12263

DEVELOPMENT AND VALIDATION OF AN INDIGENOUS INTELLIGENCE TEST

By

MAHMOOD SADIQ

Supervisor

Prof. Dr. Muhammad Jahanzeb Khan

Department of Psychology

University of Peshawar

A thesis submitted to

University of Peshawar

Khyber Pakhtonkhwa- Pakistan

in the partial fulfillment of the requirements

of the degree of

DOCTOR OF PHILOSOPHY

In

PSYCHOLOGY

2011

DEVELOPMENT AND VALIDATION OF AN INDIGENOUS INTELLIGENCE TEST

By

MAHMOOD SADIQ

Approved By

Supervisor

External Examiner

Chairperson

Department of Psychology

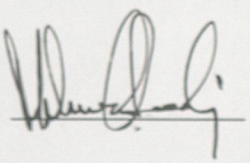
University of Peshawar

Declaration

I, MAHMOOD SADIQ, do hereby solemnly declare that the work submitted in this thesis is my own, and has not been presented previously to any other institution or university for a degree.

This work has carried out and completed at the Department of Psychology, Peshawar University.

Researcher:



DIRECTORATE OF QUALITY ENHANCEMENT
University of Peshawar, Peshawar, Pakistan

No: 12/DQE

January 2, 2013

Anti-Plagiarism Test Certificate
For MPhil/PhD Thesis/Synopsis

Plagiarism Detection Software Used:	Turnitin (HEC Recommended)
Name of Scholar:	Mr. Mahmood Sadiq
Level of Thesis/Synopsis:	PhD Thesis
Department/Institute/Centre:	Department of Psychology, University of Peshawar
Title of Thesis/Synopsis:	Development and Validation of an Indigenous Intelligence Test
Matching (Similarity) Index Found: (ASRB approved index = 25 %)	14% (Fourteen Percent only)
Comments: (Plagiarized/Not Plagiarized)	Not Plagiarized
Name & Signature of Supervisor: (Co-Supervisor)	14/1/13 Prof. Muhammad Jahanzeb Khan
Head of Department/Institute/Centre: office seal:	CHAIRPERSON Department of Psychology University of Peshawar.

Assistant Registrar (QE)
University of Peshawar

CONTENTS

List of Tables i

Acknowledgements iv

Abstract v

CHAPTER 1

INTRODUCTION

Importance of Psychological Testing 3

Nature of Intelligence 4

Different Approaches to study nature of Intelligence 7

Importance of Verbal Factor in Testing 18

Intelligence in Cultural Background 19

Intelligence Testing Across Cultures 23

Group Tests of Intelligence 25

Multilevel Group Intelligence Tests 26

Psychometric Research in Pakistan 28

Indigenous Development and Validation 33

CHAPTER II

RESEARCH DESIGN

CHAPTER III

METHOD

Layout of the Test 46

Conceptual Foundation of the test 46

Target Population 47

Test Content 47

First test draft 48

Try-outs 49

Item Analysis 52

Objectives of the Study 55

Main study 63

Reliability of the Test 63

Validity of the Test 64

Norms Development 65

CHAPTER III

RESULTS 68

CHAPTER IV

DISCUSSION 114

REFERENCES 126

APPENDIXES

LIST OF TABLES

Table No.	Description	Page No.
1.	Indices showing the discriminatory power of the items of each subtest of IIT (First Draft)	69
2.	Indices showing the difficulty level of the items of each subtest of IIT (First Draft)	70
3.	The number of items discarded from the test during the experimental tryout	71
4.	Three psychometric indices of the items of Vocabulary Test (Final Draft)	73
5.	Three psychometric indices of the items of Arithmetic Test (Final Draft)	74
6.	Three psychometric indices of the items of Analogy Test (Final Draft)	75
7.	Three psychometric indices of the items of Information Test (Final Draft)	76
8.	Three psychometric indices of the items of Comprehension Test (Final Draft)	77
9.	Three psychometric indices of the items of Similarity Test (Final Draft)	78
10.	Indices showing the discriminatory power of the items of each subtest of IIT (Final Draft)	79
11.	Indices showing the difficulty level of the items of each subtest of IIT (Final Draft)	80
12.	Average time needed for IIT	81
13.	Split Half Reliability coefficient of IIT and its subtests	83
14.	Reliability coefficient of IIT and its sub-tests (KR-20)	84
15.	Inter-correlation among subtests and IIT	85
16.	Concurrent Validity of IIT and its sub-tests	86
17.	Item total correlation for sub-test Vocabulary	87
18.	Item total correlation for sub-test Arithmetic	88

