 point & 0 point responses	Pre-testing & expert opinion
	29	Replaced with a new item	New 1 point responses	Pre-testing & expert opinion
	32	Replaced with a new item	New 1 point & 0 point responses	Pre-testing & expert opinion
Vocabulary	All items		All new Urdu sampled responses	Response option compilation (step VIII)
Picture Completion	14	Minor change in the picture (Female face)		
Arithmetic	7, 8, 10, 14,15, 17, 18, 19, 20, 22, 25, 26, 28, 33, 34	Change in names used in the items.		Pre-testing & expert opinion

Appendix E1

WISC-IV Record Form (Urdu)

Urdu Standardization Edition

Examiner's Name:________________ Site ID Number:______ Region:___________

Examinee's Name:____________________ Examinee's ID Number:__________

Teacher's Signature____________________

	Year	Month	Day
Today's Date			
Date of Birth			
Age at testing			

Gender	Female	Male	
Handedness	Left	Right	Both
Colorblind?	No	Yes	

Does the examinee have any disabling conditions? No Yes*

If yes, please explain how the examinee meets inclusion criteria in the box below No Yes

Does the examinee wear prescription lenses, glasses, or a hearing aid? No Yes*

*If yes, was he/she wearing them during testing? No Yes

Start Time ___:___ Hr. Min.

2. Similarities

Start
Ages 6–8: Sample, then Item 1
Ages 9–11: Sample, then Item 3
Ages 12–16: Sample, then Item 5

Reverse
Ages 9–16: Score of 0 or 1 on either of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 5 consecutive scores of 0

Score
Items 1–2: Score 0 or 1 point.
Items 3–23: Score 0, 1, or 2 points.
See *Administration and Scoring Manual* for sample responses.

Item	Response	Score
6–16 سُرخ-نیلا		
6–8 †1. دودھ-پانی		0 1
†2. قلم-پنسل		0 1
9–11 3. سیب-کیلا		0 1 2
4. قمیص-جوتا		0 1 2
12–16 5. بلی-چوہا		0 1 2
6. تتلی-شہد کی مکھی		0 1 2
7. سردیوں-گرمیوں		0 1 2
8. غصہ-خوشی		0 1 2
9. کہنی-گھٹنا		0 1 2
10. شہتیر-اینٹیں		0 1 2
11. پینٹنگ-مجسمہ		0 1 2
12. ماتھے پر شکن پڑنا-مسکرانا		0 1 2
13. برف-بھاپ		0 1 2
14. شاعر-مصور		0 1 2
15. سیلاب-خشک سالی		0 1 2
16. پہاڑ-جھیل		0 1 2
17. اوّل-آخر		0 1 2
18. انتقام-معافی		0 1 2

† If the child does not give a 1-point response, provide the response indicated in the *Administration and Scoring Manual.*

Continue

4. Picture Concepts

Start Time___:___
Hr. Min.

Start
Ages 6–8: Samples A & B, then Item 1
Ages 9–11: Samples A & B, then Item 5
Ages 12–16: Samples A & B, then Item 7

Reverse
Ages 9–16: Score of 0 on either of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 5 consecutive scores of 0.

Score
Score 0 or 1 point.
Correct responses are in colour

Start	Item	Response	Score
6–16		1 2 3 4 DK	
		1 2 3 4 DK	
6–8	1.	1 2 3 4 DK	0 1
	2.	1 2 3 4 DK	0 1
	3.	1 2 3 4 DK	0 1
	4.	1 2 3 4 DK	0 1
9–11	5.	1 2 3 4 DK	0 1
	6.	1 2 3 4 5 6 DK	0 1
12–16	7.	1 2 3 4 5 6 DK	0 1
	8.	1 2 3 4 5 6 DK	0 1
	9.	1 2 3 4 5 6 DK	0 1
	10.	1 2 3 4 5 6 DK	0 1
	11.	1 2 3 4 5 6 DK	0 1

Item	Response	Score
12.	1 2 3 4 5 6 7 8 9 DK	0 1
13.	1 2 3 4 5 6 7 8 9 DK	0 1
14.	1 2 3 4 5 6 7 8 9 DK	0 1
15.	1 2 3 4 5 6 DK	0 1
16.	1 2 3 4 5 6 7 8 9 DK	0 1
17.	1 2 3 4 5 6 7 8 9 DK	0 1
18.	1 2 3 4 5 6 7 8 9 DK	0 1
19.	1 2 3 4 5 6 7 8 9 10 11 12 DK	0 1
20.	1 2 3 4 5 6 7 8 9 10 11 12 DK	0 1
21.	1 2 3 4 5 6 7 8 9 10 11 12 DK	0 1
22.	1 2 3 4 5 6 7 8 9 DK	0 1
23.	1 2 3 4 5 6 7 8 9 10 11 12 DK	0 1
24.	1 2 3 4 5 6 7 8 9 10 11 12 DK	0 1
25.	1 2 3 4 5 6 7 8 9 10 11 12 DK	0 1
26.	1 2 3 4 5 6 7 8 9 10 11 12 DK	0 1
27.	1 2 3 4 5 6 7 8 9 10 11 12 DK	0 1
28.	1 2 3 4 5 6 7 8 9 DK	0 1

Stop Time___:___
Hr. Min.

Total Raw Score
(Maximum = 28)

5. Coding

(Time Limit: 120 seconds)

Start Time___:___
Hr. Min.

Start
Ages 6–7: Coding A Sample Items, then Test Items
Ages 8–16: Coding B Sample Items, then Test Items

Discontinue
After 120 seconds

Score
Use the Scoring Key to check the child's responses.
Score 1 point for each correct response.

Start	Form	Time Limit (Seconds)	Completion Time	Total Raw Score
6–7	A.	120		Max = 65
8–16	B.	120		Max = 119

Coding A Time Bonus Scores for Perfect Performance

Time in Seconds	116–120	111–115	106–110	101–105	96–100	86–95	≤8[illegible]
Score	59	60	61	62	63	64	65

Stop Time___:___
Hr. Min.

6. Vocabulary *(Continued)*

Discontinue after 5 consecutive scores of 0

Item	Response	Score
18. اعتقاد		0 1 2
19. جذب کرنا		0 1 2
20. غلاف		0 1 2
*21. جامع		0 1 2
22. نقالی		0 1 2
23. حکایت		0 1 2
24. ہجرت		0 1 2
25. رقابت		0 1 2
26. دُور اندیش		0 1 2
27. شاذ و نادر		0 1 2
28. راحت طلب		0 1 2
29. مُصِر		0 1 2
*30. ناگزیر		0 1 2
31. ترمیم		0 1 2
32. جبر کرنا		0 1 2
*33. ایذا		0 1 2
34. بڑبولا		0 1 2
35. دریاز		0 1 2
36. انحراف		0 1 2

* Responses requiring specific query are identified in the *Administration and Scoring Manual*.

Total Raw Score (Maximum = 68)

Stop Time ___:___ Hr. Min.

8. Matrix Reasoning

Start Time___:___
Hr. Min.

Start
Ages 6–8: Samples A–C, then Item 4
Ages 9–11: Samples A–C, then Item 7
Ages 12–16: Samples A–C, then Item 10

Reverse
Ages 6–16: Score of 0 on either of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained

Discontinue
After 4 consecutive scores of 0 or 4 scores of 0 on five consecutive items

Score
Score 0 or 1 point.
Correct responses are in colour.

	Item	Response						Score
6–16		1	2	3	4	5	DK	
		1	2	3	4	5	DK	
		1	2	3	4	5	DK	
	1.	1	2	3	4	5	DK	0 1
	2.	1	2	3	4	5	DK	0 1
	3.	1	2	3	4	5	DK	0 1
6–8	4.	1	2	3	4	5	DK	0 1
	5.	1	2	3	4	5	DK	0 1
	6.	1	2	3	4	5	DK	0 1
9–11	7.	1	2	3	4	5	DK	0 1
	8.	1	2	3	4	5	DK	0 1
	9.	1	2	3	4	5	DK	0 1
	10.	1	2	3	4	5	DK	0 1
12–16	11.	1	2	3	4	5	DK	0 1

Item	Response						Score
12.	1	2	3	4	5	DK	0 1
13.	1	2	3	4	5	DK	0 1
14.	1	2	3	4	5	DK	0 1
15.	1	2	3	4	5	DK	0 1
16.	1	2	3	4	5	DK	0 1
17.	1	2	3	4	5	DK	0 1
18.	1	2	3	4	5	DK	0 1
19.	1	2	3	4	5	DK	0 1
20.	1	2	3	4	5	DK	0 1
21.	1	2	3	4	5	DK	0 1
22.	1	2	3	4	5	DK	0 1
23.	1	2	3	4	5	DK	0 1
24.	1	2	3	4	5	DK	0 1
25.	1	2	3	4	5	DK	0 1

Item	Response						Score
26.	1	2	3	4	5	DK	0 1
27.	1	2	3	4	5	DK	0 1
28.	1	2	3	4	5	DK	0 1
29.	1	2	3	4	5	DK	0 1
30.	1	2	3	4	5	DK	0 1
31.	1	2	3	4	5	DK	0 1
32.	1	2	3	4	5	DK	0 1
33.	1	2	3	4	5	DK	0 1
34.	1	2	3	4	5	DK	0 1
35.	1	2	3	4	5	DK	0 1

Total Raw Score
(Maximum = 35)

Stop Time___:___
Hr. Min.

9. Comprehension

Start Time___:___
Hr. Min.

Start
Ages 6–8: Item 1
Ages 9–11: Item 3
Ages 12–16: Item 5

Reverse
Ages 9–16: Score of 0 or 1 on either of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 4 consecutive scores of 0

Score
Score 0, 1, or 2 points.
See the Administration and Scoring Manual for sample responses.

	Item	Response	Score
6–8	†1. دانت		0 1 2
	2. سبزیاں		0 1 2
9–11	3. بڑے		0 1 2
	4. سیٹ بیلٹ		0 1
12–16	5. پولیس		0 1
	*6. دھواں		0 1

† If the child does not give a 2-point response, provide the response indicated in the *Administration and Scoring Manual.*

* If the child replies with only one general idea, ask for a second response as indicated in the *Administration and Scoring Manual.*

Continue

10. Symbol Search

(Time Limit: 120 seconds)

Start Time___:___
Hr. Min.

Start
Ages 6–7: Symbol Search A Sample Items, Practice Items, then Test Items.
Ages 8–16: Symbol Search B Sample Items, Practice Items, then Test Items.

Discontinue
After 120 seconds.

Score
Use the Scoring Key to check the child's responses. Subtract Number Incorrect from Number Correct. If the Total Raw Score is < 0, enter 0.

Completion Time	Stop Time___:___ Hr. Min.	Number Correct	–	Number Incorrect	=	Total Raw Score (Ages 6–7: Max = 45) (Ages 8–16: Max = 60)

11. Picture Completion

(Time Limit: 20 seconds)

Start Time___:___
Hr. Min.

Start
Ages 6–8: Sample, then Item 1
Ages 9–11: Sample, then Item 5
Ages 12–16: Sample, then Item 10

Reverse
Ages 9–16: Score of 0 on either of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 6 consecutive scores of 0.

Score
Score 0 or 1 point.
See the Administration and Scoring Manual for sample responses.

Item	Response	Score
6–16 Sample. پنسل		
6–8 †1. جیکٹ		0 1
†2. لومڑی		0 1
3. بلی		0 1
4. گھنٹی		0 1
9–11 5. ہاتھ		0 1
6. آئینہ		0 1
7. رسی کودنا		0 1
8. پچ		0 1
9. دروازہ		0 1
12–16 10. آدمی		0 1
11. سیڑھی		0 1
12. الماری کا دروازہ		0 1
13. گھڑیال		0 1

Item	Response	Score
14. چمچ		0 1
15. پلیٹ		0 1
16. بلب		0 1
17. ڈائس		0 1
18. فٹبال		0 1
19. بائیک		0 1
20. درخت		0 1
21. قینچیاں		0 1
22. بینڈ		0 1
23. بگل		0 1
24. ہاتھ		0 1
25. سور		0 1
26. سیٹی		0 1
*27. بطخ		0 1

Item	Response	Score
28. تھرمامیٹر		0 1
29. ٹریلیس		0 1
30. سپر مارکیٹ		0 1
31. بائیسکل		0 1
32. سنہری مچھلی		0 1
33. چھتری		0 1
34. شکل		0 1
35. گھر		0 1
36. پانی		0 1
37. خاندان		0 1
38. بچہ		0 1

Total Raw Score (Maximum = 38)

Stop Time___:___
Hr. Min.

† If the child does not give a 1-point response, provide the response indicated in the *Administration and Scoring Manual.*
* Responses requiring specific query are identified in the *Administration and Scoring Manual.*

12. Cancellation

(Time Limit: 45 seconds)

Start Time___:___
Hr. Min.

Start
Ages 6–16: Sample, Practice, then Item 1

Discontinue
After 45 seconds for each item

Score
Use the Scoring Template to check the child's responses. Subtract Number Incorrect from Number Correct. If the Total Raw Score is < 0, enter 0.
CAR and CAS: Total Raw Score for Items 1 and 2, respectively.

Item	Time Limit (Seconds)	Completion Time	Number Correct	Number Incorrect	Difference	Bonus Points	Total Raw Score
6–16 1. Random	45						CAR Max = 68
2. Structured	45						CAS Max = 68

Time Bonus Points
If the child completes an item before 45 seconds and the difference is ≥ 60, award bonus points.

Time in Seconds	45	40–44	35–39	30–34	0–29
Bonus Points	0	1	2	3	4

Total Raw Score (Maximum = 136)

Stop Time___:___
Hr. Min.

15. Word Reasoning

Start Time ___:___ Hr. Min.

Start Ages 6–9: Samples A& B, then Item 1. Ages 10–16: Samples A& B, then Item 5.

Reverse Ages 10–16: Score of 0 on *either* of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue After 5 consecutive scores of 0.

Score Score 0 or 1 point. See *Administration and Scoring Manual* for sample responses.

Score	Correct	Clue	Response	
	Y N	I. یہ ایک جانور ہے جو "بھوں بھوں" کرتا ہے۔		
	Y N	I. اس کا ایک لمبا ہینڈل ہوتا ہے		
	Y N	II. اور اسے فرش صاف کرنے کے لئے استعمال کیا جاتا ہے		
0 1	Y N	I. اسے نہانے کے بعد خود کو خشک کرنے کے لئے استعمال کیا جاتا ہے۔		1.
0 1	Y N	I. ایسا کمرہ جہاں لوگ سوتے ہیں۔		2.
0 1	Y N	I. ایسا جانور جس کی لمبی سونڈ اور بڑے بڑے کان ہوتے ہیں۔		3.
0 1	Y N	I. یہ آپ کے سر کا حصہ ہے جو سونگھنے کے لئے استعمال کیا جاتا ہے۔		4.
0 1	Y N	I. اس کا ایک ہینڈل ہوتا ہے اور لوگ اسے کھول سکتے ہیں اور اس میں سے گزر سکتے ہیں۔		5.
0 1	Y N	I. یہ ایسی چیز ہے جو ہم سر ڈھانپنے کے لئے پہنتے ہیں۔		6.
0 1	Y N	I. یہ اوپر آسمان پر ہوتے ہیں۔		7.
	Y N	II. اور آپ انہیں صرف رات میں دیکھ سکتے ہیں۔		
0 1	Y N	I. یہ بہت سے رنگوں میں آتا ہے۔		8.
	Y N	II. اور یہ دیواروں پر استعمال ہوتا ہے۔		
0 1	Y N	I. یہ ایک سیکھنے کی جگہ ہے۔		9.
	Y N	II. اور اس میں ماضی کی بہت سی چیزیں ہو سکتی ہیں۔		
0 1	Y N	I. یہ بارش کے بعد زمین پر ہوتی ہے۔		10.
	Y N	II. اور یہ آپ کے کپڑوں کو گندا کر سکتی ہے۔		
0 1	Y N	I. یہ ایسا کام ہے جو لوگ چیزوں کو "نیا جیسا" کرنے کے لئے کرتے ہیں۔		11.
	Y N	II. اور لوگ اسے ناکارہ چیزوں کے ساتھ کرتے ہیں۔		
0 1	Y N	I. یہ کھانے کا ذائقہ بہتر بنا سکتا ہے۔		12.
	Y N	II. اور یہ سمندر سے مل سکتا ہے۔		
0 1	Y N	I. ان کی حفاظت جلد اور ہڈیوں سے کی گئی ہے۔		13.
	Y N	II. اور بعض اوقات انہیں تبدیل بھی کیا جا سکتا ہے۔		
0 1	Y N	I. اسے توڑا جا سکتا ہے۔		14.
	Y N	II. اور اسے لوگوں کی حفاظت کے لئے لکھا گیا ہے۔		
0 1	Y N	I. یہ قدرتی طور پر پایا جاتا ہے۔		15.
	Y N	II. اور اس کے منہ سے پانی بہتا ہے۔		

Continue

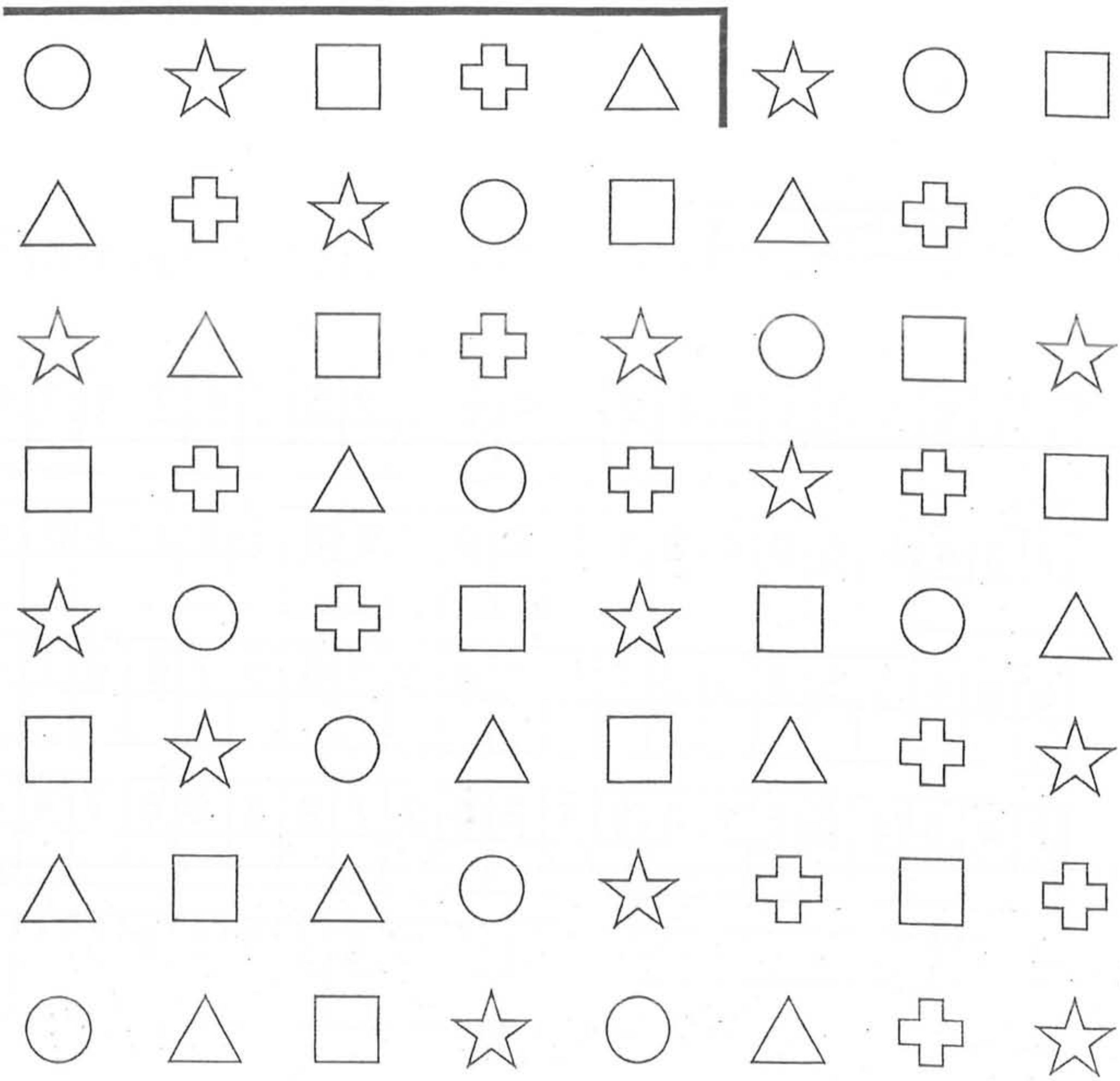
Notes / Behavioral Observations :

Coding A

Appendix E2

Ages 6–7

SAMPLE ITEMS



Symbol Search A

Ages 8–16, turn to page

Ages 6–7

SAMPLE ITEMS

<	⊕	∟	<	جی ہاں	نہیں
ǂ	∪	↝	⊗	جی ہاں	نہیں

PRACTICE ITEMS

⊞	⩃	⊞	┘	جی ہاں	نہیں
⋚	∟	~	∩	جی ہاں	نہیں

Proceed to page

A *(Continued)*

⇋	⇋	↷	⥲	جی ہاں	نہیں
⩪	≉	⥲	⊄	جی ہاں	نہیں
⊫	⊫	≬	⋔	جی ہاں	نہیں
⋔	≬	⋔	⟦	جی ہاں	نہیں
≁	⋔	≉	≁	جی ہاں	نہیں
⥳	⇌	⥷	≉	جی ہاں	نہیں
┐	└	⊣	⊪	جی ہاں	نہیں
⪽	⋐	⋗	≉	جی ہاں	نہیں
⫫	≉	⊫	⫫	جی ہاں	نہیں
‡	⊨	±	⊩	جی ہاں	نہیں
└	⊢	┐	⊫	جی ہاں	نہیں
⥲	⥲	≈	⇌	جی ہاں	نہیں
⊁	⋚	⊅	⊁	جی ہاں	نہیں
∼	⩀	∼	⋖	جی ہاں	نہیں
⊩	⟦	⊫	≉	جی ہاں	نہیں

C I

Symbol Search B

Ages 8–16

SAMPLE ITEMS

⊕	⊖	⊕	∟	<	⊢	~	جی ہاں	نہیں
↝	∟	ǂ	∩	♈	⋚	⊞	جی ہاں	نہیں

PRACTICE ITEMS

⊫	<	↝	⊫	±	⋖	⊖	جی ہاں	نہیں
≈	⊖	∩̄	±	┘	ǂ	♈	جی ہاں	نہیں

Proceed to page 8

B *(Continued)*

⊏	~	∩	≉	≁	∪	⊢	جی ہاں	نہیں
⊖	⊠	⇋	⊖	⊞	⊤	⊍	جی ہاں	نہیں
¬	⟧	~	⌋	⟧	♈	≬	جی ہاں	نہیں
⊗	⊩	⇌	⊠	±	‡	⊄	جی ہاں	نہیں
↝	~	⌒	⥲	⌋	↷	⊪	جی ہاں	نہیں
⋛	⊫	⊩	⊠	⊀	⊣	⌒	جی ہاں	نہیں
↝	⥶	⊀	⥲	⇋	≬	↷	جی ہاں	نہیں
⊃	⋖	⊀	∪	⋖	⊄	↷	جی ہاں	نہیں
⩃	⥲	≉	¬	∩	∪	⥱	جی ہاں	نہیں
↝	⊪	⊩	≉	~	⊪	≐	جی ہاں	نہیں
⊩	≐	⊣	∟	⊨	⊪	↼	جی ہاں	نہیں
≉	≬	≉	⩃	≁	♈	∮	جی ہاں	نہیں
⊅	⋐	⌓	∟	⟦	¬	⊃	جی ہاں	نہیں
⊅	≐	≐	⊩	⊏	⊪	⊀	جی ہاں	نہیں
⌒	⋖	⋖	⊢	⋗	♈	⌒	جی ہاں	نہیں

C I

B *(Continued)*

⋔	⊁	⊀	∩	≮	⊙	⊁	جی ہاں	نہیں
∮	≉	≬	↜	⇄	‡	∼	جی ہاں	نہیں
⊞	⊖	⊍	⊕	±	∪	⊢	جی ہاں	نہیں
⋛	⊅	≉	⋖	⊍	⊅	⇌	جی ہاں	نہیں
‡	⊩	⊫	±	≉	↜	⌝	جی ہاں	نہیں
⊙	∩	⊕	⊂	⊙	⊓	⋗	جی ہاں	نہیں
⋇	⊗	⊗	⊁	⊄	⇌	⋇	جی ہاں	نہیں
⟦	∟	⊧	⌝	⊏	⌟	⊍	جی ہاں	نہیں
≬	⋔	≬	↜	∩	↝	⊅	جی ہاں	نہیں
⋖	≮	⋖	⋚	⊀	≮	⋗	جی ہاں	نہیں
⧺	⩦	⊧	⊣	≉	⌝	⊫	جی ہاں	نہیں
⇄	∼	↼	⇋	⇉	∼	↝	جی ہاں	نہیں
⊖	⊍	∪	⊅	⊍	⊗	⊁	جی ہاں	نہیں
>	⊀	⋗	⇋	⊅	⌟	⋗	جی ہاں	نہیں
±	⊏	‡	⊂	⌟	⊣	±	جی ہاں	نہیں