19.	Item total correlation for sub-test Analogy	89
20.	Item total correlation for sub-test Information	90
21.	Item total correlation for sub-test Comprehension	91
22.	Item total correlation for sub-test Similarity	92
23.	Mean, Standard Deviation and t-values showing significance of gender differences on IIT and its sub-tests	94
24.	Mean, Standard Deviation and t-values showing significance of urban and rural differences on IIT and its sub-tests	95
25.	Difference in test scores province wise	96
26.	Difference in test scores income group wise	98
27.	Percentile score of Vocabulary Test	101
28.	Percentile score of Arithmetic Test	102
29.	Percentile score of Analogy Test	103
30.	Percentile score of Information Test	104
31.	Percentile score of Comprehension Test	105
32.	Percentile score of Similarity Test	106
33.	Percentile score of IIT	107
34	Percentile score Province Punjab	108
35	Percentile score Province Sindh	109
36	Percentile score Province Balochistan	110
37	Percentile score Province Khyber Pakhtoonkhwa	111
38	Percentile score Federal Area	112
39	Percentile score Province Azad Jammu Kashmir	113

LIST OF APPENDIXES

A. Sampling plan for the present study

B. Keys of the sub-tests of first try out

C. Three indices obtained through item analysis of first experimental try out

D. Keys of sub-tests of IIT (Final Draft)

E Copy of IIT (Final Draft)

F. Used Answer Sheet (sample)

G. List of Educational Institutes for data collection

ACKNOWLEDGEMENT

In the name of God Almighty kind and merciful, I am really obliged to my supervisor, Dr Muhammad Jahanzeb Khan, Vice Chancellor, University of Swat, for his exceptional guidance, support, valuable advices and convincing criticism which helped me at every stage of this endeavor.

I owe my gratitude to my professional seniors / colleagues particularly Dr Aniq Ahmed (Henderson University, USA) Col Dr Tanvir Akhtar, Lt Col Dr Sajjad Hussain, Lt Col Dr Tahir Pervaiz, Lt Col Shoaib Kiani for their affectionate/professional attitude, concern and generous suggestions, which helped me in construction of items and data pool for the study and also kept my motivation high till completion of the study. My exclusive thanks to Mr Usman (NIP) for helping me in carrying out statistical analysis for the study with the help of SPSS and ITEMAN softwares.

I also owe my special gratitude to all my professional colleagues of the three services posted at various Selection Centers and in Inter Services Selection Boards for their professional help and support in collecting data from all the provinces of the country as without their assistance the task was difficult to accomplish.

Last, but not the least, my warm and deep gratitude to my family for their full cooperation that made me one step ahead in completing the job. In the end I cannot forget to mention the moral support in form of prayers I received from my parents that always gave me extra strength in time of difficulty.

M S

ABSTRACT

The aim of the study was to construct and standardize an Indigenous Intelligence Test (IIT) to measure general intelligence 'g' and a screening tool for youth intending to apply in Armed Forces, Public Service Commission and similar organizations where English carries a status of official language. The items of IIT were developed with the help of material obtained from own culture and were constructed in simple English language. The IIT comprises of six sub-scales: Vocabulary, Arithmetic, Analogy, Information, Comprehension, and Similarity. The test was planned to measure various aspects of intelligence verbal ability, to understand concepts, general reasoning factor, element of discovery, acquired knowledge, social intelligence, eduction of relations and correlates. The usefulness of the items was judged through item analysis by administering the test to a sample of 200 subjects with minimum Intermediate qualification. Items were evaluated by considering two main aspects; discrimination power and difficulty level. Items with discriminatory power less than .30 and level of difficulty below .30 and above .70 were discarded. The final draft of the test comprised of 71 items; Vocabulary and Arithmetic sub-scales consists of 12 items each, Analogy 14, Information 11, Comprehension 12 and Similarity 10. In the main study the test was administered to 1669 students both male and female, belonging to various Government Colleges & Universities of four provinces of the country including AJK and FATA. Candidates both from urban and rural areas were included in the study. Two methods were used to establish reliability of the test i.e., KR 20 and Split-half methods. The estimated indices of reliability were, .79 and .69 respectively. To determine validity of indigenous intelligence test different validity criteria were used including urban/rural differences, province wise differences, correlation of the total test and sub-scales (construct validity) and marks in last qualified examination to determine concurrent validity Significant differences in the mean scores of the male and female candidates on vocabulary, arithmetic, analogy and information were found whereas non significant difference was observed on comprehension and similarity subscales. Significant mean differences on vocabulary, arithmetic, analogy and information and non significant difference on comprehension and similarity were also found. Inter-correlation among the subscales and IIT were calculated to determine construct validity of the test. The concurrent validity was determined by correlating scores on IIT with the college marks obtained by each student in the last annual examination. Both indices established the evidence of high validity of the test. Significant differences in urban/rural mean scores on vocabulary, analogy, information, comprehension subscales and IIT were observed.