C I

Appendix F1

LIST OF RANDOMLY SELECTED EDUCATIONAL INSTITUTES UNDER ADMINISTRATIVE CONTROL OF FEDRAL DIRECTORATE OF EDUCATION

1. Islamabad Model College for Boys, I-8/3
2. Islamabad Model College for Boys, F-11/3
3. Islamabad Model College for Girls, F-6/2
4. Islamabad Model College for Girls, G-10/2
5. Islamabad Model College for Girls, G-9/2
6. Islamabad Model School for Boys (VI - X) No. 2, I-9/4
7. Islamabad Model School for Girls (I - X), Malpur
8. Islamabad Model School for Boys (VI - X), Sangjani
9. Islamabad Model School for Girls (I - X), Shah Allah Ditta
10. Islamabad Model School for Boys (I - VIII), Kot Hathial
11. Islamabad Model School for Girls (I - VIII), Noon
12. Islamabad Model School (I - V) No. 1, G-6/2
13. Islamabad Model School (I - V) No. 4, G-9/2
14. Islamabad Model School (I - V) No. 2, I-10/1
15. Islamabad Model School (I - V) No. 2, G-9/4
16. Islamabad Model School (I - V), G-11/1
17. Islamabad Model School for Boys (I - V), Sarai Kharbooza
18. Islamabad Model School for Boys (I - X), Naugazi
19. Islamabad Model School for Girls (I - VIII), Bobari
20. Islamabad Model School (I - V), F-10/2

Appendix F2

NO. F. 1-06/2010/TP- DDT/FDE
Government of Pakistan
Federal Directorate of Education
****

Islamabad, December 28, 2011

Subject: **<u>ISSUANCE OF PERMISSION FOR TEST ADMINISTRATION IN VARIOUS SCHOOLS AND COLLEGES UNDER FEDERAL DIRECTORATE OF EDUCATION.</u>**

I am directed to convey the approval of the Competent Authority to allow Ms. Saima Ambreen (Ph.D. Scholar, NIP) Quaid-i-Azam University, Islamabad to carry out the project as part of her Ph.D. research work titled. "Wechsler Intelligence Scale for Children (WISC-IV) South Asia): Adaptation, Translation and Standardization in Pakistan"

2. You are requested to extend maximum cooperation to Ms. Saima Ambreen to make this research a success, please.

3. This is issued with the approval of Director (Traininig)

(ZAFAR IQBAL)
Assistant Director (Training)

The Principals,

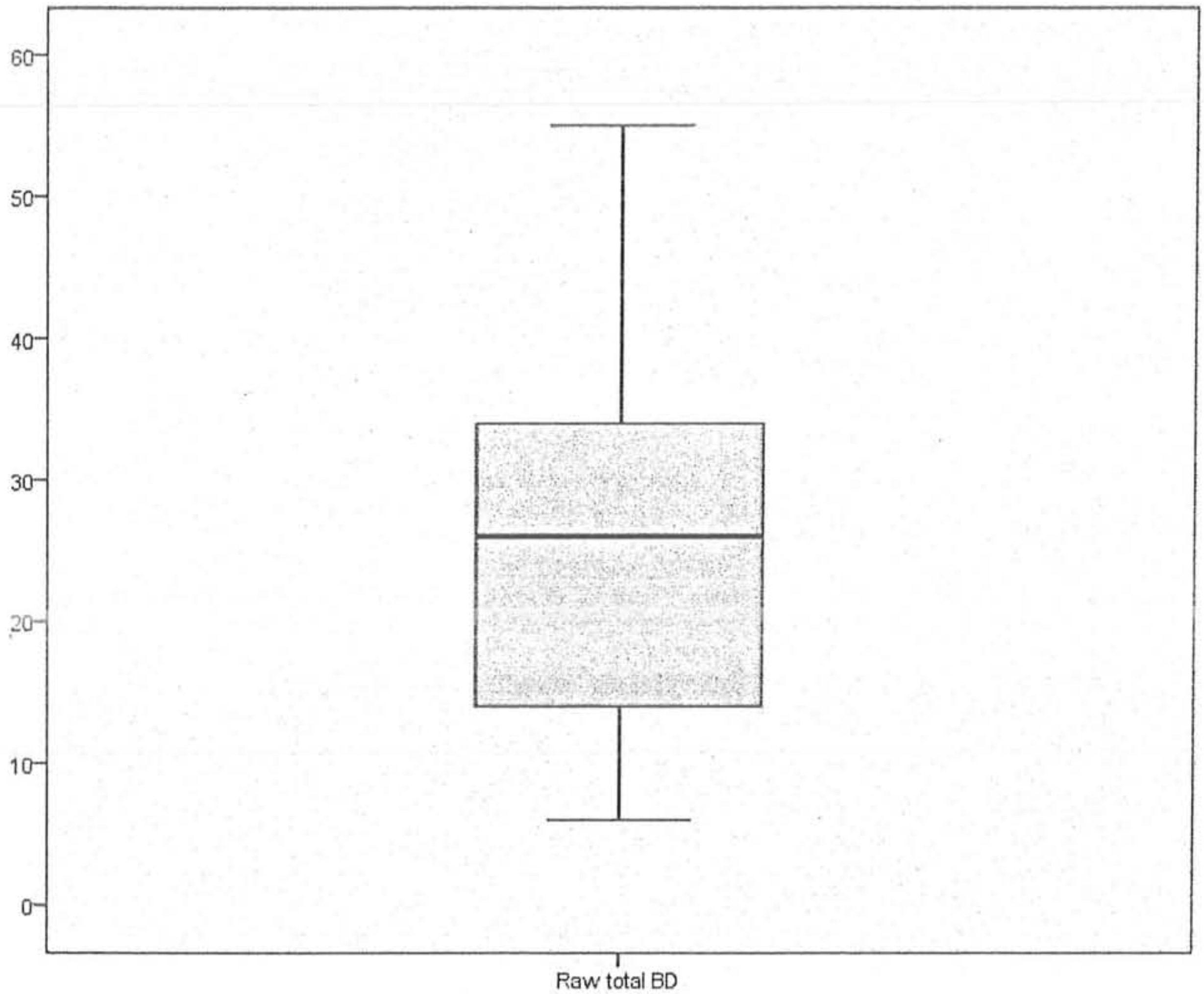
IMCB, I-8/3, IMCB, F-11/3, IMCG, F-6/2, IMCG, F-6/2, IMCG, G-10/2, IMCG, G-9/2, IMSB, (VI-X) NO.2 I-9/4, IMSG, (I-X), MALPUR, IMSB, (VI-X), SANGJANI, IMSG,(I-X), SHAH ALLAH DITTA, IMSB, (I-VIII), KOT HATHIAL, IMSG, (I-VIII), NOON, IMS, (I-V), NO.1, G-6/2, IMS, (I-V), NO.4, G-9/2, IMS, (I-V), NO.2, I-10/1, IMS, (I-V), NO.2, G-9/4, IMS, (I-V), G-11/1, IMSB, (I-V),SARAI KHARBOOZA, IMSB, (I-X), NAUGAZI, IMSG, (I-VIII), BOBRI, IMS, (I-V), F-10/2.

Copy for information: -

- *AEO,s Concerned.*
- *P.A to Director (Training) FDE IslamabaD.*
- *P.A to Director (Schools) Male FDE Islamabad.*
- *PA to Director (Schools) Female, FDE Islamabad.*

Appendix G1

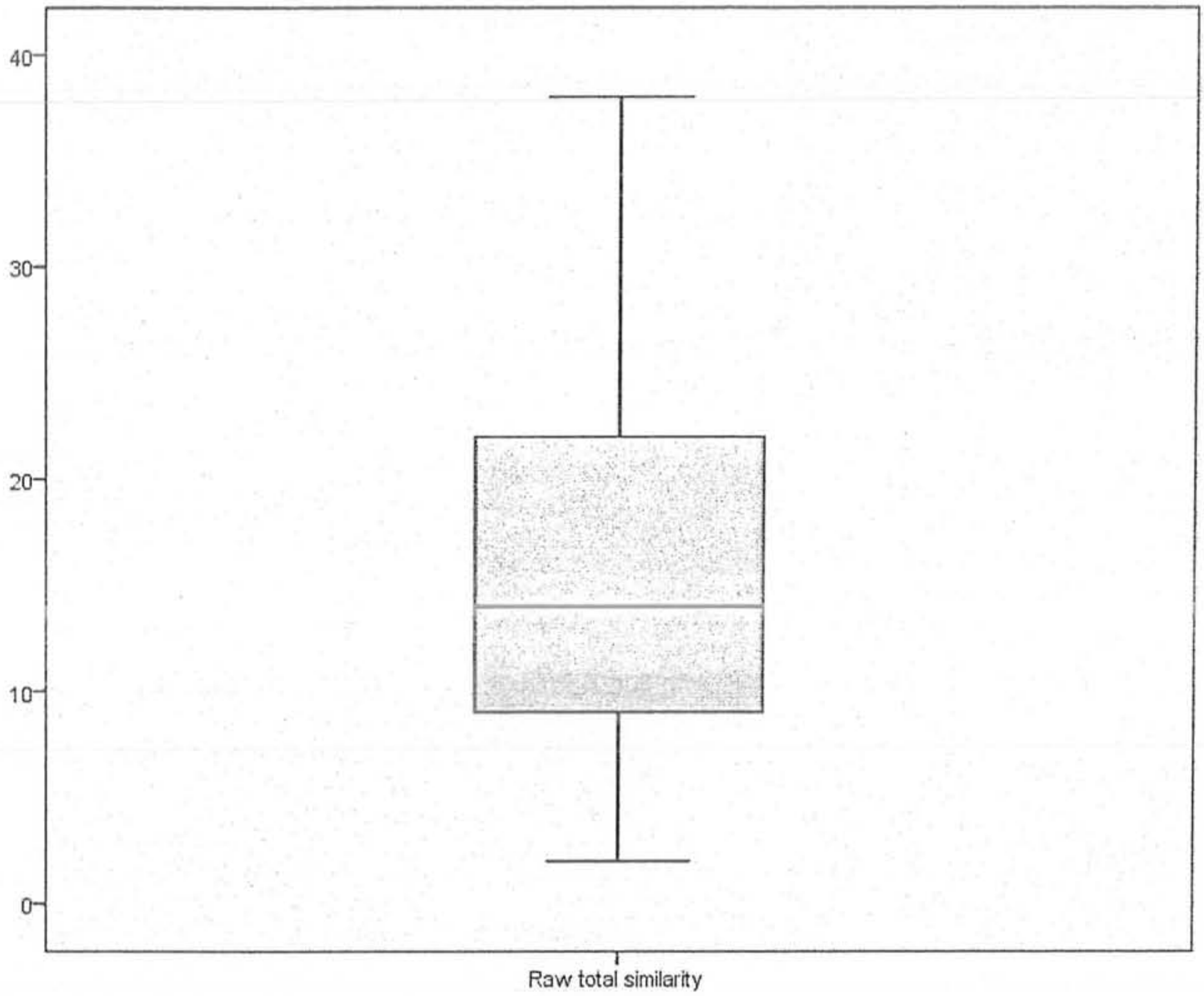
Box Plot for Block Design Subtest



Box plot depicting score distribution of Block Design subtest ($N = 88$)

Appendix G2

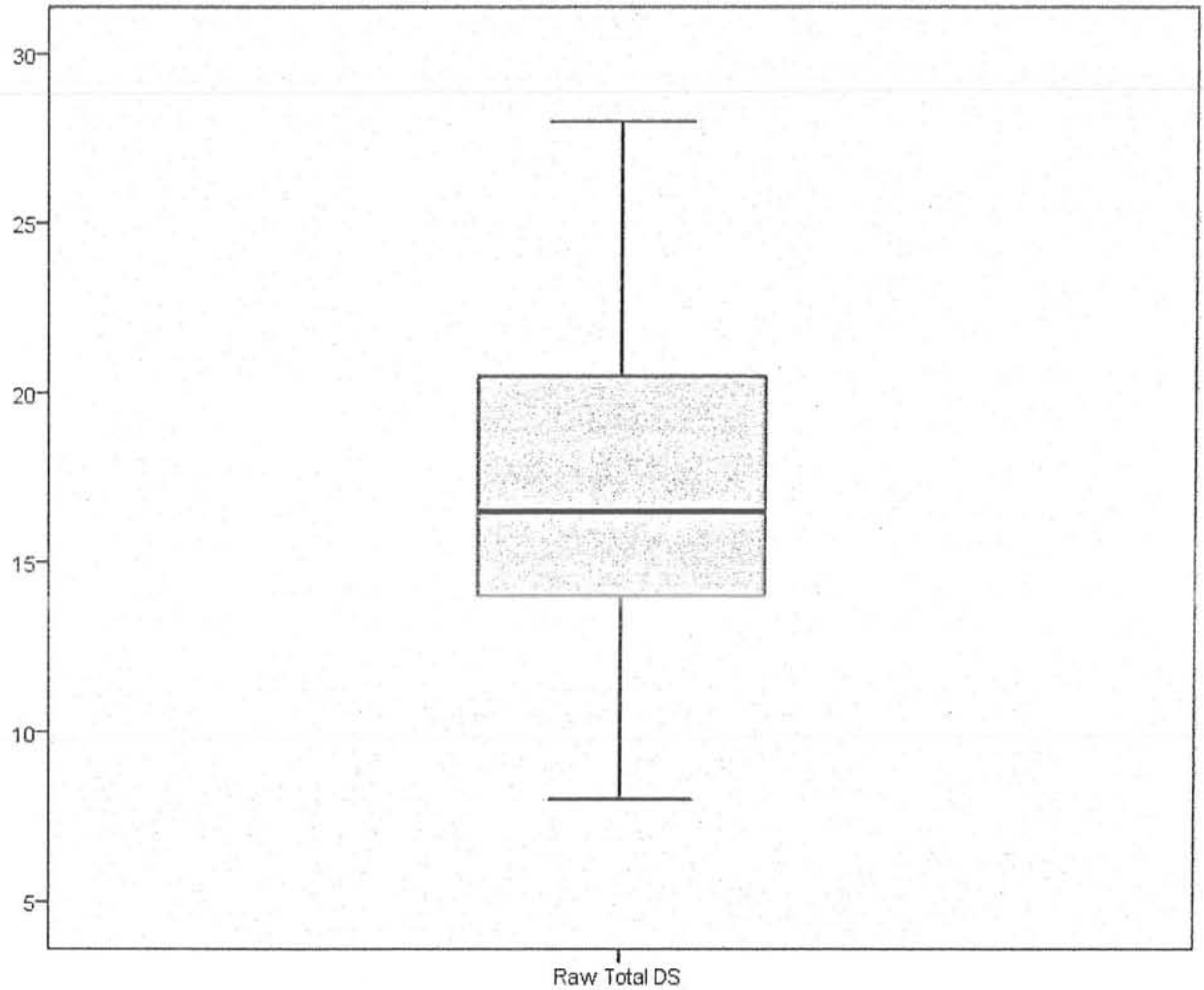
Box Plot for Similarities Subtest



Box plot depicting score distribution of Similarities subtest ($N = 88$)

Appendix G3

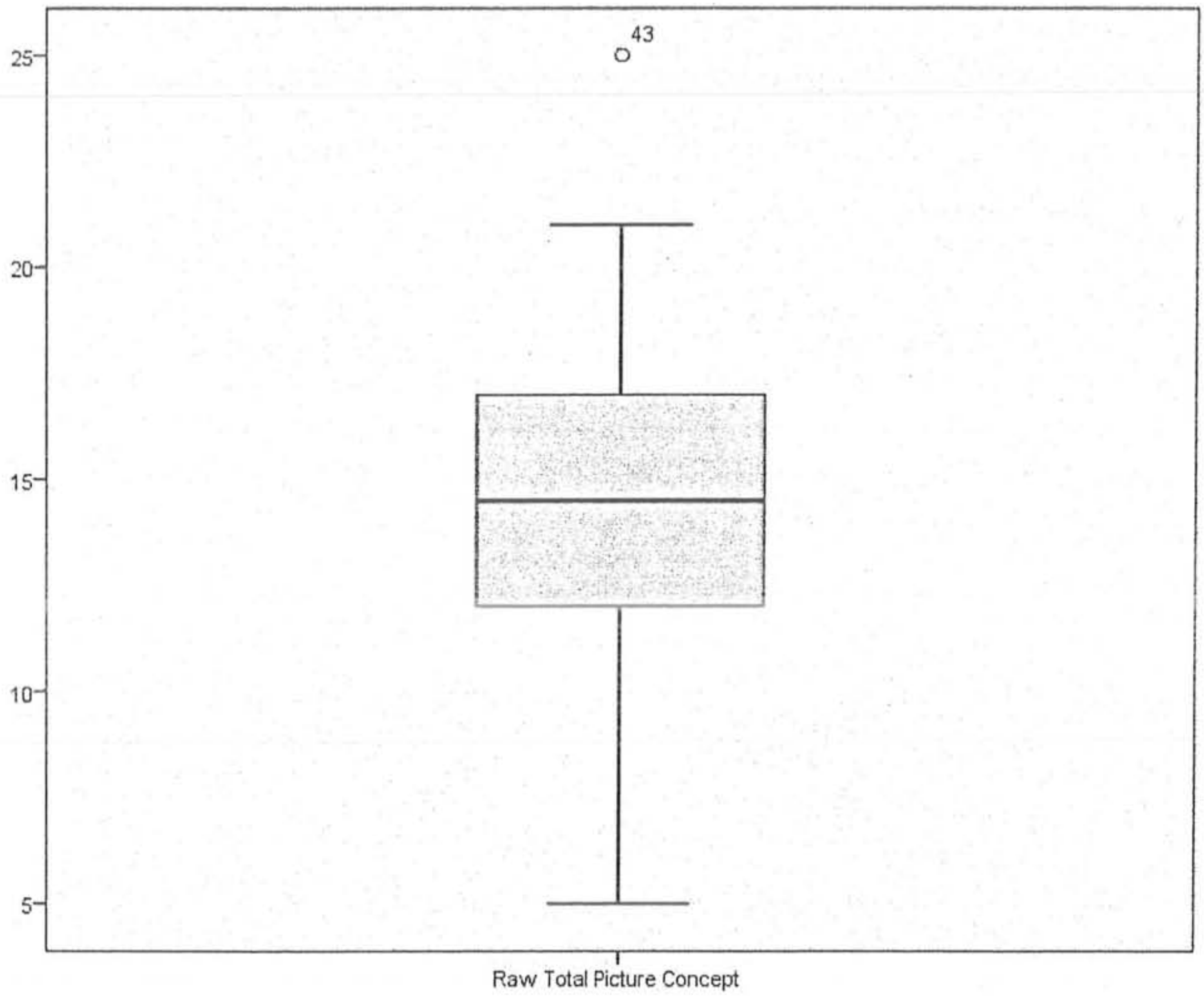
Box Plot for Digit Span Subtest



Box plot depicting score distribution of Digit Span subtest ($N = 88$)

Appendix G4

Box Plot for Picture Concept Subtest

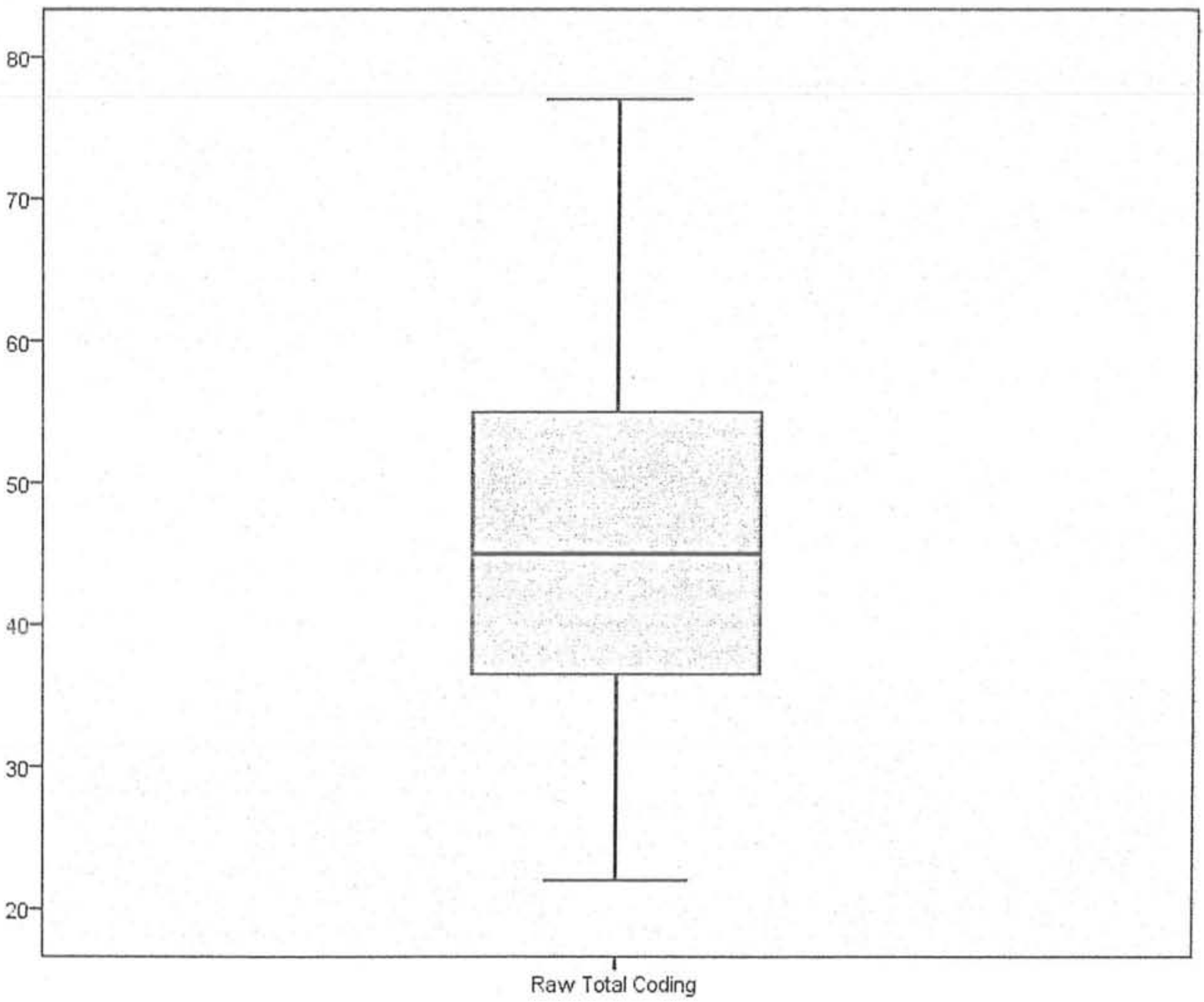


Box plot depicting score distribution and existence of an outlier in Picture Concept subtest

($N = 88$)

Appendix G5

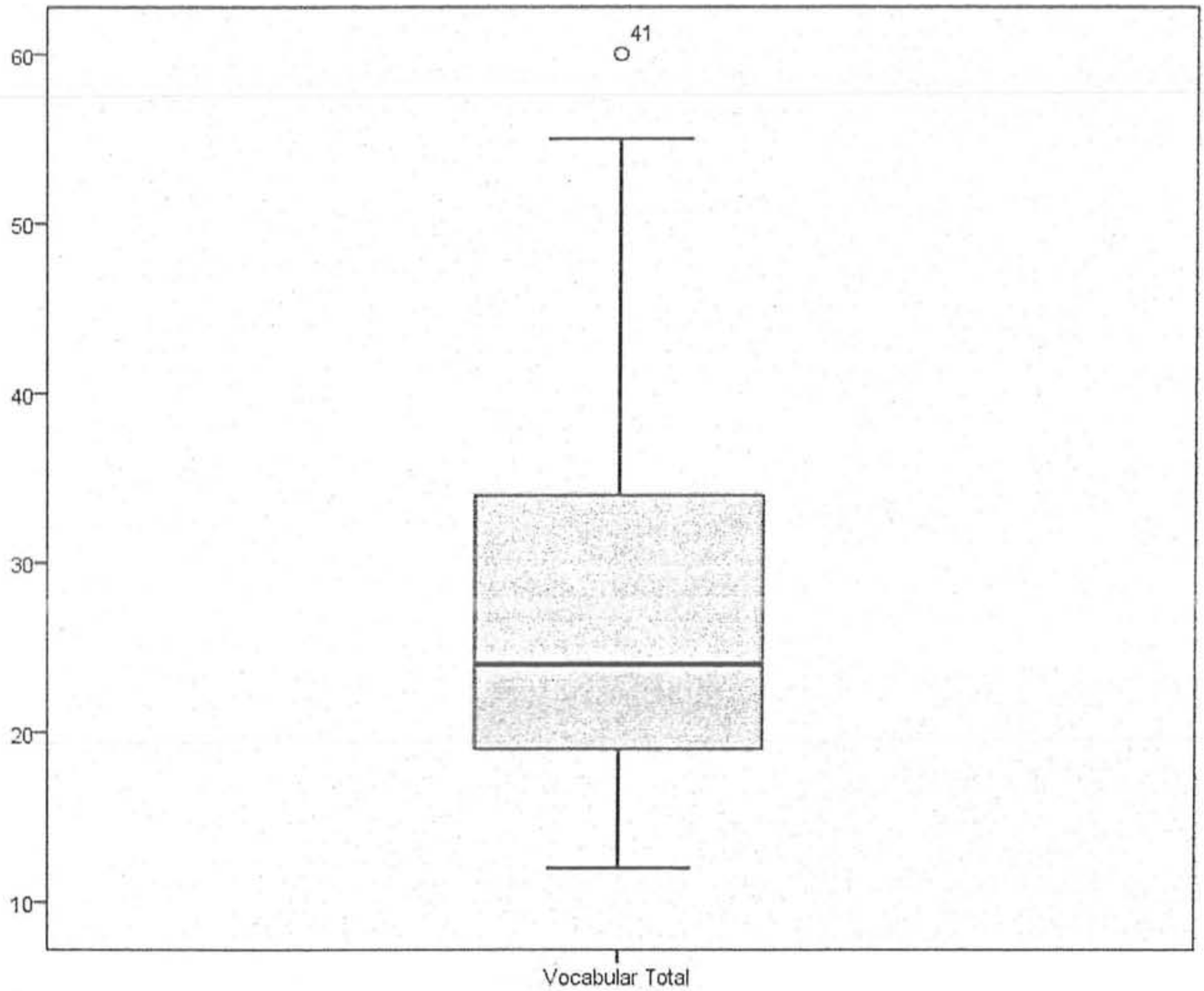
Box Plot for Coding Subtest



Box plot depicting score distribution of Coding subtest ($N = 88$)

Appendix G6

Box Plot for Vocabulary Subtest

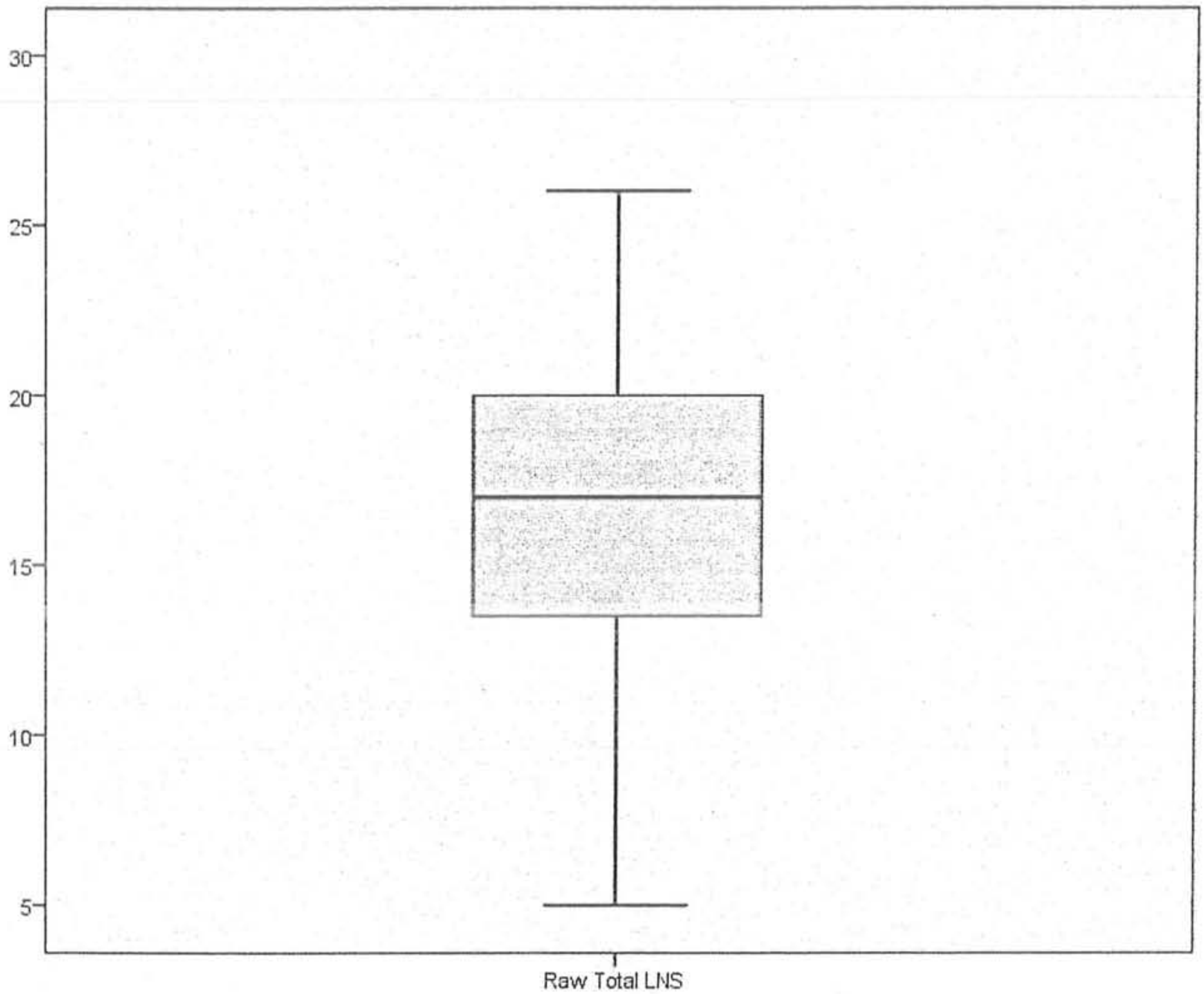


Box plot depicting score distribution and existence of an outlier in Vocabulary subtest

($N = 88$)

Appendix G7

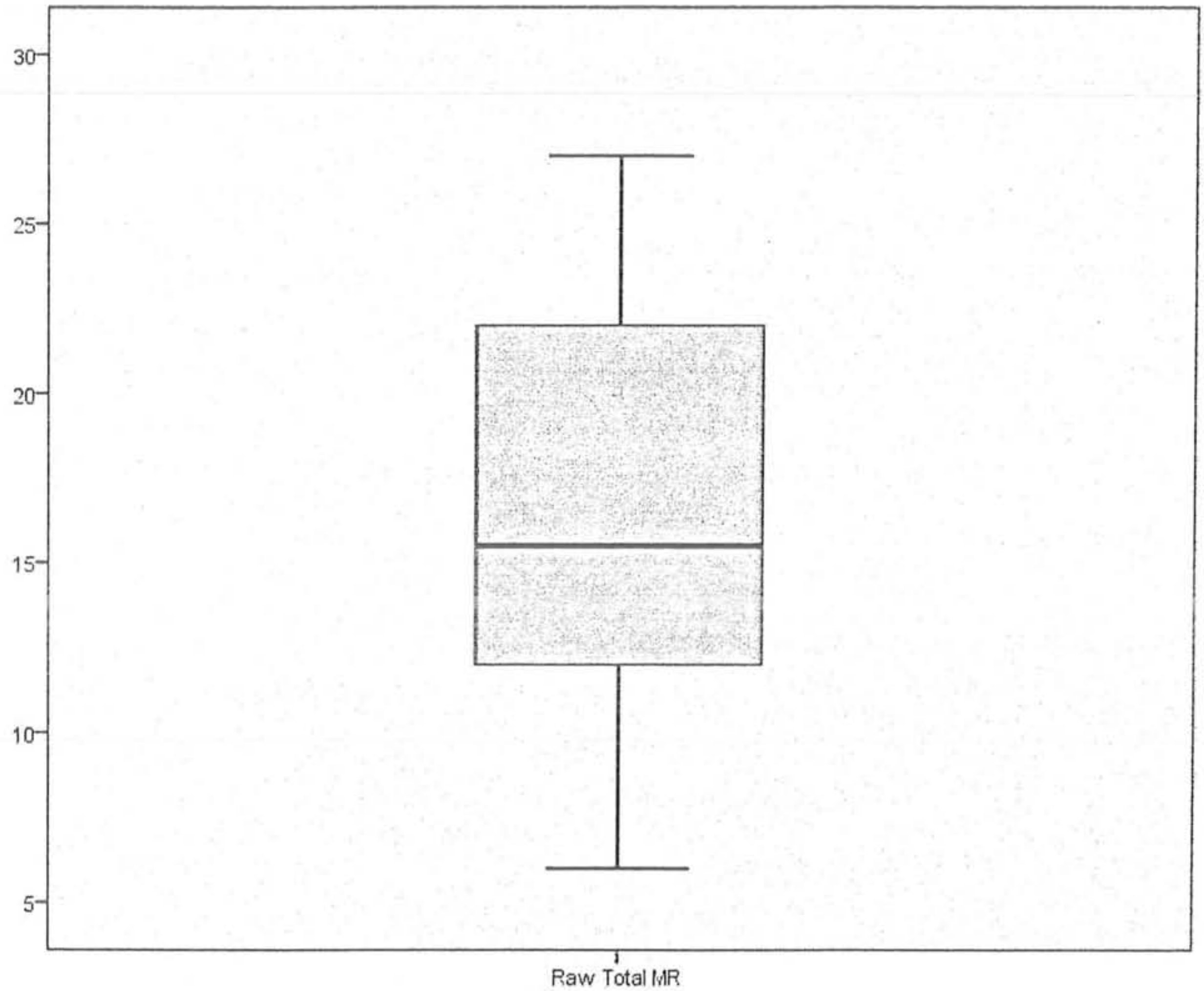
Box Plot for Letter-Number-Sequence Subtest



Box plot depicting score distribution of Letter-Number-Sequencing subtest ($N = 88$)

Appendix G8

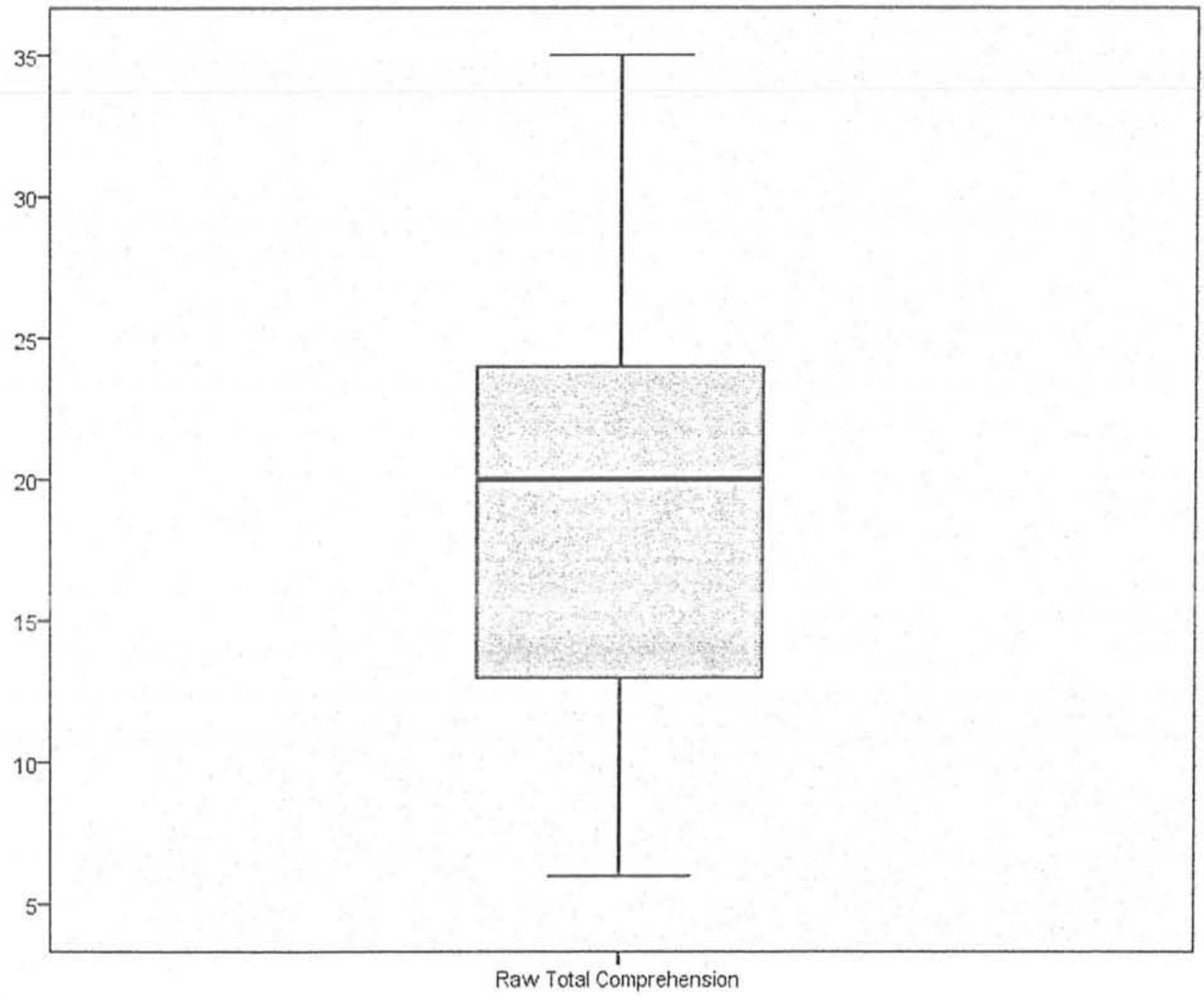
Box Plot for Matrix Reasoning Subtest



Box plot depicting score distribution of Matrix Reasoning subtest ($N = 88$)

Appendix G9

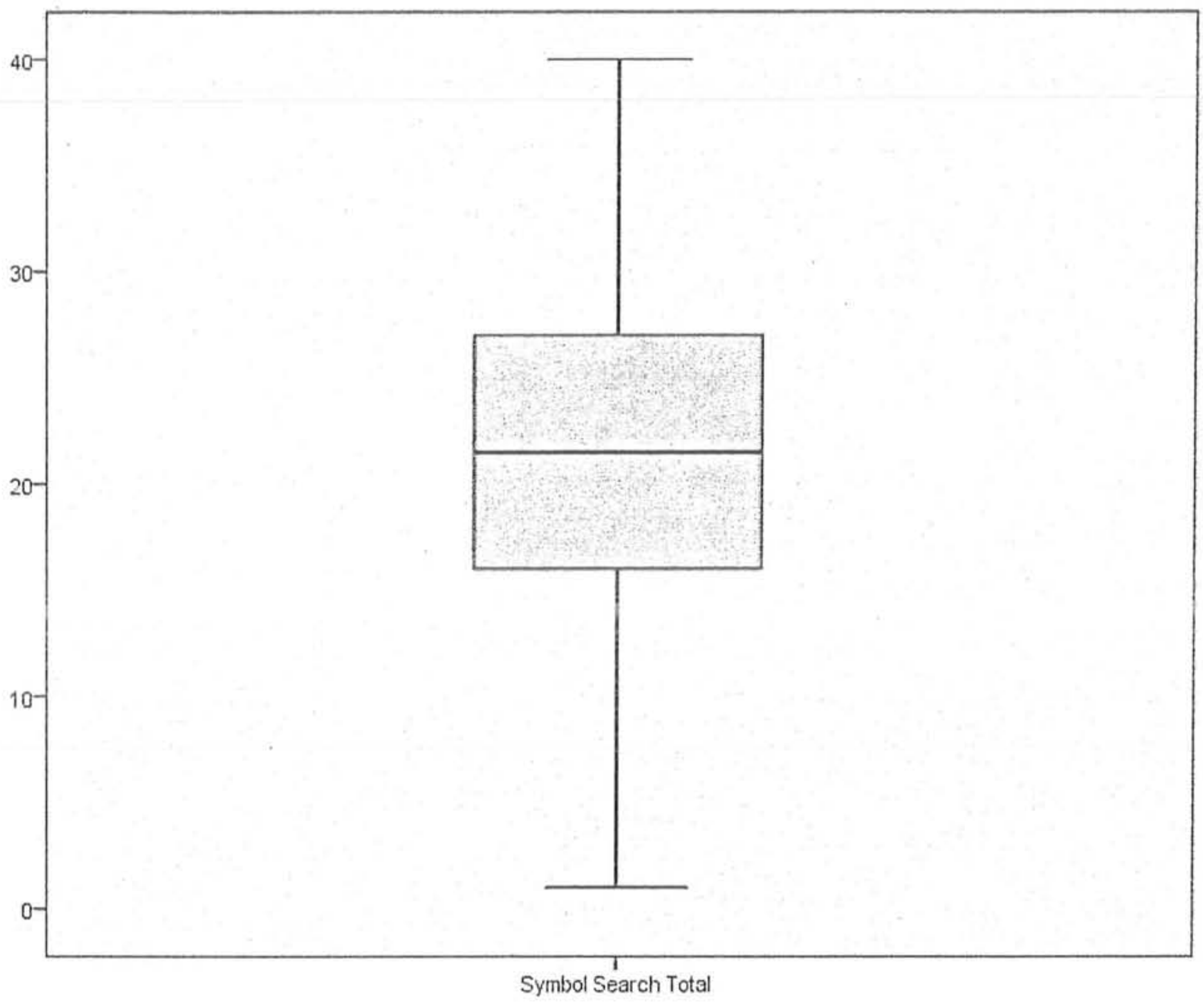
Box Plot for Comprehension Subtest



Box plot depicting score distribution of Comprehension subtest ($N = 88$)

Appendix G10

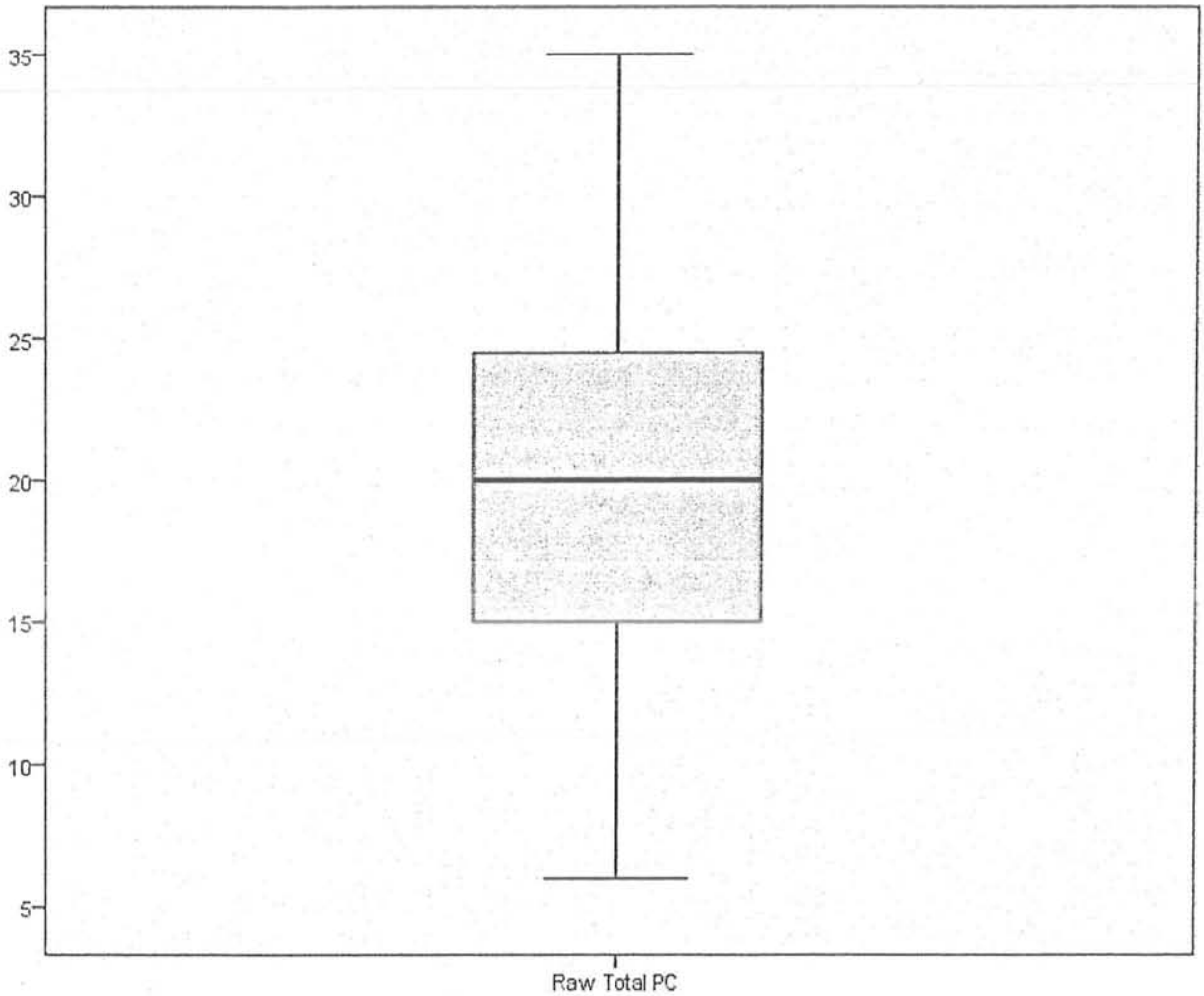
Box Plot for Symbol Search Subtest



Box plot depicting score distribution of Symbol Search subtest ($N = 88$)

Appendix G11

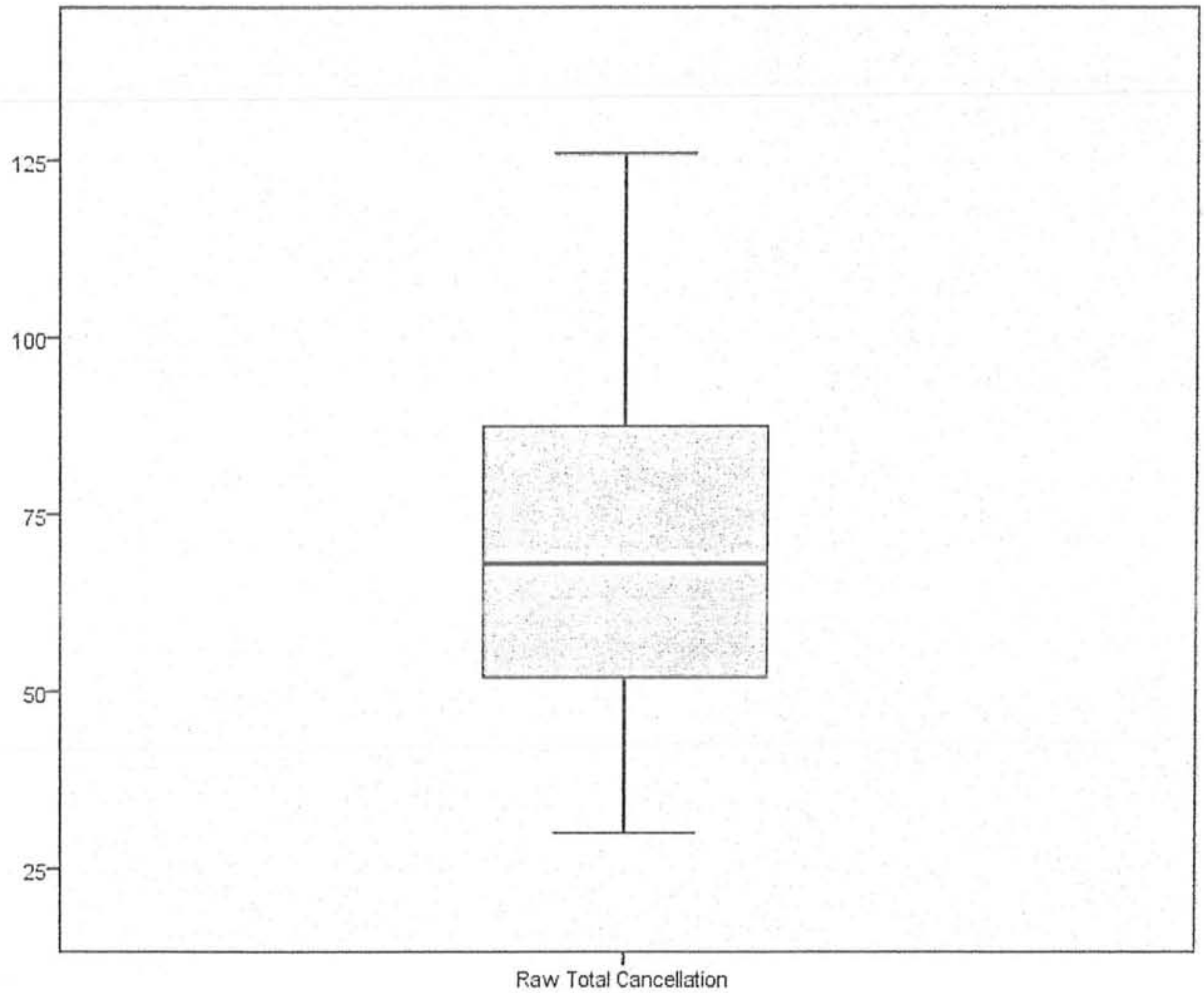
Box Plot for Picture Completion Subtest



Box plot depicting score distribution of Picture Completion subtest ($N = 88$)