Significant differences were also observed in Province wise mean scores on vocabulary, arithmetic, analogy, information and comprehension subscales. Whereas, non significant mean scores were found on similarity subscale. Income group wise comparison revealed that higher the income groups better the performance. Percentile norms were developed for IIT and subscales while administering the test on a sample of 1669 subjects representative of Pakistani youth. In addition, province wise norms were also developed to gauge true performance of candidates hailing from different provinces.

INTRODUCTION

Measurement of human personality coupled with innate abilities carries much importance in our daily practical lives. Every student of this era is much familiar with the purpose and importance of an academically oriented test. These tests are being used abundantly in every educational institute to assess particular level of knowledge or skills that are attained over a particular period of time. Consequently, it is of importance that such tests that are called Achievement Tests are properly constructed otherwise a test is likely to produce unreliable results and as such will have a negative effect on the evaluation of an examinee evaluation.

Considering testing from research point of view, if the instruments used in a study are not valid and reliable, then the results obtained with the help of such a test is likely to lead to wrong conclusions. The information gathered with the help of a research study having poor validity and reliability indices cannot be used accurately to study test hypothesis and generalize the results. Due these very reasons, it is of importance that the instruments or scales being used in a research study should exactly measure the construct what it claims to measure and its score are stable when test is administered in different settings.

It can be asserted that tests are used to make important decisions about individual (Murphy & Davidshofer, 1998). Psychological tests are not only useful in making decisions in our daily lives; rather they are also used abundantly in a research process. Hence, it is of immense important, to use appropriate techniques to construct a new psychological test so that its result are reliable and valid when administered on a population for which it has been prepared..

Psychological test are used to measure a wide variety of attributes imbued in a human personality such as, intelligence, motivation, to determine the level of knowledge seventh-grade mathematics, vocational preferences, spatial ability, anxiety, form perception and so on. Psychological testing over the years has under gone a lot of criticism. Precisely, for not been able to make completely accurate decisions about individuals. Nevertheless, a special panel of the National Academy of Sciences concluded

that psychological tests generally represent the best, fairest, and most economical method of obtaining the information necessary to make sensible decisions about individuals (Wigdor & Garner, 1982a, 1982b). The conclusion reached by the National Academy panel and as reported by (Murphy, & Davidshofer, 1998), "although psychological tests are far from perfect, they represent the best, fairest, and most accurate technology available for making many important decisions about individuals".

Psychological testing in particular is not only important, rather it is a highly specialized and somewhat technical in nature. In many of the natural sciences, measurement is a relatively straightforward process that involves assessing the physical properties of objects, such as height, weight, or velocity. On the contrary, psychological attributes, such as intelligence and creativity, cannot be measured by the same methods that are applicable to measure physical attributes. Psychological attributes in a human personality are not manifested in any simple or physical way; rather they are visible through the behavior of individuals. In addition, behavior of an individual does not exhibit any particular psychological attribute; rather it is a combination of physical, psychological, and social characteristics that exists at the same time. Due this reason, psychological measurement is considered not as simple or direct as physical measurement.

In Pakistan, most of organizations and institutions providing testing services still use Western tests with minor modification/adaptation and translation but no serious effort has been made as yet to develop standardized indigenous tests of intelligence (Riaz, 2008). All tests of intelligence, to a greater or lesser degree, reflect the culture in which they were devised and will be used. In other words intelligence tests differ in the extent primarily due to the important factor of culture loading. One of the essential aspect of an indigenous intelligence test is that how much a test incorporates the vocabulary, concepts, traditions, knowledge, and feelings associated with a particular culture.

The dearth of research work in Pakistan (Riaz, 2008), particularly in developing indigenous intelligence tests, encouraged the researcher to undertake a research work aimed in developing an indigenous Intelligence Test for adolescents. The intended test will include items (as a measuring tool) culture friendly to Pakistani youth. There is no doubt in the field of testing and measurement significant efforts have been undertaken in various universities but ironically, much of the research work carried out was limited in its scope and focused only in translation and adaptation of foreign tests.

In education system of Pakistan, after qualifying Intermediate or Graduation level a stage arrives where the majority of students do apply in various professions i.e., in Armed Forces of Pakistan or appear in Public Service Commission exams or try to seek admission in professional institutes for obtaining specialized education. It is important to note that almost, in all professional institutes in general and in Armed Forces of Pakistan in particular English enjoys a status of official language. Hence it was decided to develop an intelligence test in English for Pakistani youth having age range between 17 to 25 years with minimum qualification of Intermediate, in order to examine individual differences and the proposed test that can also be used as a screening tool during the selection process of mentioned institutes. In addition, endeavors would be made to include population of both urban and rural areas of all four provinces (including FATA & AK areas) of the country and the study will be in line with the approach known as differential or psychometric (Riaz, 2008; page 88). As a consequence, resultant Norms can be considered true representative of test performance of standardized sample.

Importance of Psychological Testing

Psychological testing has been used extensively in the developed world primarily in the field of counseling, selection and placement. Institutions like Schools / colleges, civil & military services, industry, medical clinics, and counseling centers have benefited the most from of psychological testing. Therefore, it is logical to assume that outcome of psychological tests has excessively influenced lives of individuals. For example, in order to determine the eligibility of a student for allowing admission or otherwise in a college, to select or reject a person for a job, giving diagnoses of depression to a client or not, are all relevant areas demonstrating importance of psychological testing in daily life. Therefore, it can be concluded that results of psychological tests does play important role in determining future of an individual.

Fast progression in Intelligence Testing has earned a lot of approval due its positive role in facilitating the process that helps in taking appropriate practical decisions in daily practical life. A candid literature review makes obvious that topic of Intelligence within psychology has witnessed extensive research. Thousands of research articles are published each year on the nature and measurement of intelligence, Gregory (1996).

Following literature review is likely to help in developing better understanding of nature of psychology.

Nature of Intelligence

In an attempt to understand nature of intelligence, a layperson may ask, does intelligence exists at all? The easy reply to make him understand real nature of intelligence is by giving him an example from daily life. In which it can be argued that in daily routine we do label a person intelligent or otherwise just by observing his behavior while dealing with everyday problems. When a person is engaged in handling unusual and complex situation the importance of intelligence and its influence on behavior can be observed. Nevertheless, everyday observations cannot be considered sufficient enough to establish scientific basis of intelligence.

According to experts Eysenck (1979), Humphreys (1979), Jensen (1980), Guttman & Levy (1991), a person who performs very well on one test of intelligence , such as a reading comprehension test, is likely to perform equally well on other tests such as maze tests, analogy tests, and digit-symbol substitution tests. Furthermore, Ree & Earles (1991) hypothesized that scores on any ordinary paper and pencil tests that requires active information processing on part of examinee cannot produce identical estimate of general cognitive ability.

Both above mentioned research evidences clearly signifies that generally people differ in their cognitive / perceptual abilities, which is usually called as intelligence. Longitudinal studies carried out to understand human behavior have concluded that every individual is imbued with different level of general intelligence and this variation carries certain repercussion for different behavior. Moreover, observed differences in mental ability in every individual have also made the task of presenting a widely acceptable definition of intelligence a difficult preposition. Due this very reason we find that renowned psychologists have shown little agreement in their definitions of intelligence. Both Neisser (1979) and Miles (1957) have provided sufficient arguments to highlight difficulties confounding psychologist in defining psychological constructs such as intelligence.

Intelligence is a construct, not a physical entity (Humphreys, 1979; Eysenck, 1979). Therefore, it is not easy to produce a definition of intelligence that can cover all aspects of meaning of intelligence. However, intelligence can be defined in terms of behavior that points out various levels of intelligence. Thus, it seems more logical to focus on the behavior of an intelligent person rather than to try to make an assertion what intelligence is.

According to Tuckman (1975), "The modern history of testing is the history of testing for intelligence or mental ability." In recent past, psychologists have made numerous efforts to define and measure intelligence. Despite their frequent attempts, the meaning of the term (intelligence) is not free from controversy, and still there is not a single definition which is acceptable to all experts.

Therefore, it is considered important to have a close review of the known definitions of intelligence. They same has been illustrated in successive paragraphs.

Popular definitions of Intelligence

Sternberg (1986), has identified serious shortcomings that are imbued in most of the definitions of intelligence. First, Intelligence tests were invented to measure intelligence, not to define it. Second, operational definitions block further progress in understanding the nature of intelligence, as they fail to discuss the acceptability of theories of intelligence.

Thus, it is important to understand that operational definitions of intelligence have not been offered to explain real meaning of intelligence. Whereas, a real definition of intelligence is one that tries to tell us the true nature of the thing being defined (Robinson, 1950; Sternberg, 1986).