Appendix G12

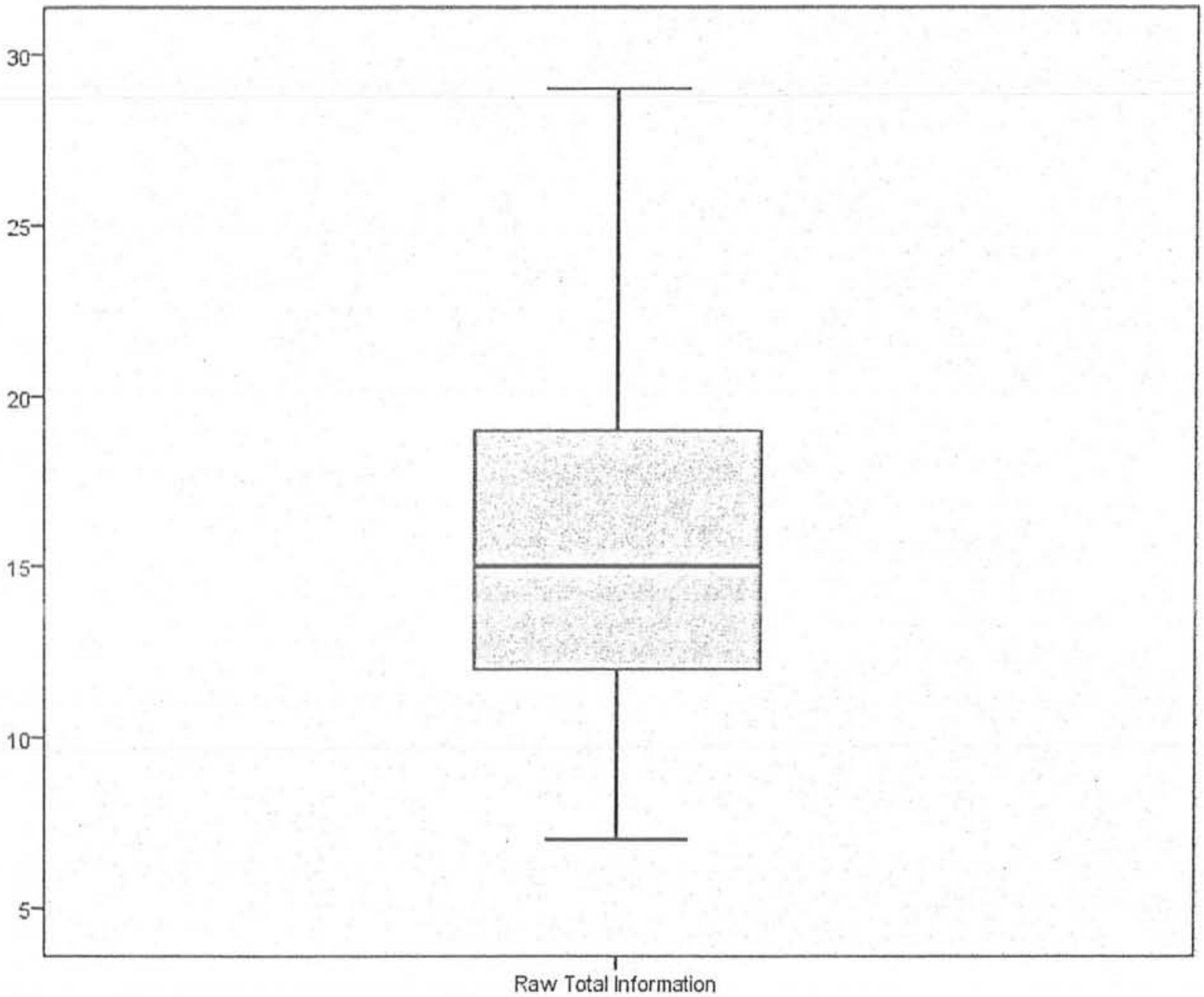
Box Plot for Cancelation Subtest



Box plot depicting score distribution of Cancelation subtest ($N = 88$)

Appendix G13

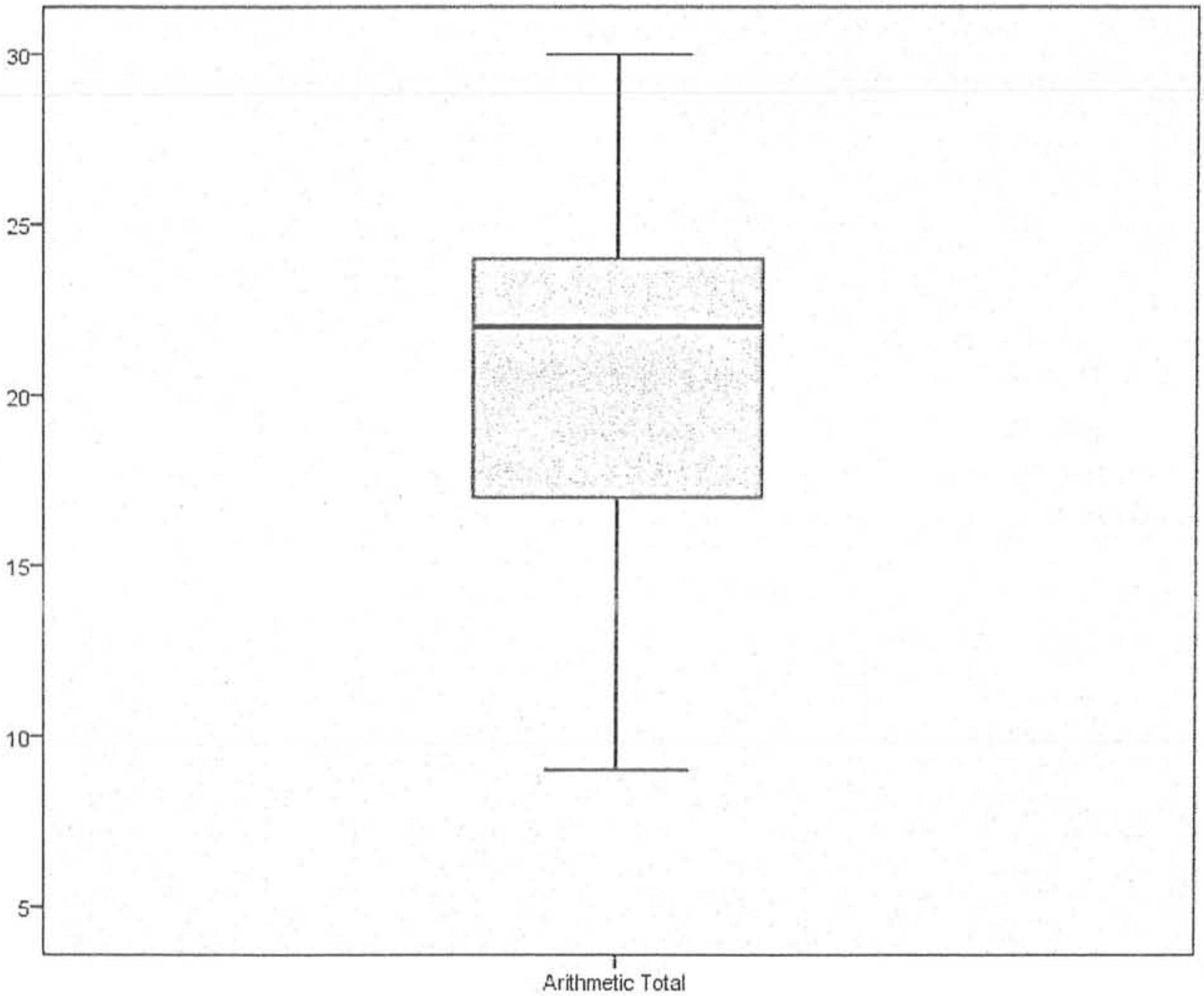
Box Plot for Information Subtest



Box plot depicting score distribution of Information subtest ($N = 88$)

Appendix G14

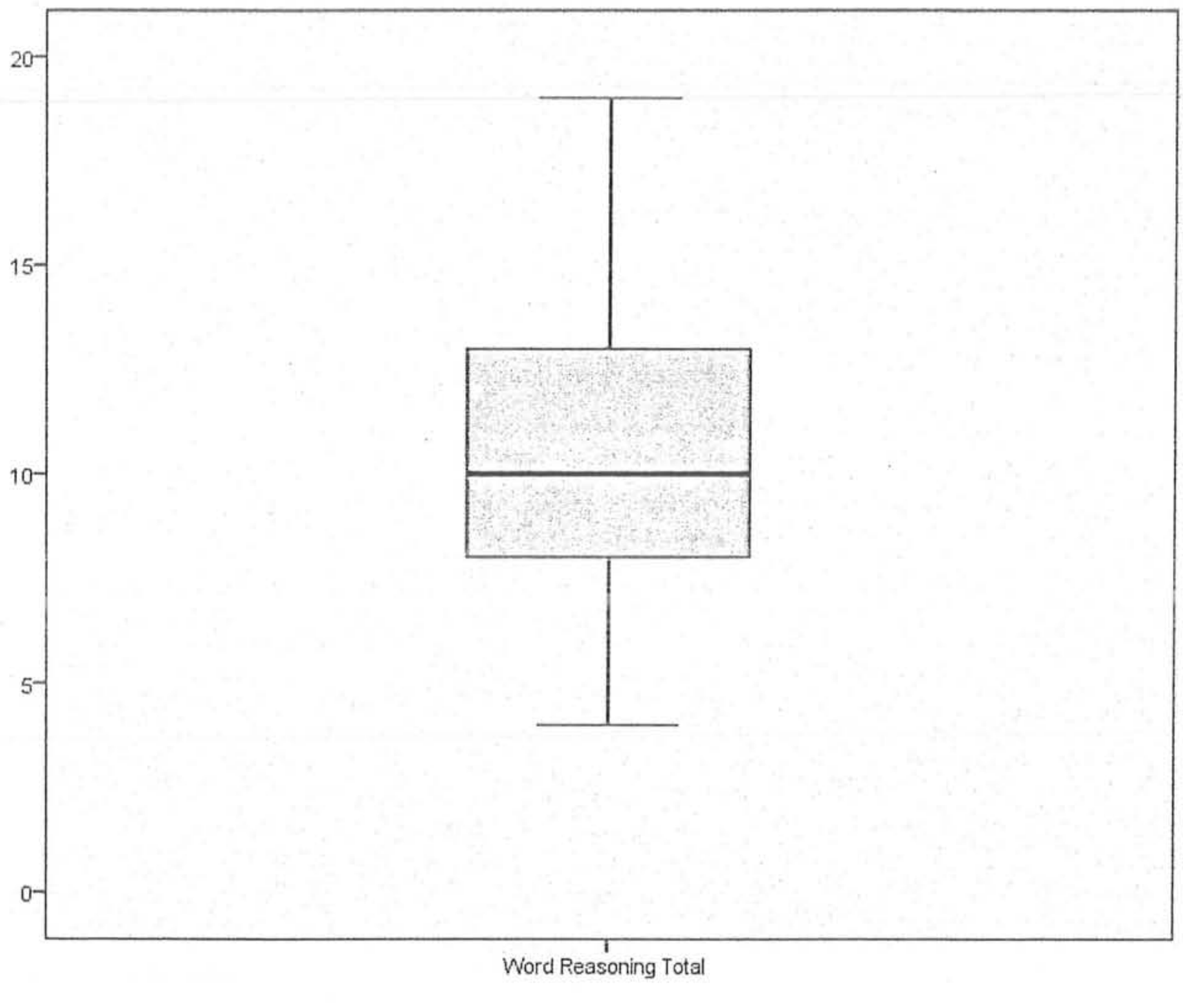
Box Plot for Arithmetic Subtest



Box plot depicting score distribution of Arithmetic subtest ($N = 88$)

Appendix G15

Box Plot for World Reasoning Subtest



Box plot depicting score distribution of Word Reasoning subtest ($N = 88$)

Schematic Plot – Block Design

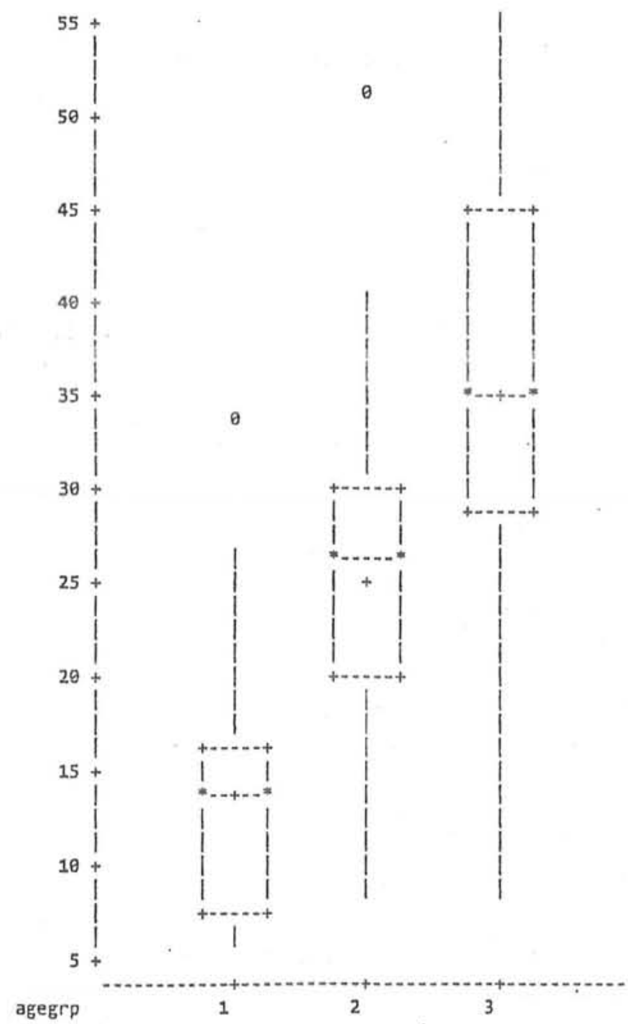
Key

- Vertical Axis= Subtest Total Raw Score
- Horizontal Axis= Age Groups

  1 = 6 – 8 years (N=24)

  2 = 9 – 12 years (N=32)

  3 = 13 – 16 years (N=32)



Schematic Plot – Similarities

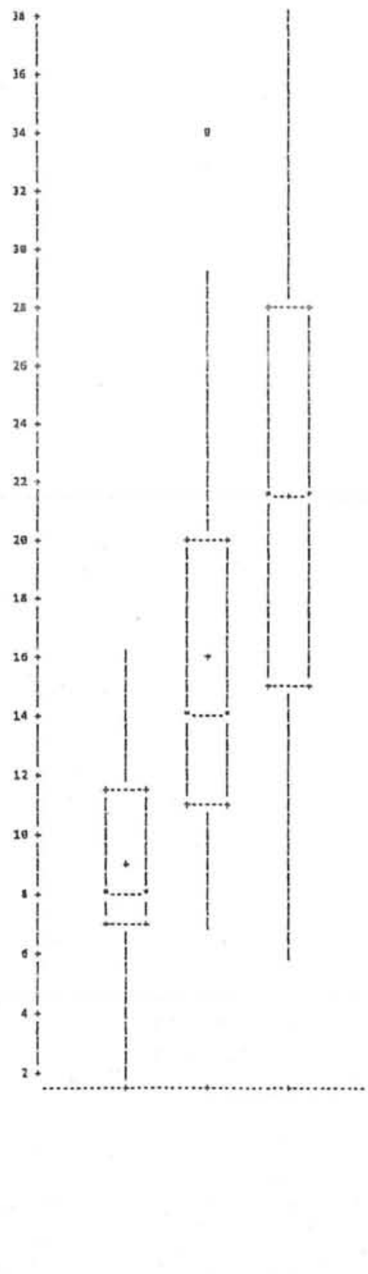
Key

- Vertical Axis= Subtest Total Raw Score
- Horizontal Axis= Age Groups

  1 = 6 – 8 years (N=24)

  2 = 9 – 12 years (N=32)

  3 = 13 – 16 years (N=32)



Schematic Plot – Vocabulary

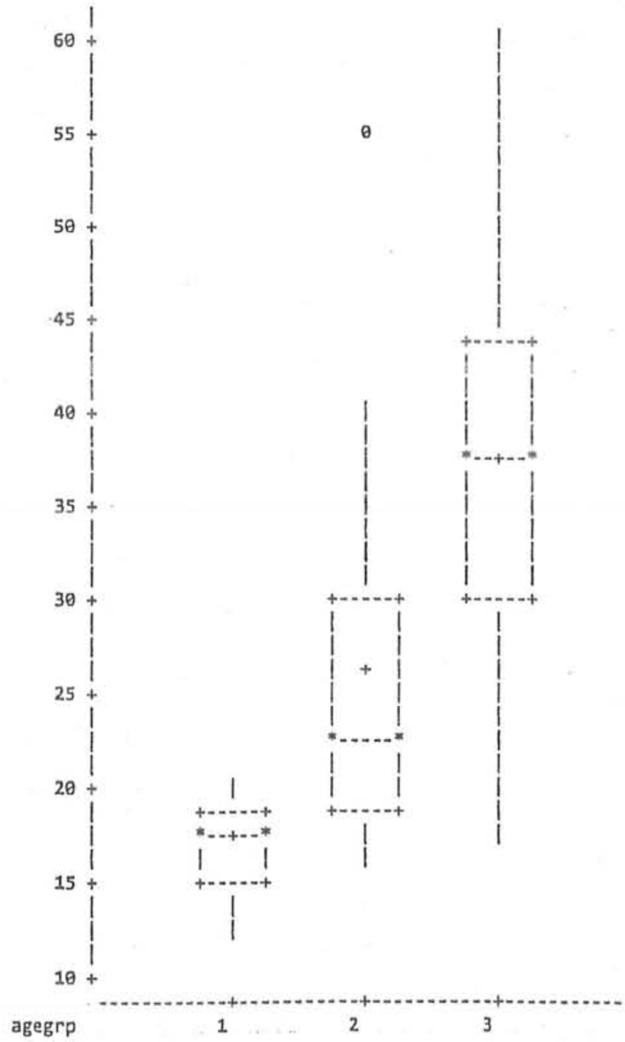
Key

- Vertical Axis= Subtest Total Raw Score
- Horizontal Axis= Age Groups

  1 = 6 – 8 years (N=24)

  2 = 9 – 12 years (N=32)

  3 = 13 – 16 years (N=32)



Letter-Number-Sequencing

Appendix H4

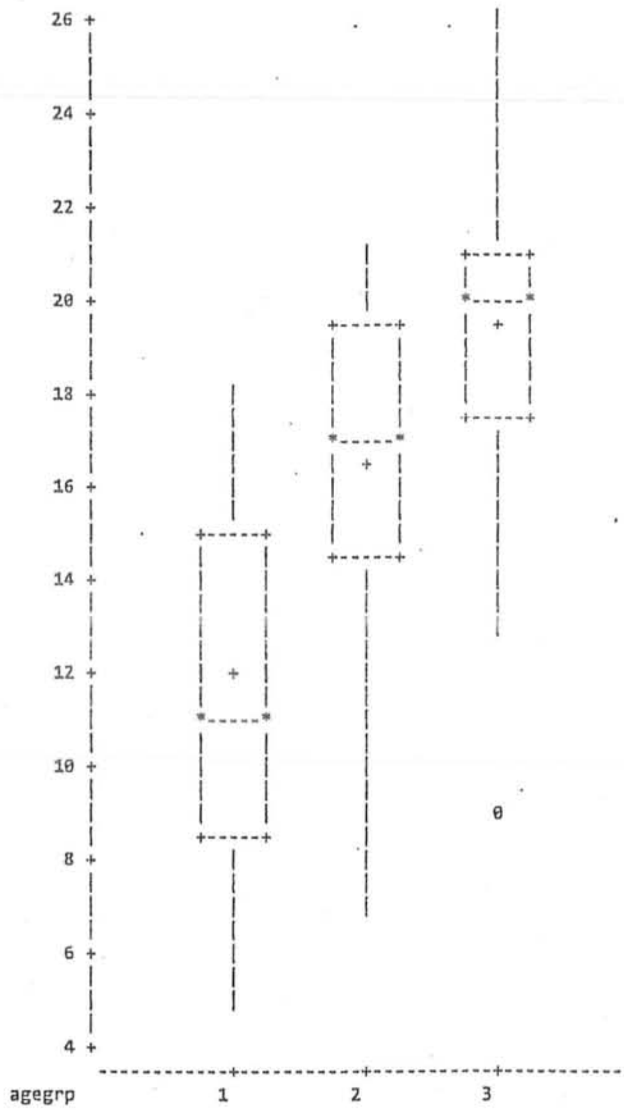
Key

- Vertical Axis= Subtest Total Raw Score
- Horizontal Axis= Age Groups

  1 = 6 – 8 years (N=24)

  2 = 9 – 12 years (N=32)

  3 = 13 – 16 years (N=32)



Schematic Plot – Matrix Reasoning

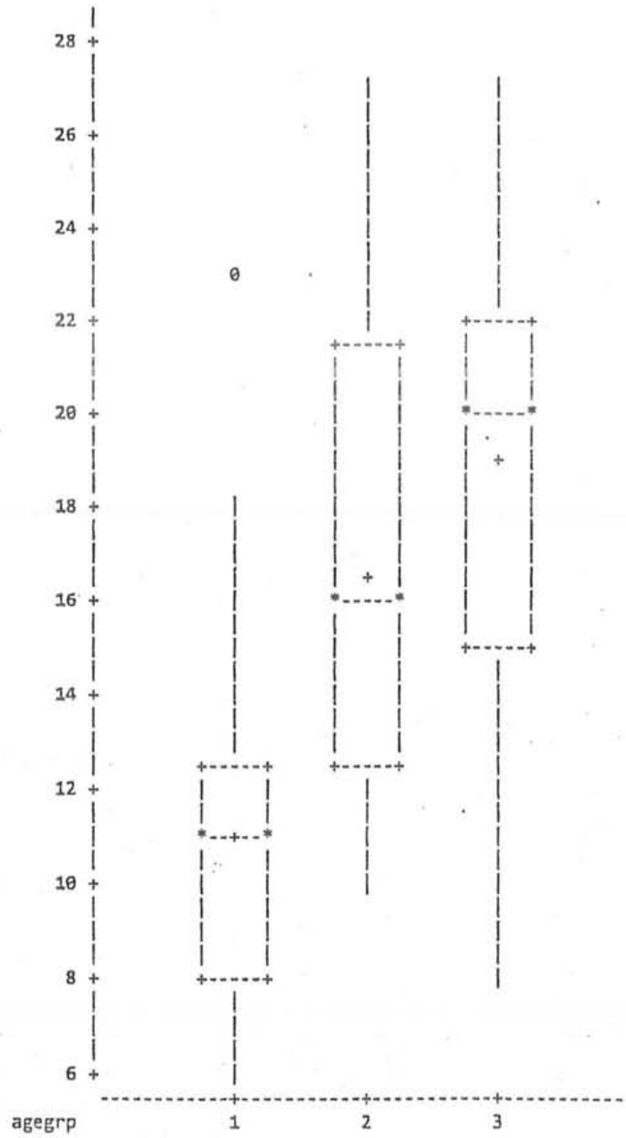
Key

- Vertical Axis= Subtest Total Raw Score
- Horizontal Axis= Age Groups

  1 = 6 – 8 years (N=24)

  2 = 9 – 12 years (N=32)