Intelligence has been defined differently by various experts. Definitions mentioned below were published in the Journal of Educational Psychology (Thorndike, 1921):

1. Spearman (1904, 1923) "a general ability which involves mainly the eduction of relations and correlates".
2. Binet and Simon (1905) "the ability to judge well, to understand well, to reason well".

3. Terman (1916) "the capacity to form concepts and to grasp their significance".
4. Pintner (1921) "the ability of the individual to adapt adequately to relatively new situations in life".
5. Thorndike (1921) "the power of good responses from the point of view of truth or fact".
6. Thurstone (1921) "the capacity to inhibit instinctive adjustments, flexibly imagine different responses, and realize modified instinctive adjustments into overt behavior".
7. Wechsler (1939) "the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with the environment".
8. Humphreys (1971) "the entire repertoire of acquired skills, knowledge, learning sets, and generalization tendencies considered intellectual in nature that are available at any one period of time".
9. Piaget (1972) "a generic term to indicate the superior forms of organization or equilibrium of cognitive structuring used for adaptation to the physical and social environment".
10. Sternberg (1985, 1986) "the mental capacity to automatize information processing and to emit contextually appropriate behavior in response to novelty; intelligence also includes Metacomponents, Performance components, and Knowledge-Acquisition components".
11. Eysenck (1986) "error-free transmission of information through the cortex".
12. Gardner (1986) "the ability or skill to solve problems or to fashion products which are valued within one or more cultural settings".
13. Ceci (1994) "multiple innate abilities which serve as a range of possibilities; these abilities develop (or fail to develop, or develop and later atrophy) depending upon motivation and exposure to relevant educational experiences".

After making a critical review of popular definitions of intelligence, two aspects were found common in each definition: (1) the capacity to learn from experience and (2) the capacity to adapt to one's environment. Thus it can be concluded that learning and adaptation are both crucial to intelligence.

As discussed earlier, researcher intends to develop an intelligence test in English language following differential or psychometric approach which is based on the assumption that the nature of intelligence can best be investigated by examining individual

differences in performance on tests of intellectual abilities (Riaz, 2008). In ensuing paragraphs attempt has been made to briefly study various approaches formulated to understand intelligence.

After reviewing the nature and definitions of intelligence it is considered important to closely examine how psychological testing emerged through pages of history prior reaching present status wherein it is considered capable enough to influence our daily life decisions.

Testing and Measurement in the field of psychology has reached to a place of significance primarily due to abundance of research work undertaken in the field of Psychometrics over the years. For that reason, it is considered of relevance to have a brief review of the theories proposed by various experts.

In addition, a researcher who intends to develop an intelligence test, foremost step is to define the construct operationally (Riaz, 2008). The next will be to take a decision regarding the approach (empirical, theoretical, rational) selected to construct items for the test.

In the following paragraphs an overview of literature available on Intelligence is being discussed.

Different Approaches to Study Nature of Intelligence

Following are approaches adopted by experts to understand nature of intelligence:

1. Differential or psychometric approach
2. Developmental approach
3. Information processing approach.

Differential or Psychometric Approach

This approach is based on the assumption that intelligence can best be measured by gauging individual differences in performance on tests of intellectual abilities. Due this very reason intelligence tests developed on the basis of this approach yields a single score

or IQ. Thus suggesting that intelligence is a global trait. However, there are others who define intelligence as a composite of separate, independent abilities.

Spearman's two-factor theory

Spearman (1927) has proposed two factors model of intelligence; the general factor (g) and specific factor (s). He found that every intellectual activity involves both a general factor (g) and a specific factor (s). Whereas, differences in intellectual abilities among individuals are due to differences in their g. Spearman suggested that a single test, highly loaded with g, may be used for assessing intelligence rather than a test developed with items collected from already available different intelligence tests. Later on, Spearman introduced another factor to his theory which is common to a group of activities and called it a Group Factor. Spearman also concluded that this factor is not universal as g and also not specific as the s factors. However, in later years due investigations carried out by his student, Spearman included broader group factors to his theory such as Arithmetic, Mechanical and Linguistic abilities (Gregory, 1996, 2004; Gross, 2005; Riaz, 2008).

Burt and Vernon's Hierarchical model

A student of Spearman, Burt (1971) also supported conclusions of his teacher and established that there is a 'g' factor common in all tests, but at the same time he concluded that the two-factor model is too simple. He and Vernon extended Spearman's model by identifying a series of Group Factors (major and minor) considered between 'g' and's' factors. According to their study 'g' is measured by all tests, whereas, group factors (major) are measured by certain tests. Similarly, group factors (minor) are also measured by exclusive tests whereas specific factors can be are measured by tests developed for specific occasions (Gregory, 1996, 2004; Gross 2005; Riaz, 2008).