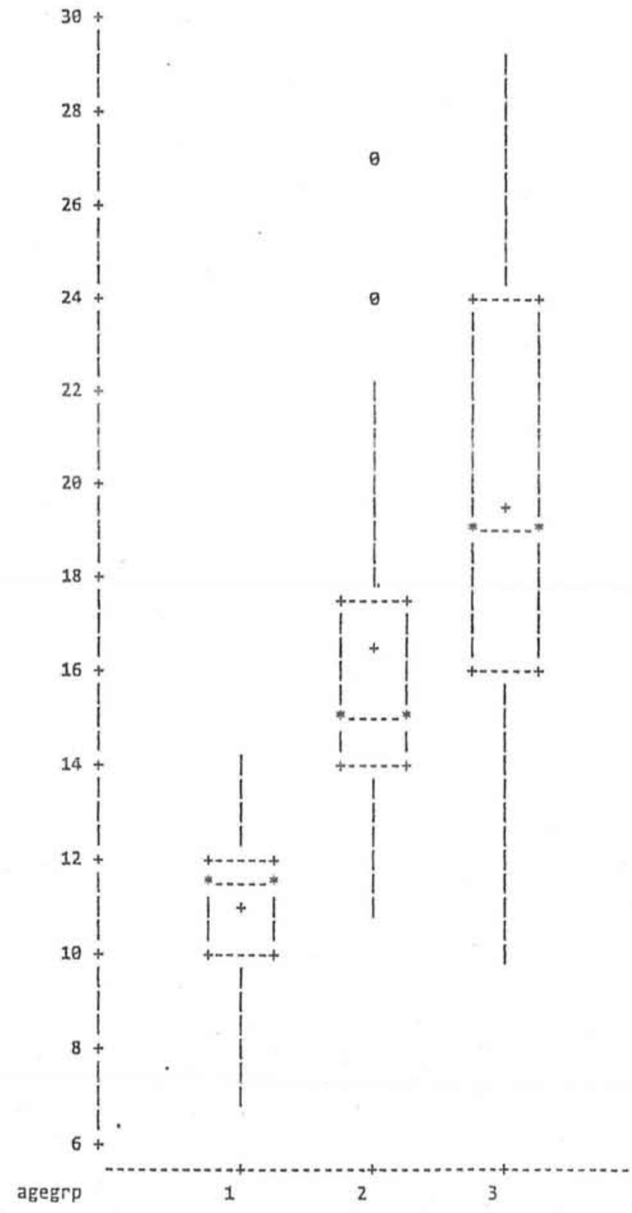
  3 = 13 – 16 years (N=32)



Schematic Plot – Information

Key

- Vertical Axis= Subtest Total Raw Score
- Horizontal Axis= Age Groups
  1 = 6 – 8 years (N=24)
  2 = 9 – 12 years (N=32)
  3 = 13 – 16 years (N=32)



Schematic Plot - Arithmetic

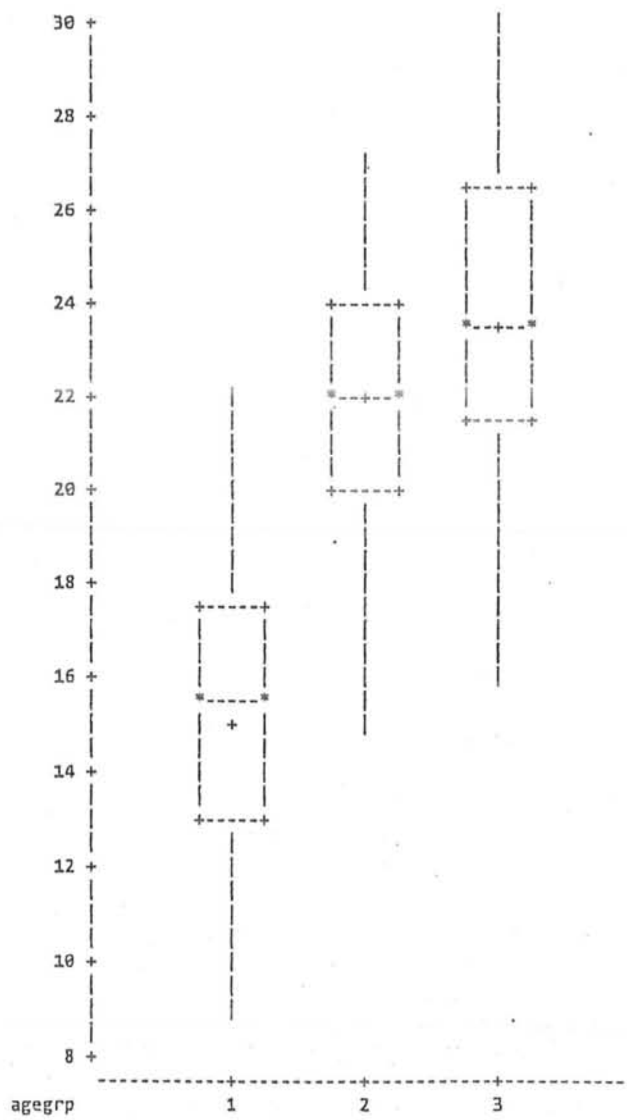
Key

- Vertical Axis= Subtest Total Raw Score
- Horizontal Axis= Age Groups

  1 = 6 – 8 years (N=24)

  2 = 9 – 12 years (N=32)

  3 = 13 – 16 years (N=32)



Appendix I

Lists of Re-ordered Items of WISC-IV Urdu Subtests along with Difficulty Index and Original Order (after Tryout II)

SIMILARITIES			INFORMATION		
New Order of Administration	pv0	Original order	New Order of Administration	pv0	Original order
1	0.92	1	1	1.00	1
2	0.97	2	2	1.00	2
3	0.98	3	3	1.00	3
4	0.96	4	4	1.00	4
5	0.92	6	5	1.00	5
6	0.87	5	6	0.97	8
7	0.71	7	7	0.97	9
8	0.64	9	8	0.95	6
9	0.54	10	9	0.95	7
10	0.36	8	10	0.89	11
11	0.31	12	11	0.89	12
12	0.29	13	12	0.89	10
13	0.30	11	13a	0.69	14
14	0.25	15	13b	-	New
15	0.22	14	14	0.56	13
16	0.21	17	15	0.51	15
17	0.14	19	16	0.45	17
18	0.14	16	17	0.33	16
19	0.13	21	18	0.32	19
20	0.11	20	19	0.26	20
21	0.10	18	20	0.24	24
22	0.05	22	21	0.19	18
23	0.00	23	22	0.19	21
			23	0.18	22
			24	0.13	26
			25	0.10	25
			26	0.09	28
			27	0.08	31
			28	0.07	30
			29	0.06	23
			30	0.06	27
			31	0.03	32
			32	0.02	29
			33	0.00	33

VOCABULARY

New Order of Administration	pv0	Original order
1	1.00	1
2	1.00	2
3	1.00	3
4	1.00	4
5	1.00	5
6	1.00	6
7	1.00	7
8	1.00	8
9	1.00	10
10	0.98	9
11	0.82	15
12	0.73	12
13	0.62	19
14a	-	New
14b	0.53	11
15	0.53	24
16	0.51	16
17	0.46	20
18	0.42	13
19	0.28	17
20	0.28	18
21	0.28	22
22	0.22	23
23	0.22	29
24	0.19	32
25	0.18	28
26	0.15	27
27	0.15	33
28	0.14	14
29	0.10	31
30	0.09	35
31	0.07	21
32	0.07	26
33	0.06	34
34	0.06	36
35	0.04	30
36a	0.00	25
36b	-	New

COMPREHENSION

New Order of Administration	pv0	Original order
1	1.00	1
2	0.97	2
3	0.90	7
4	0.88	4
5	0.89	8
6	0.87	5
7	0.86	10
8	0.81	3
9	0.75	11
10	0.67	6
11	0.69	9
12	0.63	15
13	0.57	12
14	0.48	13
15	0.34	14
16	0.15	18
17	0.14	16
18	0.14	19
19	0.09	20
20	0.06	17
21	0.06	21

PICTURE COMPLETION

New Order of Administration	MNO	Original order
1	1.00	1
2	1.00	2
3	0.99	3
4	0.97	5
5	0.92	4
6	0.91	7
7	0.91	10
8	0.91	12
9	0.85	8
10	0.84	6
11	0.83	11
12	0.84	21
13	0.80	13
14	0.77	9
15	0.68	16
16	0.64	14
17	0.65	15
18	0.56	22
19	0.49	17
20	0.45	19
21	0.44	20
22	0.42	18
23	0.40	26
24	0.43	27
25	0.31	25
26	0.27	23
27	0.24	31
28	0.22	33
29	0.20	34
30	0.19	32
31	0.17	29
32	0.15	28
33	0.11	24
34	0.07	30
35	0.06	36
36	0.06	37
37	0.05	35
38	0.03	38

PICTURE CONCEPT

New Order of Administration	MNO	Original order
1	1.00	1
2	1.00	2
3	1.00	3
4	0.98	5
5	0.90	4
6	0.90	6
7	0.86	10
8	0.84	7
9	0.81	8
10	0.78	11
11	0.77	9
12	0.84	13
13	0.72	12
14	0.60	14
15	0.63	15
16	0.45	17
17	0.27	20
18	0.22	18
19	0.22	19
20	0.20	16
21	0.13	21
22	0.10	22
23	0.09	25
24	0.07	24
25	0.05	23
26	0.02	26
27	0.01	28
28	0.00	27

MATRIX REASONING			ARITHMETICS		
New Order of Administration	MNO	Original order	New Order of Administration	MNO	Original order
1	1.00	1	1	1.00	1
2	1.00	2	2	1.00	2
3	1.00	3	3	1.00	3
4	1.00	6	4	1.00	4
5	0.99	5	5	0.99	5
6	0.94	4	6	0.97	6
7	0.85	7	7	1.00	7
8	0.80	9	8	0.99	8
9	0.83	12	9	0.99	9
10	0.77	10	10	0.97	10
11	0.68	8	11	0.93	11
12	0.63	11	12	0.92	12
13	0.51	13	13	0.92	13
14	0.51	14	14	0.83	14
15	0.42	15	15	0.77	15
16	0.70*	16	16	0.78	18
17	0.43	17	17	0.72	16
18	0.43	18	18	0.69	17
19	0.42	20	19	0.68	19
20	0.34	19	20	0.67	20
21	0.33	21	21	0.52	24
22	0.28	22	22	0.50	22
23	0.28	23	23	0.50	23
24	0.17	24	24	0.34	21
25	0.11	25	25	0.28	27
26	0.10	26	26	0.23	25
27	0.13	27	27	0.10	26
28	0.09	30	28	0.08	28
29	0.08	28	29	0.13	29
30	0.06	31	30	0.09	30
31	0.05	29	31	0.02	31
32	0.05	33	32	0.01	32
33	0.02	32	33	0.03	33
34	0.02	34	34	0.02	34
35	0.02	35			

*must be some kind of administrative or scoring error

WORD REASONING

New Order of Administration	MNO	Original order
1	0.98	1
2	0.98	4
3	0.98	6
4	0.89	3
5	0.82	2
6	0.75	5
7	0.81	7
8	0.90	8
9	0.81	10
10	0.72	9
11	0.44	12
12	0.20	15
13	0.14	11
14	0.08	13
15	0.09	14
16	0.18	16
17	0.16	18
18	0.10	19
19	0.09	17
20	0.08	20
21	0.03	21
22	0.02	23
23	0.02	24
24	0.00	22

Appendix J1

List of Added Items for Tryout-III

Vocabulary Subtest

حروف تہجی کسے کہتے ہیں؟

Information Subtest

پاکستان کے دو ہمسایہ ممالک کے نام بتائیں؟

Appendix J2

WISC-IV

Record Form (URDU)

For TRY OUT – III

India Standardization Edition

Examiner's Name:__________ Site ID Number:______ Region:__________

Examinee's Name:__________ Examinee's ID Number:__________

Teacher's Signature__________

	Year	Month	Day
Today's Date			
Date of Birth			
Age at testing			

Gender	Female	Male	
Handedness	Left	Right	Both
Colorblind?	No	Yes	

Does the examinee have any disabling conditions? No Yes*

If yes, please explain how the examinee meets inclusion criteria in the box below No Yes

Does the examinee wear prescription lenses, glasses, or a hearing aid? No Yes*

*If yes, was he/she wearing them during testing? No Yes

Start Time ___ : ___
Hr. Min.

1. Block Design (Time Limit: See Item)

Start
Ages 6–7: Item 1
Ages 8–16: Item 3

Reverse
Ages 8–16: Score of 0 or 1 on either of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 3 consecutive scores of 0.

Score
Items 1–3: Score 0, 1, or 2 points
Items 4–8: Score 0 or 4 points
Items 9–14: Score 0 or the appropriate time-bonus score
BDN
Items 1–3: Score 0, 1, or 2 points
Items 4–14: Score 0 or 4 points

Design	Presentation Method	Time Limit (Seconds)	Completion Time	Correct Design	Constructed Design	Score
7 ▶ 1. Child / Examiner	Model	30		Y N	Trial 1 Trial 2	0 1 2
2.	Model	45		Y N	Trial 1 Trial 2	0 1 2
16 ▶ 3.	Model and Picture	45		Y N	Trial 1 Trial 2	0 1 2
4.	Picture	45		Y N		0 4
5.	Picture	45		Y N		0 4
6.	Picture	75		Y N		0 4
7.	Picture	75		Y N		0 4
8.	Picture	75		Y N		0 4
9.	Picture	75		Y N		0 4 5 6 7
10.	Picture	75		Y N		0 4 5 6 7
11.	Picture	120		Y N		0 4 5 6 7
12.	Picture	120		Y N		0 4 5 6 7
13.	Picture	120		Y N		0 4 5 6 7
14.	Picture	120		Y N		0 4 5 6 7

Stop Time ___ : ___
Hr. Min.

Total Raw Score (Maximum = 68)

Block Design No Time Bonus (BDN) Total Raw Score (Maximum = 50)

2. Similarities

Start Time __ : __ Hr. Min.

Start Ages 6–8: Sample, then Item 1. Ages 9–11: Sample, then Item 3. Ages 12–16: Sample, then Item 5.

Reverse Ages 9–16: Score of 0 or 1 on either of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue After 5 consecutive scores of 0.

Score Items 1–2: Score 0 or 1 point. Items 3–23: Score 0, 1, or 2 points. See *Administration and Scoring Manual* for sample responses.

Start	Item	Response	Score
-16	سُرخ-نیلا		
-8	†1. دودھ-پانی		0 1
	†2. قلم-پنسل		0 1
-11	3. سیب-کیلا		0 1 2
	4. قمیض-جوتا		0 1 2
-16	5. تتلی-شہد کی مکھی		0 1 2
	6. بلی-چوہا		0 1 2
	7. سردیوں-گرمیوں		0 1 2
	8. کہنی-گھٹنا		0 1 2
	9. شہتیر-اینٹیں		0 1 2
	10. غصہ-خوشی		0 1 2
	11. ماتھے پر شکن پڑنا-مسکرانا		0 1 2
	12. برف-بھاپ		0 1 2
	13. پینٹنگ-مجسمہ		0 1 2
	14. سیلاب-خشک سالی		0 1 2
	15. شاعر-مصور		0 1 2
	16. اوّل-آخر		0 1 2
	17. ربڑ-کاغذ		0 1 2
	18. پہاڑ-جھیل		0 1 2

† If the child does not give a 1-point response, provide the response indicated in the *Administration and Scoring Manual*.

Continue

2. Similarities *(Continued)*

Discontinue after 5 consecutive scores of 0

Item	Response	Score
19. نمک-پانی		0 1 2
20. اجازت-بندش		0 1 2
21. انتقام-معافی		0 1 2
22. حقیقت-خواب		0 1 2
23. مکان-زمان		0 1 2

Stop Time ___:___ Hr. Min.

Total Raw Score (Maximum = 44)

Start Time ___:___ Hr. Min.

3. Digit Span

Start
Age 6–16:
Forwards: Item 1
Backwards: Sample, then Item 1

Discontinue
Forwards: After scores of 0 on both trials of an item
Backwards: After scores of 0 on both trials of an item

Score
Score 0 or 1 point for each trial
DSF & DSB
Total Raw Score for DS Forwards and Backwards, respectively
LDSF & LDSB
Number of digits recalled on last trial scored 1 point for DS Forwards and Backwards, respectively

Forwards

Trial		Response	Trial Score	Item Score
6–16 ▶ 1.	2 – 9		0 1	0 1 2
	4 – 6		0 1	
2.	3 – 8 – 6		0 1	0 1 2
	6 – 1 – 2		0 1	
3.	3 – 4 – 1 – 7		0 1	0 1 2
	6 – 1 – 5 – 8		0 1	
4.	8 – 4 – 2 – 3 – 9		0 1	0 1 2
	5 – 2 – 1 – 8 – 6		0 1	
5.	3 – 8 – 9 – 1 – 7 – 4		0 1	0 1 2
	7 – 9 – 6 – 4 – 8 – 3		0 1	
6.	5 – 1 – 7 – 4 – 2 – 3 – 8		0 1	0 1 2
	9 – 8 – 5 – 2 – 1 – 6 – 3		0 1	
7.	1 – 8 – 4 – 5 – 9 – 7 – 6 – 3		0 1	0 1 2
	2 – 9 – 7 – 6 – 3 – 1 – 5 – 4		0 1	
8.	5 – 3 – 8 – 7 – 1 – 2 – 4 – 6 – 9		0 1	0 1 2
	4 – 2 – 6 – 9 – 1 – 7 – 8 – 3 – 5		0 1	

Backwards

Trial		Response	Trial Score	Item Score
6–16 ▶	8 – 2			
	5 – 6			
1.	2 – 1		0 1	0 1 2
	1 – 3		0 1	
2.	3 – 5		0 1	0 1 2
	6 – 4		0 1	
3.	5 – 7 – 4		0 1	0 1 2
	2 – 5 – 9		0 1	
4.	7 – 2 – 9 – 6		0 1	0 1 2
	8 – 4 – 9 – 3		0 1	
5.	4 – 1 – 3 – 5 – 7		0 1	0 1 2
	9 – 7 – 8 – 5 – 2		0 1	
6.	1 – 6 – 5 – 2 – 9 – 8		0 1	0 1 2
	3 – 6 – 7 – 1 – 9 – 4		0 1	
7.	8 – 5 – 9 – 2 – 3 – 4 – 6		0 1	0 1 2
	4 – 5 – 7 – 9 – 2 – 8 – 1		0 1	
8.	6 – 9 – 1 – 7 – 3 – 2 – 5 – 8		0 1	0 1 2
	3 – 1 – 7 – 9 – 5 – 4 – 8 – 2		0 1	

LDSF Max = 9

Digit Span Forwards (DSF) Total Raw Score (Maximum = 16)

LDSB Max = 8

Digit Span Backwards (DSB) Total Raw Score (Maximum = 16)

Stop Time ___:___ Hr. Min.

Total Raw Score (Maximum = 32)

Start Time___:___
Hr. Min.

4. Picture Concepts

Start
Ages 6–8: Sample A & B, then Item 1
Ages 9–11: Sample A & B, then Item 5
Ages 12–16: Sample A & B, then Item 7

Reverse
Ages 9–16: Score of 0 on either of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 5 consecutive scores of 0.

Score
Score 0 or 1 point. Correct responses are in color.

Item	Response	Score
6–16 A.	1 2 | 3 4 DK	
B.	1 2 | 3 4 DK	
6–8 1.	1 2 | 3 4 DK	0 1
2.	1 2 | 3 4 DK	0 1
3.	1 2 | 3 4 DK	0 1
4.	1 2 | 3 4 DK	0 1
9–11 5.	1 2 | 3 4 DK	0 1
6.	1 2 3 | 4 5 6 DK	0 1
12–16 7.	1 2 3 | 4 5 6 DK	0 1
8.	1 2 3 | 4 5 6 DK	0 1
9.	1 2 3 | 4 5 6 DK	0 1
10.	1 2 3 | 4 5 6 DK	0 1
11.	1 2 3 | 4 5 6 DK	0 1

Item	Response	Score
12.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
13.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
14.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
15.	1 2 3 | 4 5 6 DK	0 1
16.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
17.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
18.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
19.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
20.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
21.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
22.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
23.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
24.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
25.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
26.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
27.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
28.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1

Stop Time___:___
Hr. Min.

Total Raw Score
(Maximum = 28)

5. Coding

(Time Limit: 120 seconds)

Start Time___:___
Hr. Min.

Start
Ages 6–7: Coding A Sample Items, then Test Items
Ages 8–16: Coding B Sample Items, then Test Items

Discontinue
After 120 seconds

Score
Use the scoring key to check the child's responses.
Score 1 point for each correct response.

Form	Time Limit (Seconds)	Completion Time	Total Raw Score
6–7 A.	120		(Max = 65)
8–16 B.	120		(Max = 119)

Coding A Time Bonus Scores for Perfect Performance							
Time in Seconds	116–120	111–115	106–110	101–105	96–100	86–95	≤85
Score	59	60	61	62	63	64	65

Stop Time___:___
Hr. Min.

6. Vocabulary

Start Time ___:___
Hr. Min.

Start
Ages 6–8: Item 5
Ages 9–11: Item 7
Ages 12–16: Item 9

Reverse
Ages 6–16: Score of 0 or 1 on either of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 5 consecutive scores of 0

Score
Items 1–4: Score 0 or 1 point.
Items 5–[illegible]: Score 0, 1, or 2 points.
[illegible] Administration and Scoring Manual [illegible] sample responses.

Item	Response	Score
Picture Items		
1. کار		0 1
2. پھول		0 1
3. ریل گاڑی		0 1
4. بالٹی		0 1
Verbal Items		
8 → †5. ٹوپی		0 1 2
†6. چھتری		0 1 2
11 → 7. گھڑیال		0 1 2
8. گائے		0 1 2
16 → 9. سائیکل		0 1 2
10. چور		0 1 2
11. بہادر		0 1 2
*12. رخصت		0 1 2
13. جذب کرنا		0 1 2
*14. حروف / حروف تہجی		0 1 2
15. ہجرت		0 1 2
16. فرماں برداری		0 1 2
17. دفاع		0 1 2

† If the child does not give a 2-point response, provide the response indicated in the *Administration and Scoring Manual*.
* Responses requiring specific query are identified in the *Administration and Scoring Manual*.

Continue

5. Vocabulary *(Continued)*

Discontinue after 5 consecutive scores of 0

Item	Response	Score
18. قدیم		0 1 2
19. جزیرہ		0 1 2
20. احمقانہ		0 1 2
*21. نقال		0 1 2
22. حکایت		0 1 2
23. مُختصر		0 1 2
24. جبر کرنا		0 1 2
25. وقت طلب		0 1 2
26. شاذ و نادر		0 1 2
27. ایذا		0 1 2
28. چھال		0 1 2
29. ترمیم		0 1 2
*30. دربار		0 1 2
31. جامع		0 1 2
32. دُوراندیش		0 1 2
*33. بڑبولا		0 1 2
34. انحراف		0 1 2
35. ناگزیر		0 1 2
36. مسابقت / رقابت		0 1 2

* Responses requiring specific query are identified in the *Administration and Scoring Manual.*

Total Raw Score (Maximum = 68)

Stop Time___:___ Hr. Min.

7. Letter–Number Sequencing

Start Time___:___ Hr. Min.

Start
Ages 6–7: Qualifying Items, Sample Item, then Item [illegible]
Ages 8–16: Sample Item, then Item [illegible]

Discontinue
If child is unable to respond correctly to either Qualifying Item or after scores of 0 on [illegible]

Score
Score 0 or 1 point for each trial

Qualifying Item	Correct Response	Correct
6–7 ▶ Counting	Child counts to three.	Y N
Alphabet	Child recites alphabet to the letter C.	Y N

Item		Trial	Correct Responses		Verbatim Response	Trial Score	Item Score
8–16 ▶ S	1.	A – 2	2 – A	A – 2			
	2.	B – 3	3 – B	B – 3			
1.	1.	A – 3	3 – A	A – 3		0 1	0 1 2 3
		If the child responds A–3, immediately correct the child as instructed in the manual.					
	2.	B – 1	1 – B	B – 1		0 1	
	3.	2 – C	2 – C	C – 2		0 1	
2.	1.	C – 4	4 – C	C – 4		0 1	0 1 2 3
	2.	5 – E	5 – E	E – 5		0 1	
	3.	D – 3	3 – D	D – 3		0 1	
3.	1.	B – 1 – 2	1 – 2 – B	B – 1 – 2		0 1	0 1 2 3
	2.	1 – 3 – C	1 – 3 – C	C – 1 – 3		0 1	
	3.	2 – A – 3	2 – 3 – A	A – 2 – 3		0 1	
4.	1.	D – 2 – 9	2 – 9 – D	D – 2 – 9		0 1	0 1 2 3
	2.	R – 5 – B	5 – B – R	B – R – 5		0 1	
		If the child responds 5–R–B or R–B–5, say, Remember to say the letters in order.					
	3.	H – 9 – K	9 – H – K	H – K – 9		0 1	
5.	1.	3 – E – 2	2 – 3 – E	E – 2 – 3		0 1	0 1 2 3
		If the child responds 3–2–E or E–3–2, say, Remember to say the numbers in order.					
	2.	9 – J – 4	4 – 9 – J	J – 4 – 9		0 1	
	3.	B – 5 – F	5 – B – F	B – F – 5		0 1	
6.	1.	1 – C – 3 – J	1 – 3 – C – J	C – J – 1 – 3		0 1	0 1 2 3
	2.	5 – A – 2 – B	2 – 5 – A – B	A – B – 2 – 5		0 1	
	3.	D – 8 – M – 1	1 – 8 – D – M	D – M – 1 – 8		0 1	
7.	1.	1 – B – 3 – G – 7	1 – 3 – 7 – B – G	B – G – 1 – 3 – 7		0 1	0 1 2 3
	2.	9 – V – 1 – T – 7	1 – 7 – 9 – T – V	T – V – 1 – 7 – 9		0 1	
	3.	P – 3 – J – 1 – M	1 – 3 – J – M – P	J – M – P – 1 – 3		0 1	
8.	1.	1 – D – 4 – E – 9 – G	1 – 4 – 9 – D – E – G	D – E – G – 1 – 4 – 9		0 1	0 1 2 3
	2.	H – 3 – B – 4 – F – 8	3 – 4 – 8 – B – F – H	B – F – H – 3 – 4 – 8		0 1	
	3.	7 – Q – 6 – M – 3 – Z	3 – 6 – 7 – M – Q – Z	M – Q – Z – 3 – 6 – 7		0 1	
9.	1.	S – 3 – K – 4 – Y – 1 – G	1 – 3 – 4 – G – K – S – Y	G – K – S – Y – 1 – 3 – 4		0 1	0 1 2 3
	2.	7 – S – 9 – K – 1 – T – 6	1 – 6 – 7 – 9 – K – S – T	K – S – T – 1 – 6 – 7 – 9		0 1	
	3.	L – 2 – J – 6 – Q – 3 – G	2 – 3 – 6 – G – J – L – Q	G – J – L – Q – 2 – 3 – 6		0 1	
10.	1.	4 – B – 8 – R – 1 – M – 7 – H	1 – 4 – 7 – 8 – B – H – M – R	B – H – M – R – 1 – 4 – 7 – 8		0 1	0 1 2 3
	2.	J – 2 – U – 8 – A – 5 – C – 4	2 – 4 – 5 – 8 – A – C – J – U	A – C – J – U – 2 – 4 – 5 – 8		0 1	
	3.	6 – L – 1 – Z – 5 – H – 2 – W	1 – 2 – 5 – 6 – H – L – W – Z	H – L – W – Z – 1 – 2 – 5 – 6		0 1	

Stop Time___:___ Hr. Min.