Thurston's Primary Mental Abilities

Thurstone (1935, 1938, 1947) found that not all mental tests correlate highly to each other. Opposite to the findings of Spearman's theory, Thurstone (1931) introduced seven important group factors that he named as Primary Mental Abilities. These factors include V (Verbal comprehension), W (Word fluency), N (Number), S (Space), M (Associated memory), P (Perceptual Speed), and I or R (Induction or General Reasoning).

Later on, Thurstone and his associates developed a series of tests and named it as the Primary Mental Abilities Tests which is capable of measuring each factor. In addition, Thurstone developed factor-analysis procedures to find correlation matrices for the existence of group factors. This method helps a researcher to discover practically the numbers of factors present in a matrix and to define each factor in terms of the mental quality loaded on it.

Thurstone's analysis of PMAs continues to influence test development even today. Schaie (1983, 1985) has revised and modified the Primary Mental Abilities Test and used these measures in longitudinal study of Adult Intelligence. According to him, if intelligence involves only the g, then the group factors should also change at about the same rate as with age. On the contrary, he noted that some PMAs show little decrease with age (Verbal Comprehension, Word Fluency, Inductive Reasoning), whereas other PMAs decline more rapidly with age (Space, Number). Thus, this may be a good reason for reporting group factors and not limiting intelligence with a single general factor (Gregory, 1996, 2004; Gross 2005; Riaz, 2008).

Guilford's Structure of Intellect model

After World War II, J. P. Guilford (1967, 1985) started his research to determine factors of intelligence that had been proposed by Thurstone. However, Guilford soon concluded that the number of distinct mental abilities is in excess of the seven proposed by Thurstone. Guilford proposed 150 independent abilities or factors and also classified a cognitive task under three major dimensions. These are:

1. Content (what must participant think about.)
2. Operations (what kind of thinking is participant being asked to perform).
3. Products (what kind of answer is required).

Guilford proposed an elegant structure-of-intellect (SI) model. In total, Guilford identified five types of operations, five types of contents, and six types of products, for a total of 5 x 5 x 6 or 150 factors of intellect. Guilford claims to have verified over 100 of these factors in his research. However, (1992, 2003), claimed that individual's scores are often correlated, suggesting that the number of basic abilities are much smaller than assumed by Guilford (Gregory, 1996, 2004; Gross, 2005; Riaz, 2008).

Vernon's Hierarchical Model

British psychologist including Burt (1949), Vernon (1960) and Humphreys (1962) presented an alternative method for grouping of factors. According to Vernon (1950, 1971), intelligence is neither a single general mental ability nor a number of more specific independent abilities rather it is a combination of both. He was of the view, that there is no doubt that general intelligence is actively involved in all mental activities; but more specific abilities are required when complex tasks are to be performed. In hierarchical model, Vernon gave most importance to Spearman's 'g' factor by placing it on the top. Humphreys (1962, 1970) also preferred a hierarchical model for when grouping of specific factors. He suggested that each test user may choose from hierarchy, the level which he or she considered most relevant to its objectives. For example, if a user wishes to assess verbal ability, may use test including items developed to tap verbal facility, such as vocabulary, analogies, and series completion. Whereas, if a user wants to tap an individual's ability to solve analogies problems, tests comprising verbal, numerical, pictorial and spatial analogies items can be used.

The hierarchical model of intelligence both for theoretical and practical purposes has gained wide popularity. As a theoretical model, it favors presence of single general factor (Spearman's 'g') with associated multiple-factors. In addition, at practical level such tests have the capacity to cover wide-range of aptitudes and as such flexible enough in administration. Examples of tests developed on hierarchical model are Differential Ability Scale (Elliott, 1990) and Multidimensional Aptitude Battery (Jackson, 1993).

R. Cattell and the Fluid/Crystallized Distinction

Raymond Cattell (1941, 1971) theory of intelligence has enjoyed a status of popularity among test constructors and test users. The theory advocates presence of two major types of cognitive abilities i.e., fluid intelligence (Gf) and crystallized intelligence (Gc). To understand difference between Gf and Gc is important for every test developer (Riaz, 2008). Gc ability can be measured by items requiring knowledge and skills that are expected out of individuals of a particular age and cultural background. Whereas, Gf ability is closely related to an individual's inherent potentials necessary for acting intelligently. It also depends more on psychological and biological factors than on formal schooling. Fluid and Crystallized abilities are considered highly correlated at an early age

of 2-3 years (Riaz, 2008). But these abilities get apart from each other as a child grows older and come across various experiences in school and family. The more intelligent children exposed to rich environmental experiences are observed to invest most of their fluid talents in the crystallized skills required in their culture. Gf abilities can be measured by tests comprising of items that are relatively culture-free. On the other hand, Gc abilities can be effectively measured by tests including items taping vocabulary, comprehension, information etc.