Total Raw Score (Maximum = 30)

8. Matrix Reasoning

Start Time ___ : ___
Hr. Min.

Start
Ages 6–8: Samples A–C, then Item 4
Ages 9–11: Samples A–C, then Item 7
Ages 12–16: Samples A–C, then Item 11

Reverse
Ages 6–16: Score of 0 on either of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 4 consecutive scores of 0 or 4 scores of 0 on five consecutive items

Item	Response						Score
	1	2	3	4	5	DK	
	1	2	3	4	5	DK	
	1	2	3	4	5	DK	
1.	1	2	3	4	5	DK	0 1
2.	1	2	3	4	5	DK	0 1
3.	1	2	3	4	5	DK	0 1
4.	1	2	3	4	5	DK	0 1
5.	1	2	3	4	5	DK	0 1
6.	1	2	3	4	5	DK	0 1
7.	1	2	3	4	5	DK	0 1
8.	1	2	3	4	5	DK	0 1
9.	1	2	3	4	5	DK	0 1
10.	1	2	3	4	5	DK	0 1
11.	1	2	3	4	5	DK	0 1

Item	Response						Score
12.	1	2	3	4	5	DK	0 1
13.	1	2	3	4	5	DK	0 1
14.	1	2	3	4	5	DK	0 1
15.	1	2	3	4	5	DK	0 1
16.	1	2	3	4	5	DK	0 1
17.	1	2	3	4	5	DK	0 1
18.	1	2	3	4	5	DK	0 1
19.	1	2	3	4	5	DK	0 1
20.	1	2	3	4	5	DK	0 1
21.	1	2	3	4	5	DK	0 1
22.	1	2	3	4	5	DK	0 1
23.	1	2	3	4	5	DK	0 1
24.	1	2	3	4	5	DK	0 1
25.	1	2	3	4	5	DK	0 1

Item	Response						Score
26.	1	2	3	4	5	DK	0 1
27.	1	2	3	4	5	DK	0 1
28.	1	2	3	4	5	DK	0 1
29.	1	2	3	4	5	DK	0 1
30.	1	2	3	4	5	DK	0 1
31.	1	2	3	4	5	DK	0 1
32.	1	2	3	4	5	DK	0 1
33.	1	2	3	4	5	DK	0 1
34.	1	2	3	4	5	DK	0 1
35.	1	2	3	4	5	DK	0 1

Total Raw Score
(Maximum = 35)

Stop Time ___ : ___
Hr. Min.

9. Comprehension

Start Time ___ : ___
Hr. Min.

Start

Reverse

Discontinue

Score

Item	Response	Score
†1. دانت		0 1 2
2. سبزیاں		0 1 2
3. لڑائی		0 1 2
4. سیٹ بیلٹ		0 1 2
5. ورزش		0 1 2
6. پولیس		0 1 2

† If the child does not give a 2-point response, provide the response indicated in the *Administration and Scoring Manual*.
* If the child replies with only one general idea, ask for a second response as indicated in the *Administration and Scoring Manual*.

Continue

Comprehension *(Continued)*

Discontinue after 4 consecutive scores of 0

Item	Response	Score
7. لائسنس		0 1 2
8. بٹوے		0 1 2
9. معائنہ		0 1 2
*10. دھواں		0 1 2
11. معذرت		0 1 2
*12. وعدہ		0 1 2
*13. لائبریریاں		0 1 2
*14. ڈاکٹر		0 1 2
15. اخبار		0 1 2
16. ملکیت		0 1 2
17. جلدحق		0 1 2
*18. جمہوریت		0 1 2
*19. ٹیکنالوجی		0 1 2
20. ٹکٹ		0 1 2
*21. ابلاغ		0 1 2

If the child replies with one general idea, ask for a second response as indicated in the *Administration and Scoring Manual*.

Total Raw Score (Maximum = 42)

Stop Time___:___
Hr. Min.

10. Symbol Search

(Time Limit: 120 seconds)

Start Time ___:___ Hr. Min.

Start Ages 6–7: Symbol Search A Sample Items, Practice Items, then Test Items
Ages 8–16: Symbol Search B Sample Items, Practice Items, then Test Items

Discontinue After 120 seconds

Score

Completion Time	Stop Time ___:___ Hr. Min.	Number Correct	–	Number Incorrect	=	Total Raw Score (Ages 6–7; Max = 45) (Ages 8–16; Max = 60)

11. Picture Completion

(Time Limit: 20 seconds)

Start Time ___:___ Hr. Min.

Start Ages 6–8: Sample, then Item 1
Ages 9–11: Sample, then Item 5
Ages 12–16: Sample, then Item 10

Reverse

Discontinue

Score

Item	Response	Score
6–16 Sample پنسل		
6–8 †1. جیکٹ		0 1
†2. لومڑی		0 1
3. بلی		0 1
4. ہاتھ		0 1
9–11 5. گھنٹی		0 1
6. رسی کودنا		0 1
7. آری		0 1
8. الماری کا دروازہ		0 1
9. پتا		0 1
12–16 10. آئینہ		0 1
11. سیڑھی		0 1
12. قینچیاں		0 1
13. گھڑیال		0 1
14. دروازہ		0 1
15. بلب		0 1
16. چہرہ		0 1
17. بیلٹ		0 1
18. بینڈ		0 1
19. ڈائس		0 1
20. ہانیک		0 1
21. درخت		0 1
22. فٹ بال		0 1
23. سیٹی		0 1
*24. بط		0 1
25. سور		0 1
26. پل		0 1
27. ہائیکل		0 1
28. چھتری		0 1
29. قفل		0 1
30. سنہری مچھلی		0 1
31. ویلیس		0 1
32. تھرمامیٹر		0 1
33. ہاتھ		0 1
34. سپر مارکیٹ		0 1
35. پانی		0 1
36. خاندان		0 1
37. گھر		0 1
38. تاج		0 1

Total Raw Score (Maximum = 38)

Stop Time ___:___ Hr. Min.

† If the child does not give a 1-point response, provide the response indicated in the *Administration and Scoring Manual.*
* Responses requiring specific query are identified in the *Administration and Scoring Manual.*

12. Cancellation

(Time Limit: 45 seconds)

Start Time ___:___ Hr. Min.

Start Ages 6–16: Sample, Practice, then Item 1

Discontinue After 45 seconds for each item

Score

Item	Time Limit (Seconds)	Completion Time	Number Correct	Number Incorrect	Difference	Bonus Points	Total Raw Score
6–16 1. Random	45						CAR Max = 68
2. Structured	45						CAS Max = 68

Time Bonus Points If the child completes an item before 45 seconds and the difference is ≥ 60, award bonus points					
Time in Seconds	45	40–44	35–39	30–34	0–29
Bonus Points	0	1	2	3	4

Total Raw Score (Maximum = 136)

Stop Time ___:___ Hr. Min.

13. Information

Start Time ___ : ___
Hr. Min.

Start
Ages 6–8: Item 5
Ages 9–11: Item 10
Ages 12–16: Item 12

Reverse
Ages 9–16: Score of 0 on either of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 5 consecutive scores of 0

Score
Score 0 or 1 point.
See Administration and Scoring Manual for sample responses.

Item	Response	Score
†1. پاؤں		0 1
†2. ناک		0 1
3. کھانے کی چیز		0 1
*4. کان		0 1
6–8 ▶ 5. عمر		0 1
6. بچے		0 1
7. مارچ		0 1
8. ٹانگیں		0 1
*9. جمعرات		0 1
9–11 ▶ 10. ابالنا		0 1
*11. موسم		0 1
12–16 ▶ *12. بخار		0 1
13. ہمسایہ / دوزجن		0 1
*14. معدہ (stomach)		0 1
15. مہینے		0 1
*16. سال		0 1
17. فوسل		0 1

Item	Response	Score
18. آکسیجن		0 1
19. بیج		0 1
*20. آبادی		0 1
21. خطاطی (کیلیگرافی)		0 1
22. کولمبس		0 1
23. لوزون		0 1
24. بیرومیٹر		0 1
*25. زنگ		0 1
26. چھوٹا دن		0 1
27. فیشن		0 1
28. ہیرے		0 1
29. ملائشیا		0 1
30. اردن!		0 1
31. سقراط (socrates)		0 1
*32. کراچی		0 1
*33. تارکین		0 1

† If the child does not give a 1-point response, provide the response indicated in the *Administration and Scoring Manual*.

* Responses requiring specific query are identified in the *Administration and Scoring Manua*

Stop Time ___ : ___
Hr. Min.

Total Raw Score
(Maximum = 33)

14. Arithmetic (Time Limit: 30 seconds)

Start Time ___ : ___
Hr. Min.

Start
Ages 6–7: Item 3
Ages 8–9: Item 9
Ages 10–16: Item 12

Reverse
Score of 0 on either of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 4 consecutive scores of 0

Score
Score 0 or 1 point.

Item	Correct Response	Response	Score
†1. ریچھ	1, 2, 3		0 1
†2. چوزے	1, 2, 3, 4, 5		0 1
6–7 ▶ †3. درخت	1, 2,...10		0 1
4. تتلیاں	9		0 1
5. سیب	2		0 1
6. رنگ دار پنسل	5		0 1
7. کتابیں	4		0 1
8. بسکٹ	3		0 1
8–9 ▶ 9. ٹکڑے	2		0 1
10. روپے	6		0 1
11. پنسل	6		0 1
10–16 ▶ 12. چاکلیٹ	7		0 1

Item	Correct Response	Response	Score
13. کاریں	16		0 1
14. قلم	14		0 1
15. کیک	25		0 1
16. غبارے	7		0 1
17. ماکس	5		0 1
18. سیب	9		0 1
19. ٹکٹ	6		0 1
20. پوائنٹس	32		0 1
21. طالب علموں	19		0 1
22. رسالے	3		0 1
23. انعامات	24		0 1
24. سوالات	20 seconds		0 1

Item	Correct Response	Response	Score
25. کلاس روم	20		0 1
26. جالیا	7		0 1
27. دام	Rs.8.50		0 1
28. کتاب	30		0 1
29. گاڑی چلانے	60		0 1
30. درجہ حرارت	3		0 1
31. کھلونا	34		0 1
32. کاریگر	48		0 1
33. پرواز	2:00		0 1
34. کام	40		0 1

Total Raw Score
(Maximum = 34)

† If the child does not give a 1-point response, provide the response indicated in the *Administration and Scoring Manual*.

Stop Time ___ : ___
Hr. Min.

15. Word Reasoning

Start Time ___ : ___
Hr. Min.

Start Ages 6-9: Samples A & B, then Item 1. Ages 10-16: Samples A & B, then Item [illegible]

Reverse Ages 10-16: Score of 0 on first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue After 5 consecutive scores of 0.

Score Score 0 or 1 point. See Administration and Scoring Manual for sample responses.

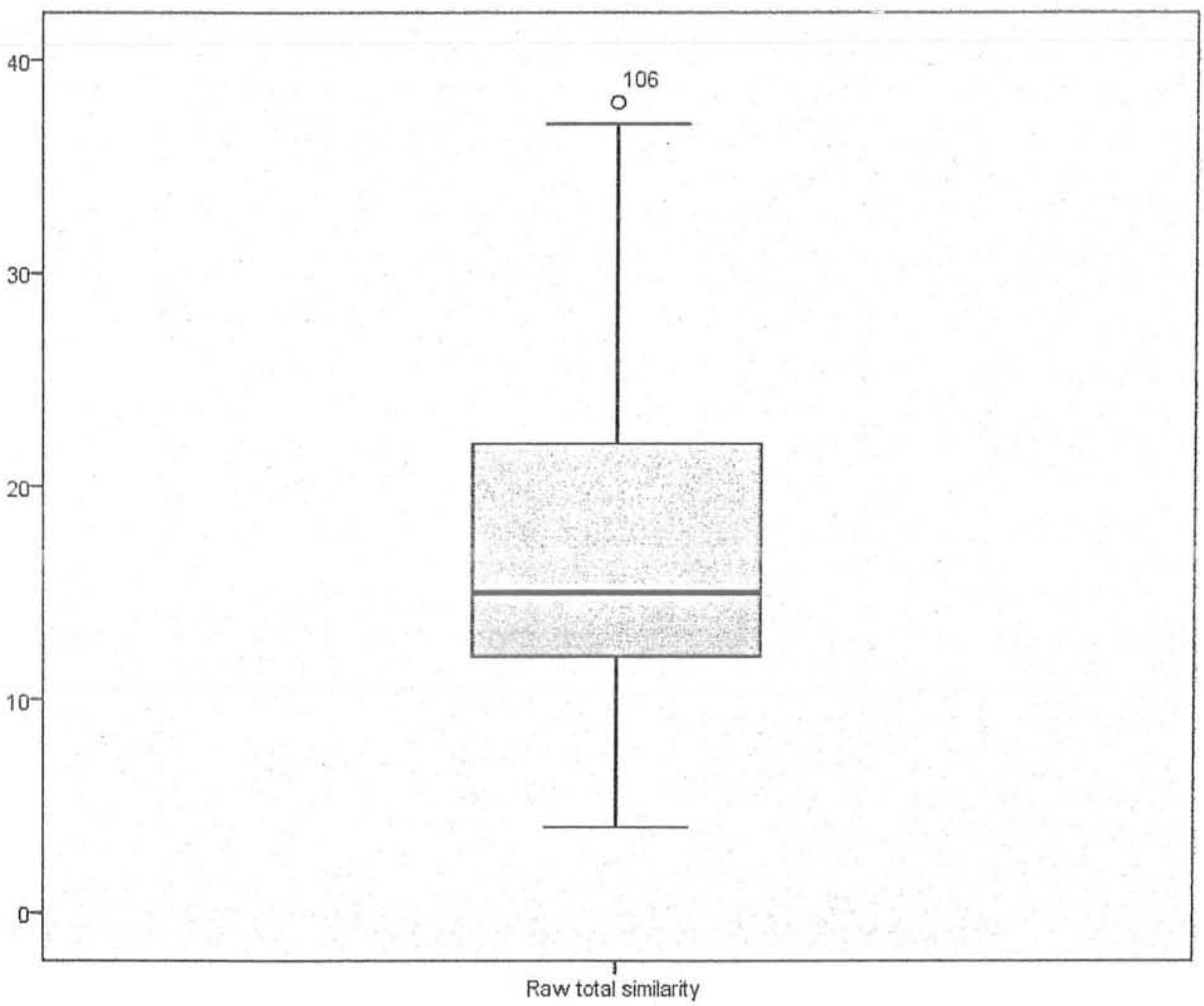
Score	Correct	Response	Clue	
	Y N		یہ ایک جانور ہے جو "بھوں ووں" کرتا ہے۔	A.
	Y N		اس کا ایک لمبا ہینڈل ہوتا ہے	B.
	Y N		اور اسے فرش صاف کرنے کے لئے استعمال کیا جاتا ہے	
0 1	Y N		اسے نہانے کے بعد خود کو خشک کرنے کے لئے استعمال کیا جاتا ہے۔	1.
0 1	Y N		یہ آپ کے سر کا حصہ ہے جو سونگھنے کے لئے استعمال کیا جاتا ہے۔	2.
0 1	Y N		یہ ایسی چیز ہے جو ہم سر ڈھاپنے کے لئے پہنتے ہیں۔	3.
0 1	Y N		ایسا جانور جس کی لمبی سونڈ اور بڑے بڑے کان ہوتے ہیں۔	4.
0 1	Y N		ایسا کمرہ جہاں لوگ سوتے ہیں۔	5.
0 1	Y N		اس کا ایک ہینڈل ہوتا ہے اور لوگ اسے کھول سکتے ہیں اور اس میں سے گزر سکتے ہیں۔	6.
0 1	Y N		یہ اوپر آسمان پر ہوتے ہیں۔	7.
	Y N		اور آپ انہیں صرف رات میں دیکھ سکتے ہیں۔	
0 1	Y N		یہ بہت سے رنگوں میں آتا ہے۔	8.
	Y N		اور یہ دیواروں پر استعمال ہوتا ہے۔	
0 1	Y N		یہ بارش کے بعد زمین پر ہوتی ہے۔	9.
	Y N		اور یہ آپ کے کپڑوں کو گندا کر سکتی ہے۔	
0 1	Y N		یہ ایک سیکھنے کی جگہ ہے۔	10.
	Y N		اور اس میں ماضی کی بہت سی چیزیں ہو سکتی ہیں۔	
0 1	Y N		یہ کھانے کا ذائقہ بہتر بنا سکتا ہے۔	11.
	Y N		اور یہ سمندر سے مل سکتا ہے۔	
0 1	Y N		یہ قدرتی طور پر پایا جاتا ہے۔	12.
	Y N		اور اس کے منہ سے پانی بہتا ہے۔	
0 1	Y N		یہ ایسا کام ہے جو لوگ چیزوں کو "نیا جیسا" کرنے کے لئے کرتے ہیں۔	13.
	Y N		اور لوگ اسے ناکارہ چیزوں کے ساتھ کرتے ہیں۔	
0 1	Y N		ان کی حفاظت جلد اور ہڈیوں سے کی گئی ہے۔	14.
	Y N		اور بعض اوقات انہیں تبدیل بھی کیا جا سکتا ہے۔	
0 1	Y N		اسے توڑا جا سکتا ہے۔	15.
	Y N		اور اسے لوگوں کی حفاظت کے لئے لکھا گیا ہے۔	

	Y N			یہ نئی دریافتوں کی طرف رہنمائی کرتا ہے۔	
0 1	Y N			اور یہ ایک مرحلہ وار عمل ہے۔	16.
	Y N			اور اس میں تجربے بھی ہو سکتے ہیں۔	
	Y N			یہ ایسی چیز ہے جو ہر کسی کی ہوتی ہے۔	
0 1	Y N			اور ہر سال بدلتی ہے۔	17.
	Y N			اور یہ کبھی کم نہیں ہو سکتی۔	
	Y N			اسے چھوا نہیں جا سکتا۔	
0 1	Y N			اور یہ پلٹ نہیں سکتا۔	18.
	Y N			اور یہ رُک نہیں سکتا۔	
	Y N			اس سے لوگوں کو کچھ کرنے یا رکھنے کی اجازت ملتی ہے۔	
0 1	Y N			اور یہ حکومت کی طرف سے دیا جاتا روی جاتی ہے۔	19.
	Y N			اور ہو سکتا ہے اسے حاصل کرنے کے لئے لوگوں کو ٹیسٹ دینا پڑے۔	
	Y N			یہ ایسی چیز ہے جو زیادہ تر لوگ چاہتے ہیں۔	
0 1	Y N			اور تنازعے سے یہ ختم ہو جاتا ہے۔	20.
	Y N			اور حکومتیں اسے قائم رکھنے کی کوشش کرتی ہیں۔	
	Y N			یہ ایک دریا ہو سکتا ہے۔	
0 1	Y N			اور جنگیں اسے تبدیل کر سکتی ہیں۔	21.
	Y N			اور دو ملک اسے مل کر استعمال کر سکتے ہیں۔	
	Y N			یہ ایک جگہ ہے۔	
0 1	Y N			اور یہ موسم سے محفوظ ہے۔	22.
	Y N			اور یہ کسی اور چیز کے اندر پایا جاتا ہے۔	
	Y N			اسے پہلے کبھی دیکھا یا کیا نہیں گیا۔	
0 1	Y N			اس سے ہماری زندگی بہتر یا آسان ہو سکتی ہے۔	23.
	Y N			اور یہ ذہن کی پیداوار ہے۔	
	Y N			یہ لوگوں سے بنتا رہتی ہے۔	
0 1	Y N			اور اسے چھوا نہیں جا سکتا۔	24.
	Y N			اور یہ پہلے ہی وقوع پزیر ہو چکا ہوتا رہو چکی ہوتی ہے۔	

Notes / Behavioral Observations:

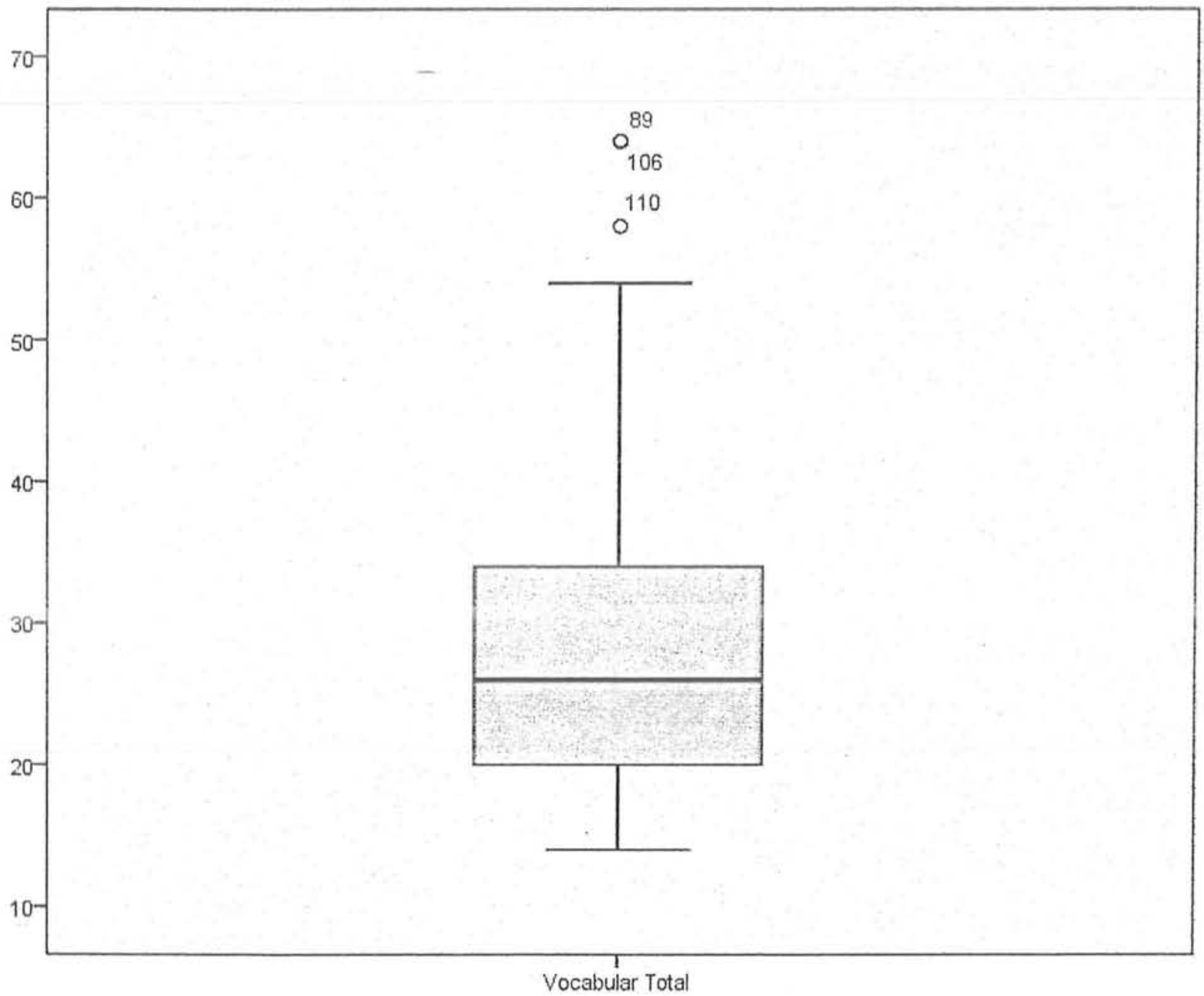
Appendix K1

Similarities- Box Plot



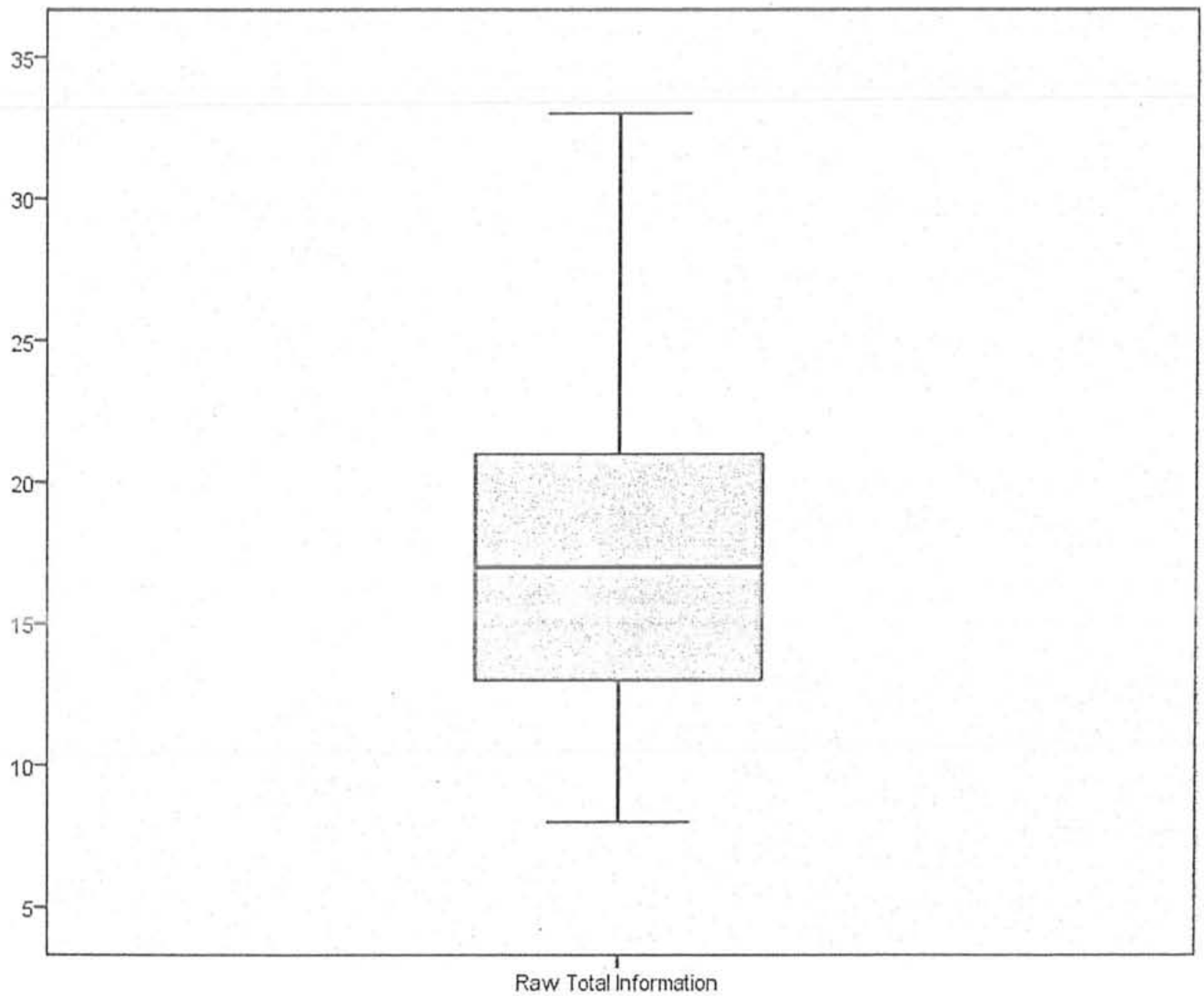
Appendix K2

Vocabulary - Box Plot



Appendix K3

Information - Box Plot



Appendix L

WISC-IV Record Form (Urdu)

Urdu Standardization Edition

Examiner's Name:________________ Site ID Number:______ Region:__________

Examinee's Name:____________________ Examinee's ID Number:__________

Teacher's Signature____________________

	Year	Month	Day
Today's Date			
Date of Birth			
Age at testing			

Gender	Female	Male	
Handedness	Left	Right	Both
Colorblind?	No	Yes	

Does the examinee have any disabling conditions? No Yes*

If yes, please explain how the examinee meets inclusion criteria in the box below No Yes

Does the examinee wear prescription lenses, glasses, or a hearing aid? No Yes*

*If yes, was he/she wearing them during testing? No Yes

2. Similarities

Start Time ___ : ___
Hr. Min.

Start
Ages 6–8: Sample, then Item 1
Ages 9–11: Sample, then Item 3
Ages 12–16: Sample, then Item 5

Reverse
Ages 9–16: Score of 0 or 1 on *either* of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 5 consecutive scores of 0.

Score
Items 1–2: Score 0 or 1 point
Items 3–23: Score 0, 1, or 2 points
See *Administration and Scoring Manual* for sample responses.

Item	Response	Score
–16	سُرخ۔نیلا	
6–8 †1.	دودھ۔پانی	0 1
†2.	قلم۔پینسل	0 1
–11 3.	سیب۔کیلا	0 1 2
4.	قمیض۔جوتا	0 1 2
2–16 5.	تتلی۔شہد کی مکھی	0 1 2
6.	بلی۔چوہا	0 1 2
7.	کہنی۔گھٹنا	0 1 2
8.	سردیوں۔گرمیوں	0 1 2
9.	شہتیر۔اینٹیں	0 1 2
10.	غصہ۔خوشی	0 1 2
11.	برف۔بھاپ	0 1 2
12.	ماتھے پر شکن پڑنا۔مسکرانا	0 1 2
13.	سیلاب۔خشک سالی	0 1 2
14.	پینٹنگ۔مجسمہ	0 1 2
15.	اول۔آخر	0 1 2
16.	شاعر۔مصور	0 1 2
17.	نمک۔پانی	0 1 2
18.	اجازت۔بندش	0 1 2

† If the child does not give a 1-point response, provide the response indicated in the *Administration and Scoring Manual.*

Continue

4. Picture Concepts

Start Time___:___
Hr. Min.

Start
Ages 6–8: Samples A& B, then Item 1
Ages 9–11: Samples A& B, then Item 5
Ages 12–16: Samples A& B, thenItem 7

Reverse
Ages 9–16: Score of 0 on *either* of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 5 consecutive scores of 0

Score
Score 0 or 1 point
Correct responses are in colour.

	Item	Response							Score
–16		1	2 |	3	4			DK	
		1	2 |	3	4			DK	
6–8	1.	1	2 |	3	4			DK	0 1
	2.	1	2 |	3	4			DK	0 1
	3.	1	2 |	3	4			DK	0 1
	4.	1	2 |	3	4			DK	0 1
9–11	5.	1	2 |	3	4			DK	0 1
	6.	1	2	3 |	4	5	6	DK	0 1
2–16	7.	1	2	3 |	4	5	6	DK	0 1
	8.	1	2	3 |	4	5	6	DK	0 1
	9.	1	2	3 |	4	5	6	DK	0 1
	10.	1	2	3 |	4	5	6	DK	0 1
	11.	1	2	3 |	4	5	6	DK	0 1

Item	Response	Score
12.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
13.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
14.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
15.	1 2 3 | 4 5 6 DK	0 1
16.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
17.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
18.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
19.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
20.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
21.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
22.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
23.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
24.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
25.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
26.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1
27.	1 2 3 | 4 5 6 | 7 8 9 DK	0 1
28.	1 2 3 4 | 5 6 7 8 | 9 10 11 12 DK	0 1

Stop Time___:___
Hr. Min.

Total Raw Score
(Maximum = 28)

5. Coding

(Time Limit: 120 seconds)

Start Time___:___
Hr. Min.

Start
Ages 6–7: Coding A Sample Items, then Test Items
Ages 8–16: Coding B Sample Items, then Test Items

Discontinue
After 120 seconds

Score
Use the Scoring Key to check the child's responses.
Score 1 point for each correct response.

	Form	Time Limit (Seconds)	Completion Time	Total Raw Score
6–7	A.	120		Max = 65
8–16	B.	120		Max = 119

Coding A Time Bonus Scores for Perfect Performance

Time in Seconds	116–120	111–115	106–110	101–105	96–100	86–95	≤85
Score	59	60	61	62	63	64	65

Stop Time___:___
Hr. Min.

6. Vocabulary *(Continued)*

Discontinue after 5 consecutive scores of 0

Item	Response	Score
18. قدیم		0 1 2
19. جزیرہ		0 1 2
20. احمقانہ		0 1 2
21. دقت طلب		0 1 2
22. جبر کرنا		0 1 2
23. نقالی		0 1 2
24. مُخففہ		0 1 2
25. حکایت		0 1 2
26. ایذا		0 1 2
27. شاذونادر		0 1 2
28. چھال		0 1 2
29. دوراندیش		0 1 2
30. دیرباز		0 1 2
31. جامع		0 1 2
32. بڑبولا		0 1 2
33. ترمیم		0 1 2
34. انحراف		0 1 2
35. ناگزیر		0 1 2
36. رقابت		0 1 2

* Responses requiring specific query are identified in the *Administration and Scoring Manual*.

Total Raw Score (Maximum = 68) ____

Stop Time ___:___ Hr. Min.

8. Matrix Reasoning

Start Time ___:___
Hr. Min.

Start
Ages 6–8: Samples A–C, then Item 4
Ages 9–11: Samples A–C, then Item 7
Ages 12–16: Samples A–C, then Item 11

Reverse
Ages 6–16: Score of 0 on *either* of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 4 consecutive scores of 0 *or* 4 scores of 0 on *five* consecutive items

Score
Score 0 or 1 point
Correct responses are in colour.

	Item	Response						Score	
6–16		**1**	2	3	4	5	DK		
		1	2	3	4	**5**	DK		
		1	2	3	**4**	5	DK		
	1.	1	**2**	3	4	5	DK	0	1
	2.	1	2	**3**	4	5	DK	0	1
	3.	1	2	3	**4**	5	DK	0	1
6–8	4.	**1**	2	3	4	5	DK	0	1
	5.	1	2	3	4	**5**	DK	0	1
	6.	1	**2**	3	4	5	DK	0	1
9–11	7.	1	2	3	**4**	5	DK	0	1
	8.	1	2	3	**4**	5	DK	0	1
	9.	**1**	2	3	4	5	DK	0	1
	10.	1	2	**3**	4	5	DK	0	1
12–16	11.	1	**2**	3	4	5	DK	0	1

Item	Response						Score	
12.	1	2	3	4	**5**	DK	0	1
13.	**1**	2	3	4	5	DK	0	1
14.	1	2	3	**4**	5	DK	0	1
15.	**1**	2	3	4	5	DK	0	1
16.	1	**2**	3	4	5	DK	0	1
17.	1	**2**	3	4	5	DK	0	1
18.	1	2	3	**4**	5	DK	0	1
19.	1	2	3	**4**	5	DK	0	1
20.	1	2	3	4	**5**	DK	0	1
21.	1	2	**3**	4	5	DK	0	1
22.	1	2	**3**	4	5	DK	0	1
23.	1	2	3	4	**5**	DK	0	1
24.	1	2	3	**4**	5	DK	0	1
25.	1	2	**3**	4	5	DK	0	1

Item	Response						Score	
26.	1	**2**	3	4	5	DK	0	1
27.	**1**	2	3	4	5	DK	0	1
28.	1	**2**	3	4	5	DK	0	1
29.	**1**	2	3	4	5	DK	0	1
30.	1	2	3	4	**5**	DK	0	1
31.	1	**2**	3	4	5	DK	0	1
32.	1	2	3	**4**	5	DK	0	1
33.	1	2	**3**	4	5	DK	0	1
34.	**1**	2	3	4	5	DK	0	1
35.	1	2	3	4	**5**	DK	0	1

Total Raw Score (Maximum = 35)

Stop Time ___:___
Hr. Min.

9. Comprehension

Start Time ___:___
Hr. Min.

Start
Ages 6–8: Item 1
Ages 9–11: Item 3
Ages 12–16: Item 5

Reverse
Ages 9–16: Score of 0 on *either* of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 4 consecutive scores of 0

Score
Score 0, 1, or 2 points
See the *Administration and Scoring Manual* for sample responses.

	Item	Response	Score		
6–8	†1. دانت		0	1	2
	2. سبزیاں		0	1	2
9–11	3. لڑائی		0	1	2
	4. سیٹ بیلٹ		0	1	2
12–16	5. ورزش		0	1	2
	6. پولیس		0	1	2

† If the child does not give a 2-point response, provide the response indicated in the *Administration and Scoring Manual*.
* If the child replies with only one general idea, ask for a second response as indicated in the *Administration and Scoring Manual*.

Continue

10. Symbol Search

(Time Limit: 120 seconds)

Start Time___:___ Hr. Min.

Start
Ages 6–7: Symbol Search A Sample Items, Practice Items, then Test Items
Ages 8–16: Symbol Search B Sample Items, Practice Items, then Test Items

Discontinue
After 120 seconds

Score
Use the Scoring Key to check the child's responses.
Subtract Number Incorrect from Number Correct.
If the Total Raw Score is <0, enter 0.

Completion Time	Stop Time___:___ Hr. Min.	Number Correct	−	Number Incorrect	=	Total Raw Score (Ages 6–7: Max = 45) (Ages 8–16: Max = 60)

11. Picture Completion

(Time Limit: 20 seconds)

Start Time___:___ Hr. Min.

Start
Ages 6–8: Sample, then Item 1
Ages 9–11: Sample, then Item 5
Ages 12–16: Sample, then Item 10

Reverse
Ages 9–16: Score of 0 on the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 6 consecutive scores of 0

Score
Score 0 or 1 point
See the *Administration and Scoring Manual* for sample responses.

Item	Response	Score
6–16	پنسل	
6–8 †1.	جیکٹ	0 1
†2.	لومڑی	0 1
3.	بلی	0 1
4.	ہاتھ	0 1
9–11 5.	گھنٹی	0 1
6.	رسی کودنا	0 1
7.	آری	0 1
8.	الماری کا دروازہ	0 1
9.	چ	0 1
12–16 10.	آئینہ	0 1
11.	سیڑھی	0 1
12.	قینچیاں	0 1
13.	گھڑیال	0 1

Item	Response	Score
14.	دروازہ	0 1
15.	بلب	0 1
16.	چہرہ	0 1
17.	بیلٹ	0 1
18.	بیڈ	0 1
19.	ڈائس	0 1
20.	بائیک	0 1
21.	درخت	0 1
22.	فٹ بال	0 1
23.	سیٹی	0 1
*24.	بلا	0 1
25.	سور	0 1
26.	پتی	0 1
27.	بائیسکل	0 1

Item	Response	Score
28.	چھتری	0 1
29.	شکل	0 1
30.	سنہری مچھلی	0 1
31.	ٹریلیس	0 1
32.	تھرمامیٹر	0 1
33.	ہاتھ	0 1
34.	سپر مارکیٹ	0 1
35.	پائی	0 1
36.	خاندان	0 1
37.	گھر	0 1
38.	جوتا	0 1

Total Raw Score (Maximum = 38)

Stop Time___:___ Hr. Min.

† If the child does not give a 1-point response, provide the response indicated in the *Administration and Scoring Manual*.
* Responses requiring specific query are identified in the *Administration and Scoring Manual*.

12. Cancellation

(Time Limit: 45 seconds)

Start Time___:___ Hr. Min.

Start
Ages 6–16: Sample, Practice, then Item 1

Discontinue
After 45 seconds for each item

Score
Use the Scoring Template to check the child's responses.
Subtract Number Incorrect from Number Correct. If the Total Raw Score is <0, enter 0.
CAR and CAS: Total Raw Score for Items 1 and 2, respectively.

Item	Time Limit (Seconds)	Completion Time	Number Correct	Number Incorrect	Difference	Bonus Points	Total Raw Score
6–16 1. Random	45						CAR Max = 68
2. Structured	45						CAS Max = 68

Time Bonus Points If the child completes an item before 45 seconds and the difference is ≥ 60, award bonus points					
Time in Seconds	45	40–44	35–39	30–34	0–29
Bonus Points	0	1	2	3	4

Total Raw Score (Maximum = 136)

Stop Time___:___ Hr. Min.

15. Word Reasoning

Start Time ___ : ___
Hr. Min.

Start
Ages 6-9: Samples A & B, then Item 1
Ages 10-16: Samples A & B, then Item 5

Reverse
Ages 10–16: Score of 0 on *either* of the first two items given, administer preceding items in reverse order until two consecutive perfect scores are obtained.

Discontinue
After 5 consecutive scores of 0

Score
Score 0 or 1 point
See *Administration and Scoring Manual* for sample responses.

Score	Correct		Response	
	Y N	I. یہ ایک جانور ہے جو "بھوں ووں" کرتا ہے۔		
	Y N	I. اس کا ایک لمبا ہینڈل ہوتا ہے۔۔۔		
	Y N	II. اور اسے فرش صاف کرنے کے لئے استعمال کیا جاتا ہے۔		
0 1	Y N	I. اسے نہانے کے بعد خود کو خشک کرنے کے لئے استعمال کیا جاتا ہے۔		1.
0 1	Y N	I. یہ آپ کے سر کا حصہ ہے جو سونگھنے کے لئے استعمال کے جاتا ہے۔		2.
0 1	Y N	I. یہ ایسی چیز ہے جو ہم سر ڈھانپنے کے لئے پہنتے ہیں۔		3.
0 1	Y N	I. ایسا جانور جس کی لمبی سونڈ اور بڑے بڑے کان ہوتے ہیں۔		4.
0 1	Y N	I. ایسا کمرہ جہاں لوگ سوتے ہیں۔		5.
0 1	Y N	I. اس کا ایک ہینڈل ہوتا ہے اور لوگ اسے کھول سکتے ہیں اور اس میں سے گزر سکتے ہیں۔		6.
0 1	Y N	I. یہ اوپر آسمان پر ہوتے ہیں۔۔۔		7.
	Y N	II. اور آپ انہیں صرف رات میں دیکھ سکتے ہیں۔		
0 1	Y N	I. یہ بہت سے رنگوں میں آتا ہے۔۔۔		8.
	Y N	II. اور یہ دیواروں پر استعمال ہوتا ہے۔		
0 1	Y N	I. یہ بارش کے بعد زمین پر ہوتی ہے۔۔۔		9.
	Y N	II. اور یہ آپ کے کپڑوں کو گندا کر سکتی ہے۔		
0 1	Y N	I. یہ ایک سیکھنے کی جگہ ہے۔۔۔		10.
	Y N	II. اور اس میں ماضی کی بہت سی چیزیں ہو سکتی ہیں۔		
0 1	Y N	I. یہ کھانے کا ذائقہ بہتر بنا سکتا ہے۔۔۔		11.
	Y N	II. اور یہ سمندر سے مل سکتا ہے۔		
0 1	Y N	I. یہ قدرتی طور پر پایا جاتا ہے۔۔۔		12.
	Y N	II. اور اس کے منہ سے پانی بہتا ہے۔		
0 1	Y N	I. یہ ایسا کام ہے جو لوگ چیزوں کو "نیا جیسا" کرنے کے لئے کرتے ہیں۔۔۔		13.
	Y N	II. اور لوگ اسے ناکارہ چیزوں کے ساتھ کرتے ہیں۔		
0 1	Y N	I. ان کی حفاظت جلد اور ہڈیوں سے کی گئی ہے۔۔۔		14.
	Y N	II. اور بعض اوقات انہیں تبدیل بھی کیا جا سکتا ہے۔		
0 1	Y N	I. اسے توڑا جا سکتا ہے۔۔۔		15.
	Y N	II. اور اسے لوگوں کی حفاظت کے لئے لکھا گیا ہے۔		

Continue

Notes / Behavioral Observations :

Appendix M

Home Back

POPULATION BY PROVINCE/REGION SINCE 1951

Province / Region		Population 1951	1961	1972	1981	1998
Pakistan		33,740,167	42,880,378	65,309,340	84,253,644	132,352,279
	Rural	27,754,670	33,225,806	48,715,689	60,412,173	89,315,875
	Urban	5,985,497	9,654,572	16,593,651	23,841,471	43,036,404
NWFP		4,556,545	5,730,991	8,388,551	11,061,328	17,743,645
	Rural	4,051,800	4,972,475	7,192,896	9,395,675	14,749,561
	Urban	504,745	758,516	1,195,655	1,665,653	2,994,084
FATA		1,332,005	1,847,195	2,491,230	2,198,547	3,176,331
	Rural	1,332,005	1,822,547	2,477,930	2,198,547	3,090,858
	Urban	-	24,648	13,300	-	85,473
Punjab		20,540,762	25,463,974	37,607,423	47,292,441	73,621,290
	Rural	16,972,686	19,988,052	28,424,728	34,240,795	50,602,265
	Urban	3,568,076	5,475,922	9,182,695	13,051,646	23,019,025
Sindh		6,047,748	8,367,065	14,155,909	19,028,666	30,439,893
	Rural	4,279,621	5,200,047	8,430,133	10,785,630	15,600,031
	Urban	1,768,127	3,167,018	5,725,776	8,243,036	14,839,862
Balochistan		1,167,167	1,353,484	2,428,678	4,332,376	6,565,885
	Rural	1,022,618	1,125,016	2,029,094	3,655,604	4,997,105
	Urban	144,549	228,468	399,584	676,772	1,568,780
Islamabad		95,940	117,669	237,549	340,286	805,235
	Rural	95,940	117,669	160,908	135,922	276,055
	Urban	-	-	76,641	204,364	529,180

E-mail:
censuspak@yahoo.com

Website:
http://www.census.gov.pk

Appendix N.

FAUJI FOUNDATION

CREATED UNDER THE CHARITABLE ENDOWMENTS ACT 1890
FOR THE BENEFIT OF EX-SERVICEMEN AND THEIR FAMILIES

Welfare Division (Edn)
Education Department
68 Tipu Road Chaklala
P.O Box-84
Rawalpindi
Tel:5951821-40 Ext-1721

NoWD/Edn/1/GC-Colleges

7 Mar 2013

To:
FF College for Boys
FF College for Girls
FFMS Chaklala-III
FFMS Renala Khurd
FFMS Quetta-1&II
FFMS Abbotabad
FFMS Lahore
FFMS Karachi I&II
FFMS Kohat
FFMS Multan
FFMS Hyderabad

Subject: **Issuance of Permission for Test Administration in Various Schools and Colleges of Fauji Foundation (Affiliated with FBISE)**

1. Miss Saima Ambreen D/O Major (Retd) Imtiaz Hussain who is a Ph.D scholar at Quaid-i-Azam Univeristy Islamabad has been given research work regarding translation and standardization of Wechsler Intelligence Scale for Children (WISC-IV South Asia) in collaboration with Pearson Assessment (NCS Pearson India private Limited Bangalore). Children from 6-16 years of age would be tested in intelligence.

2. Scope/aspect req to be seen in the students are as under:-
   a. Verbal comprehension index.
   b. Perceptual reasoning.
   c. Working memory.
   d. Processing speed

3. Miss Saima Ambreen has been permitted to visit Colleges/ as per the time schedule given by the Principal for completion of her assignment. Contact no is 0336-5134289 for coordination.