Fluid intelligence has been identified depending on physiological efficiency therefore is considered relatively free from the influence of formal education (Horn, 1967). Therefore, fluid intelligence relates to a person's inherent capacity to learn and solve problems. Thus, fluid intelligence is used when a task requires adaptation to a new situation. On the other hand, Crystallized abilities develops from learning and exposure to various other cultures (acculturation), and can be measured by tests of knowledge, general information, use of language (vocabulary) and a wide variety of acquired (Horn & Cattell, 1967). In addition, Personality factors, motivation, educational and cultural opportunity are also important in its development. However, its dependence on physiological influence is indirect and mainly affects fluid abilities. Crystallized intelligence represents what one has learned by investing fluid intelligence in cultural settings (e.g., learning algebra in school). Crystallized intelligence is highly culturally dependent and is used for tasks which require a learned or habitual response.

Cattell (1940) believed that measures of fluid intelligence are primarily culture-free and due this very reason he devised the Culture-Fair Intelligence Test to eliminate cultural bias in testing. It is commonly found that measures of crystallized and fluid intelligence correlated moderately ($r = .5$). Catell's theory of fluid and crystallized intelligence was subsequently modified by Horn (1968, 1985, 1988, 1991,1994) and several other factors were included in the theory i.e., visual processing (Gv), auditory processing (Ga), quantitative processing (Gq), speed of processing (Gs), facility with reading and writing (Grw), short-term memory (Gsm) and long term storage and retrieval (Gir).

Three-Stratum Theory of Cognitive Ability

Three-stratum theory of cognitive ability is another influential multiple intelligence model based on factor-analytic studies proposed by Carroll (1993). In this model again 'g' or general intelligence is on the top level. Whereas, eight abilities and processes i.e., fluid intelligence (Gf), crystallized intelligence (Gc), general memory and learning (Y), broad retrieval capacity (R), broad cognitive speediness (S), and processing/decision speed (T) are identified in the second level. In addition, there are many level factors or speed factors under each of the abilities in this level. For example, general reasoning, quantitative reasoning, and Piagetian reasoning are level factors linked to Gf. Speed of reasoning is a speed factor linked to Gf, whereas oral fluency and writing ability are speed factor linked with Gc. Language development, comprehension, spelling ability, and communication ability includes four level factors linked to Gc.

The three-stratum theory is a hierarchical theory which means all of the abilities listed in a layer are included in the level above.

Sensory Keenness Theory of Intelligence

The sensory keenness theory of intelligence proposed by Sir Francis Galton and his student J. Mc Keen Cattell is another theory based on psychometric approach. In this theory intelligence is being studied with the help of an apparatus i.e., Reaction Time - Movement Time (RT-MT) apparatus. An experimental method recommended by Jensen (1980) to undertake culture-reduced study of intelligence. In RT-MT studies, the subject is to place on the home button, index finger of the hand he/she normally uses while managing daily working. Subsequently, an auditory warning signal will sound and in 1 to 4 seconds one of the eight green light will flash. Upon this the subject must turn off the light as quickly as possible by touching the micro-switch button directly below it. RT is the time the subject takes to remove his / her finger from the home button after a green light goes on. MT is the interval between removing the finger from the home button and touching the button that turns off the green light. Jensen (1980) reported that indices of RT and MT correlated as high as .50 with traditional psychometric tests of intelligence. Vernon (1994) has also reported substantial relationships as high as .70 for multiple correlations between speed-of-processing RT-type measures and traditional measures of intelligence (Vernon, 1994, Vernon & Mori, 1990). These findings suggest that by

including speed-of-processing measures such as RT can enhance testing capacity of standardized intelligence test batteries.

Cattell-Horn-Carroll (CHC) Model of Cognitive Abilities

The CHC model mix together Cattell-Horn theory with Carroll's three-stratum theory. McGrew (1997) originally proposed this model and subsequently modified by McGrew and Flanagan (1998) in light of their findings acquired through subsequent factor-analytic studies.

The McGrew-Flanagan CHC model includes ten broad level abilities such as; Fluid Intelligence (Gf), Crystallized Intelligence (Gc), Quantitative Knowledge (Gq), Reading/writing ability (Grw), Short-term Memory (Gsm), Visual processing (Gv), Processing Speed (Gs), and Decision/Reaction time or Speed (Gt).