4. Fwd for info/nec action, please.

Col Muhammad Zubair Masood Khan (Retd)
Senior Manager (Edn)
(Colleges & Computers)

Copy to : Miss Saima Ambreen Ph.D Scholar
National Institute of Psychology (NIP)
Quaid-i-Azam University Islamabad

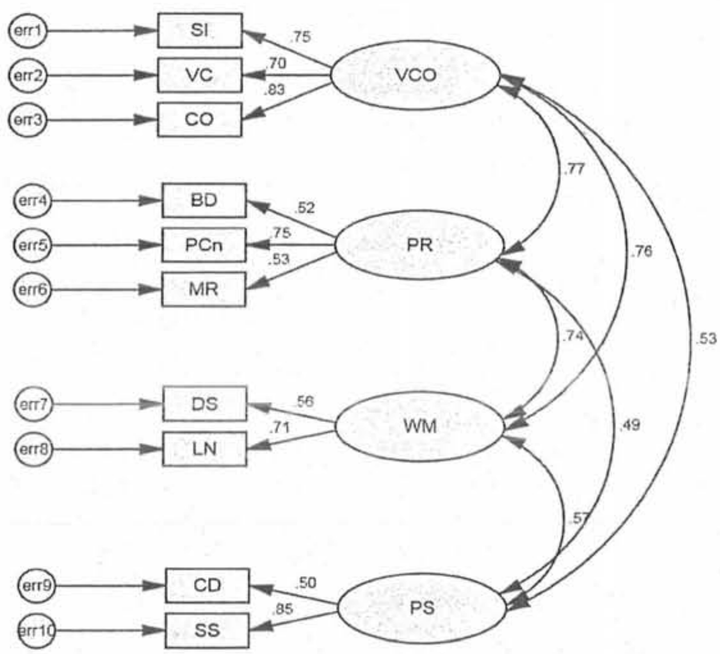
Item-total Correlations of Supplemental Subtests of WISC-IV PAK (N= 800)

Item	Picture Completion		Information		Arithmetic		Word Reasoning	
No.	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
1	.04	.246	.09	.014	-	-	.28	.000
2	.06	.080	-	-	-	-	.31	.000
3	.17	.000	.06	.081	-	-	.24	.000
4	.32	.000	-	-	.04	.190	.39	.000
5	.38	.000	.14	.000	.24	.000	.42	.000
6	.39	.000	.21	.000	.32	.000	.47	.000
7	.39	.000	.29	.000	.28	.000	.33	.000
8	.44	.000	.35	.000	.33	.000	.39	.000
9	.40	.000	.36	.000	.45	.000	.50	.000
10	.44	.000	.44	.000	.43	.000	.58	.000
11	.42	.000	.51	.000	.55	.000	.67	.000
12	.42	.000	.65	.000	.51	.000	.62	.000
13	.45	.000	.72	.000	.54	.000	.60	.000
14	.42	.000	.70	.000	.63	.000	.62	.000
15	.52	.000	.72	.000	.65	.000	.68	.000
16	.58	.000	.71	.000	.68	.000	.65	.000
17	.49	.000	.74	.000	.65	.000	.65	.000
18	.55	.000	.71	.000	.64	.000	.67	.000
19	.46	.000	.76	.000	.67	.000	.53	.000
20	.59	.000	.72	.000	.67	.000	.61	.000
21	.61	.000	.72	.000	.69	.000	.43	.000
22	.61	.000	.72	.000	.67	.000	.40	.000
23	.61	.000	.61	.000	.75	.000	.38	.000
24	.61	.000	.64	.000	.62	.000	.31	.000
25	.57	.000	.62	.000	.65	.000		.000

26	.58	.000	.65	.000	.58	.000
27	.53	.000	.53	.000	.48	.000
28	.57	.000	.57	.000	.40	.000
29	.47	.000	.46		.54	.000
30	.50	.000	.37		.50	.000
31	.55	.000	.41		.30	.000
32	.44	.000	.31		.29	.000
33	.44	.000	.25		.27	.000
34	.47	.000			.20	.000
35	.39	.000				.000
36	.37	.000				
37	.32	.000				
38	.32	.000				

Appendix P1

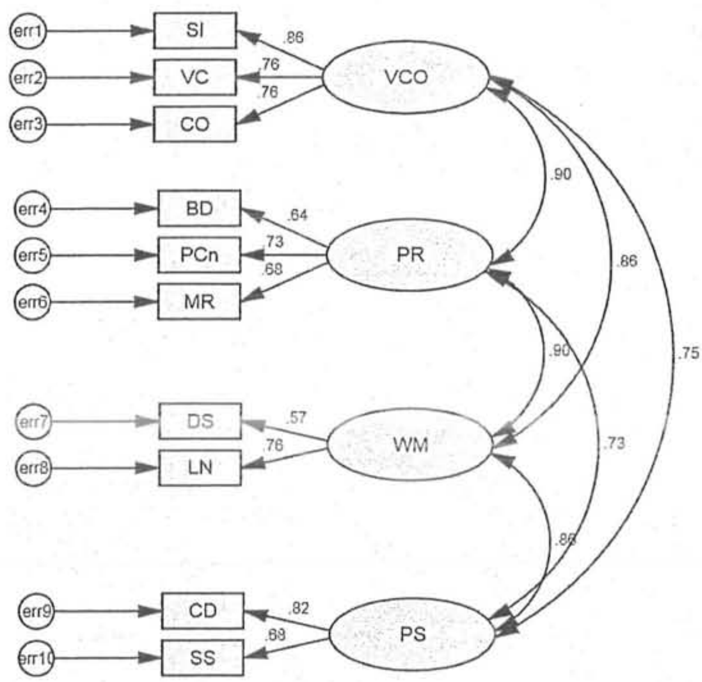
Figural Presentation of Four-Factor Structure of WISC-IV Core Subtests for 6 to 8 Years



The oval ellipse represent latent factor; the rectangles represents the subtests; the small circles represent the error variance

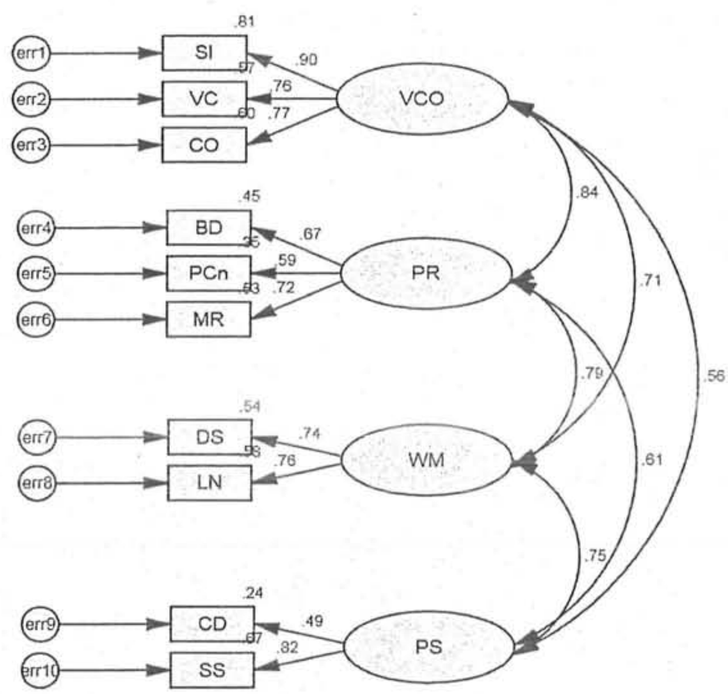
Appendix P2

Figural Presentation of Four-Factor Structure of WISC-IV Core Subtests for 9 to 12 Years



The oval ellipse represent latent factor; the rectangles represents the subtests; the small circles represent the error variance

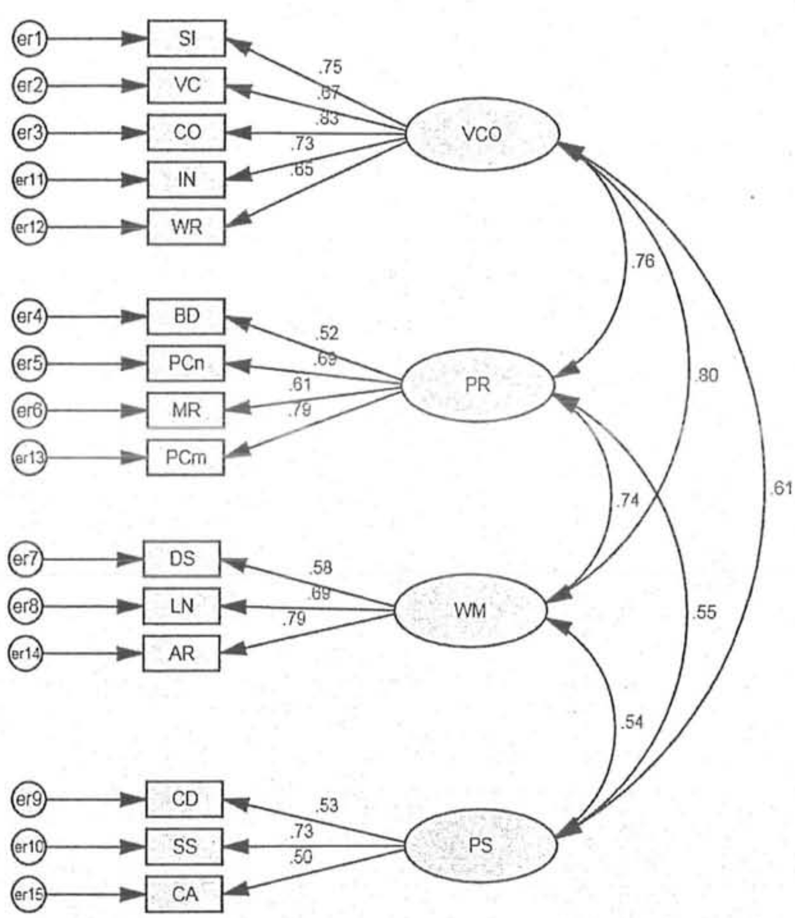
Figural Presentation of Four-Factor Structure of WISC-IV Core Subtests for 13 to 16 Years



The oval ellipse represent latent factor; the rectangles represents the subtests; the small circles represent the error variance

Appendix Q1

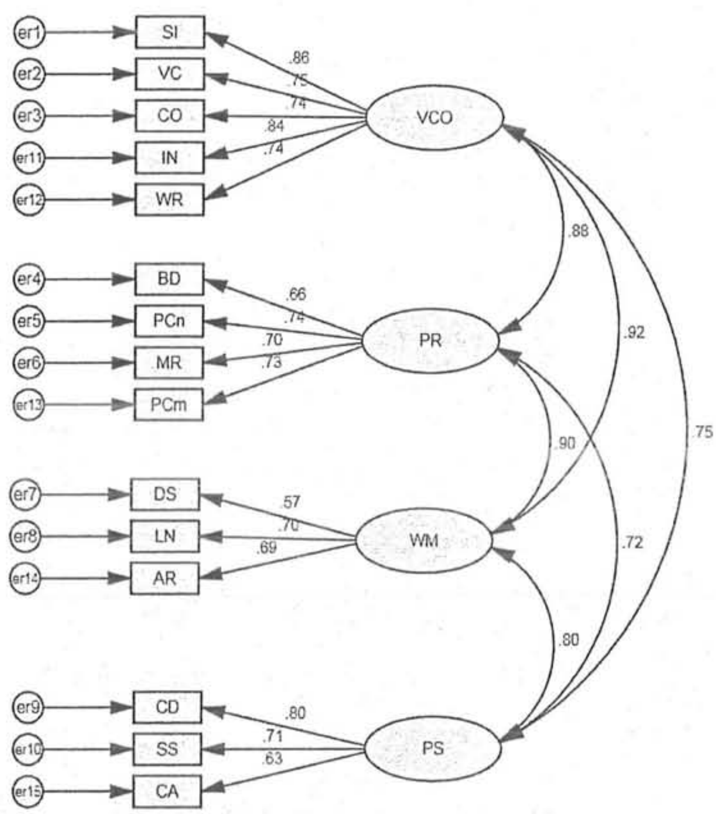
Figural Presentation of Four-Factor Structure of WISC-IV All Subtests for 6 to 8 Years



The oval ellipse represent latent factor; the rectangles represents the subtests; the small circles represent the error variance

Appendix Q2

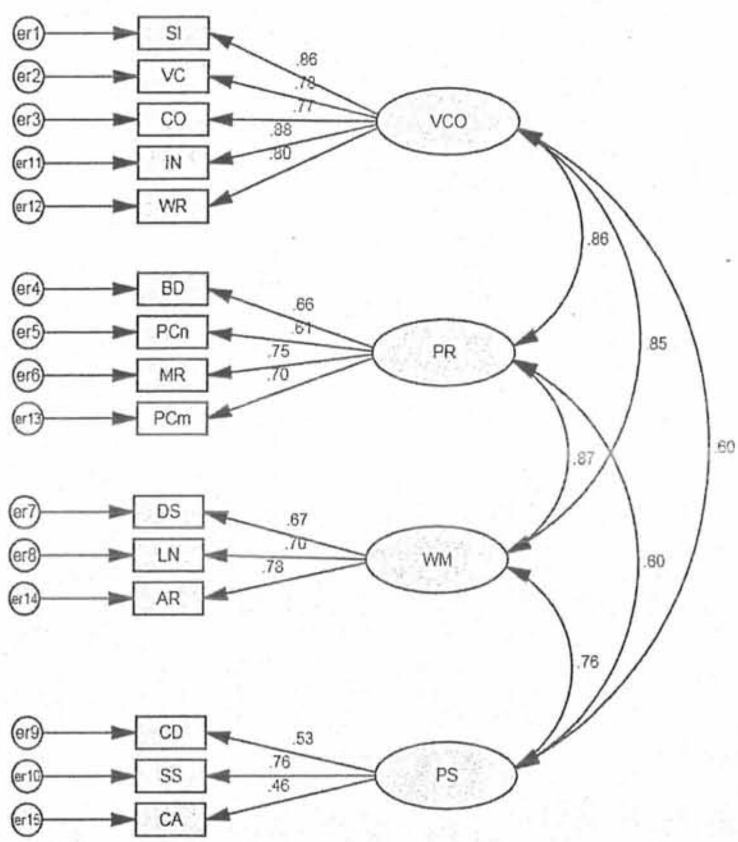
Figural Presentation of Four-Factor Structure of WISC-IV All Subtests for 9 to 12 Years



The oval ellipse represent latent factor; the rectangles represents the subtests; the small circles represent the error variance

Appendix Q3

Figural Presentation of Four-Factor Structure of WISC-IV All Subtests for 13 to 16 Years



The oval ellipse represent latent factor; the rectangles represents the subtests; the small circles represent the error variance

Appendix R1: Scaled Score Equivalents of Total Raw Scores for Subtests, by Age Group

AGES 6:00 – 6:11

Scaled Score	BD	SI	DS	PCn	CD	VC	LN	MR	CO	SS	PCm	CA	IN	AR	WR
1	-	-	-	-	0-5	0-7	-	-	-	-	-	0	0	0	0
2	-	-	-	-	6-9	8	0	-	-	0	0-2	-	-	-	-
3	-	0	7	0	10-13	9	1	0	0	1-3	3-4	-	5	-	-
4	0	1	8	1	14-17	10	2	1-2	1	4	5-6	10-15	-	5	2
5	1-2	2	9	2-3	18-20	11	3	3	2	5-6	7-8	16-21	6	6	3
6	3-4	3-4	10	4	22-26	12	4	4-5	3-4	7-9	9	22-24	7	7	4
7	5-6	5	-	5	27-30	13	5-6	6	5-6	10-11	10-11	30-32	-	8	5
8	7-8	6	11	6-7	31-34	14	7	7-8	7-8	12-13	12	33-38	8	9	-
9	9-10	7	12	8	35-38	15	8-9	9	9	14-16	13	39-43	9	10-11	6
10	11-12	8	13	9	39-42	16	10	10-11	10-11	17-18	14-16	44-49	-	12	7
11	13	9	14	10-11	43-46	17	11	12	12-13	19-20	17-18	50-55	10	13	8
12	14-15	10	15	12	47-50	18	12-13	13	14	21-23	19-20	56-58	11	14	9
13	16-17	11	-	13	51-54	19	14	14-15	15-16	24-25	21	62-66	-	15	-
14	18-19	12	16	14-15	55-58	20	15	16	17-18	26-27	22	67	12	16-17	10
15	20-21	13-14	17	16	59-62	21	16-17	17-18	19-20	28-30	23-24	73	13	18	11
16	22-23	15	18	17	63	22	18	19	21	31-33	25	82	14	19	-
17	24-25	16	19	18	-	23	19	20-21	22-23	34-35	26	87	15	20	12
18	26-27	17	20	19	-	24-25	20-21	22	24-25	36-37	27-29	92-103	16	-	13
19	28-68	18-44	21-32	20-28	65	26-68	22-30	23-35	26-45	38-45	30-38	104-136	17-33	34	14-24

Appendix R11: **Scaled Score Equivalents of Total Raw Scores for Subtests, by Age Group** AGES 16:00 – 16:11

Scaled Score	BD	SI	DS	PCn	CD	VC	LN	MR	CO	SS	PCm	CA	IN	AR	WR
1	0-13	0-5	0-9	0-9	0-30	0-22	0-13	0-10	0-12	0-6	0-12	0-40	0-12	0-10	0-5
2	14-16	6-9	10-11	10-11	31-34	23-25	-	11	13-14	7-10	13-15	41-50	-	11-18	6-7
3	17	10-12	12	12	35-36	26-27	14	12	15-16	11-12	16-18	51-56	13	19	8
4	18-21	13	13	-	37-40	28-30	15	13-14	17-18	13-15	19-20	57-58	14	20	9
5	22-25	14-15	14	13	41-43	31	16	15	19-20	16-17	21-22	64	15	21	10
6	26-27	16-17	15	14	44-47	32-34	17	16-17	21-22	18-19	23-24	67-68	16-17	22	11
7	28-30	18-20	16-17	15	48-51	35-37	18	18	23	20-22	25-26	72-76	18-19	23	12
8	31-34	21-22	18	16	52-55	38-39	19	19	24-25	23-25	27-28	77-82	20	24	13
9	35-37	23-25	19	17	56-58	40-42	-	20-21	26-27	26-27	29	83-87	21-22	25	14-15
10	38-40	26-27	20-21	18	59-62	43-44	20	22	28-29	28-30	30	88-92	23-24	26	16
11	41-43	28-29	22	-	63-66	45-47	21	23-24	30-31	31-32	31-32	93-97	25	27	17
12	44-47	30-32	23	19	67-70	48-50	22	25	32-33	33-35	33	98-102	26-27	-	18
13	48-50	33-34	24	20	71-73	51-53	23	26-27	34-35	36-38	34	103-107	28	28	19
14	51-53	35-37	25-26	21	74-77	54-55	24	28	36	39-40	35	108-113	29-30	29	20-21
15	54-55	38-39	27	22	78-81	56-58	25	29-30	37-38	41-44	36	114-117	31	30	22
16	59-60	40	28	23-26	82-84	59-60	26	31	39	45	37	118-121	32	31	23
17	61-66	-	-	-	85-88	61	27	32-33	40	46	-	122-125	-	32	-
18	67	-	-	27	89-92	62	28	34	41	47-49	-	126-128	-	33	-
19	68	41-44	29-32	28	93-119	63-68	29-30	35	42	50-60	38	129-136	33	34	24

Appendix S1

VCI Composites of Sums of Scaled Scores

Sum of Scaled Scores	VCI	Confidence Level 95%	Sum of Scaled Scores	VCI	Confidence Level 95%
13	72	67-77	34	107	102-112
14	74	69-79	35	108	103-113
15	75	70-80	36	110	105-115
16	77	72-82	37	112	107-117
17	78	73-83	38	113	108-118
18	80	75-85	39	115	110-120
19	82	77-87	40	117	112-122
20	83	78-88	41	118	113-123
21	85	80-90	42	120	115-125
22	87	82-92	43	122	117-127
23	88	83-93	44	123	118-128
24	90	85-95	45	125	120-130
25	92	87-97	46	127	122-132
26	93	88-98	47	128	123-133
27	95	90-100	48	130	125-135
28	97	92-102	49	132	127-137
29	98	93-103	50	133	128-138
30	100	95-105	51	135	130-140
31	102	97-107	52	137	132-142
32	103	98-108	53	139	134-144
33	105	100-110			

PRI Composites of Sums of Scaled Scores

Sum of Scaled Scores	PRI	Confidence Level 95%	Sum of Scaled Scores	PRI	Confidence Level 95%
9	66	57-75	30	100	91-109
10	67	58-76	31	102	93-111
11	68	59-77	32	103	94-112
12	70	61-79	33	105	96-114
13	72	63-81	34	107	98-116
14	74	65-83	35	108	99-117
15	76	67-85	36	110	101-119
16	77	68-86	37	112	103-121
17	78	69-87	38	113	104-122
18	80	71-89	39	115	106-124
19	82	73-91	40	117	108-126
20	83	74-92	41	118	109-127
21	85	76-94	42	120	111-129
22	87	78-96	43	121	112-130
23	88	79-97	44	123	114-132
24	90	81-99	45	125	116-134
25	92	83-101	46	126	117-135
26	93	84-102	47	128	119-137
27	95	86-104	48	130	121-139
28	97	88-106	49	132	123-141
29	98	89-107	50	133	124-142

WMI Composites of Sums of Scaled Scores

Sum of Scaled Scores	WMI	Confidence Level 95%	Sum of Scaled Scores	WMI	Confidence Level 95%
4	60	50-70	25	112	102-122
5	63	53-73	26	115	105-125
6	66	56-76	27	117	107-127
7	68	58-78	28	120	110-130
8	71	61-81	29	123	113-133
9	73	63-83	30	125	115-135
10	76	66-86	31	127	117-137
11	79	69-89	32	130	120-140
12	81	71-91	33	132	122-142
13	84	74-94	34	134	124-144
14	86	76-96	35	137	127-147
15	88	78-98	36	139	129-149
16	91	81-101	37	142	132-152
17	93	83-103			
18	96	86-106			
19	98	88-108			
20	101	91-111			
21	103	93-113			
22	105	95-115			
23	107	97-117			
24	110	100-120			

Appendix S4

PSI Composites of Sums of Scaled Scores

Sum of Scaled Scores	PSI	Confidence Level 95%	Sum of Scaled Scores	PSI	Confidence Level 95%
2	55	39-71	23	107	91-123
3	57	41-73	24	110	94-126
4	60	44-76	25	112	96-128
5	63	47-79	26	115	99-131
6	65	49-81	27	117	101-133
7	67	51-83	28	120	104-136
8	70	54-86	29	122	106-138
9	73	57-89	30	125	109-141
10	75	59-91	31	127	111-143
11	77	61-93	32	130	114-146
12	80	64-96	33	132	116-148
13	83	67-99	34	135	119-151
14	86	70-102	35	137	121-153
15	88	72-104	36	140	124-156
16	90	74-106			
17	92	76-108			
18	95	79-111			
19	97	81-113			
20	100	84-116			
21	102	86-118			
22	105	89-121			

Appendix S5

FSIQ Composites of Sums of Scaled Scores

Sum of Scaled Scores	FSIQ	Confidence Level 95%	Sum of Scaled Scores	FSIQ	Confidence Level 95%
38	69	61-77	61	80	72-88
39	69	61-77	62	81	73-89
40	70	62-78	63	81	73-89
41	70	62-78	64	82	74-90
42	71	63-79	65	82	74-90
43	71	63-79	66	83	75-91
44	72	64-80	67	83	75-91
45	72	64-80	68	84	76-92
46	73	65-81	69	84	76-92
47	73	65-81	70	85	77-93
48	74	66-82	71	85	77-93
49	74	66-82	72	86	78-94
50	75	67-83	73	86	78-94
51	75	67-83	74	87	79-95
52	76	68-84	75	87	79-95
53	76	68-84	76	88	80-96
54	77	69-85	77	88	80-96
55	77	69-85	78	89	81-97
56	78	70-86	79	89	81-97
57	78	70-86	80	90	82-98
58	79	71-87	81	90	82-98
59	79	71-87	82	91	83-99
60	80	72-88	83	91	83-99

FSIQ Composites of Sums of Scaled Scores (cont....)

Sum of Scaled Scores	FSIQ	Confidence Level 95%	Sum of Scaled Scores	FSIQ	Confidence Level 95%
84	92	84-100	107	103	95-111
85	92	84-100	108	104	96-112
86	93	85-101	109	104	96-112
87	93	85-101	110	105	97-113
88	94	86-102	111	105	97-113
89	94	86-102	112	106	98-114
90	95	87-103	113	106	98-114
91	95	87-103	114	107	99-115
92	96	88-104	115	107	99-115
93	96	88-104	116	108	100-116
94	97	89-105	117	108	100-116
95	97	89-105	118	109	101-117
96	98	90-106	119	109	101-117
97	98	90-106	120	110	102-118
98	99	91-107	121	110	102-118
99	99	91-107	122	111	103-119
100	100	92-108	123	111	103-119
101	100	92-108	124	112	104-120
102	101	93-109	125	112	104-120
103	101	93-109	126	113	105-121
104	102	94-110	127	113	105-121
105	102	94-110	128	114	106-122
106	103	95-111	129	114	106-122

FSIQ Composites of Sums of Scaled Scores

Sum of Scaled Scores	FSIQ	Confidence Level 95%	Sum of Scaled Scores	FSIQ	Confidence Level 95%
130	115	107-123	146	123	115-131
131	115	107-123	147	123	115-131
132	116	108-124	148	124	116-132
133	116	108-124	149	125	117-133
134	117	109-125	150	125	117-133
135	117	109-125	151	126	118-134
136	118	110-126	152	126	118-134
137	118	110-126	153	127	119-135
138	119	111-127	154	127	119-135
139	119	111-127	155	128	120-136
140	120	112-128	156	128	120-136
141	120	112-128	157	129	121-137
142	121	113-129	158	129	121-137
143	121	113-129	159	129	121-137
144	122	114-130	160	130	122-138
145	122	114-130			