The McGrew-Flanagan CHC model does not include the general intellectual ability factor 'g'. This exclusion does not mean that the integrated model does not accept existence of 'g'. In fact, it was omitted by McGrew (1997) since it has little practical relevance to cross-battery assessment and interpretation (McGrew & Flanagan, 1998, p, 14).

The CHC model has placed researchers at an advantage as they are at liberty to assess as many human abilities as they wish. It also encourages researchers to reexamine existing theories by means of statistical methods like factor analysis. Such model can also help in identifying best characteristics in existing theories that can be combined to create a useful and practicable model of human intelligence.

Developmental Approach

In the following paragraphs theory based on developmental approach will be discussed:

Theory of Cognitive Development

The Swiss psychologist Jean Piaget (1926, 1952, 1967, 1972, 1980) devised a theory of cognitive development, the best known theory based on developmental model. Piaget's approach is basically concerned with qualitative aspects of intelligence and not with measuring individual differences.

According to Piaget, intelligence increases in a child by a process in which balance continuously shifts between the assimilation of new information into existing cognitive arrangements and also by accommodating these arrangements to the new information. Assimilation and accommodation are the two corresponding processes identified by Piaget through which a child internalized awareness of the outside world.

Piaget (1963,) identified four major stages of cognitive development: Sensorimotor, Preoperational, Concrete Operational, and Formal Operational. During sensorimotor stage (birth to two years), a child explores external environment and develop own schemata through their senses and motor (muscular) activities.

Preoperational stage (2 to 7 years) is characterized by the use of words and symbols to represent objects and relationships among them. In this stage, a child begins to symbolize the world mentally, but is egocentric (self-centered). By the age of 7 years, a child enters the stage of concrete operations. In this stage, which lasts for about 12 years of age, a child starts showing capacity for undertaking logical thinking. Their moral judgments, management of self and other intellectual abilities begins to develop at this stage.

The stage of formal operations (12 years and above) is the final stage in Jean Piaget's theory of cognitive development and during this stage cognitive maturity is acquired by an adolescents. Mature adult thought process starts to emerge in this stage characterized by deductive logic, and abstract thinking.

Piaget's theory has dominated child psychology of twentieth century. However, some critics argue that Piaget has underestimated a child's abilities (Meltzoff 1989; Moore, 1989, Gopnik, 1997 & Bjorklund, 2004). Some researchers with the help of different methods have also found out that preschool children are less egocentric and capable of conservation at earlier ages than Piaget believed. Furthermore, it has been

argued by critics that formal education and specific cultural experiences can significantly influence cognitive development, a factor overlooked by Piaget (Berry et al., 1992; Brislin, 1993). Despite criticisms, Meltzoff, (1997) believed that "Piaget's theory is critical for getting the field of cognitive development off the ground and it's time to move on".

However, some psychologists are moving on to information process approach as they view children and adults similar to computer system. Like computers, children receive information (input) from their environments, store it, retrieve it, manipulate it, and then respond to it clearly in terms of their behavior (output). Brief detail of theories developed on the basis of this approach is appended below:

Information-Processing Approach

Now, the theories based on information-processing approach will be discussed in the following paragraphs:

Aleksander Luria Information-Processing Theory

The work of Russian neuropsychologist Aleksander Luria (1966a, 1966b, 1973, 1977, 1980) gave the way for information processing approach. The mechanism by which the information is processed is the main focus of the approach. Two kinds of information-processing style have been identified, i.e., simultaneous processing and successive processing. In simultaneous processing, information is incorporated and produced at once and as a whole for example, when we visit any exhibition, the information communicated by each displayed item, is processed in a manner that can be called simultaneous process. In successive processing, each bit of information is individually processed but in sequence. Kaufman Assessment Battery for children (Kaufman & Kaufman, 1983a, 1983b) is based on information-processing style presented by Luria.

Triarchic Theory of Intelligence

Robert Sternberg (1985b, 1986, and 1994) proposed another information-processing approach to study intelligence. According to Sternberg "the essence of intelligence is that it provides a mean to govern ourselves so that our thoughts and actions

are organized, coherent, and responsive to both our internally driven needs and to the needs of the environment" (Sternberg, 1986, p. 141). The Triarchic theory includes both Spearman's 'g' and information processing components. The theory recognized three aspects i.e., the componential (analytical) subtheory, experiential (creative) subtheory, and contextual (practical) subtheory.

Componential subtheory

According to Sternberg, componential subtheory is determined by three sets of mental processes these are, Meta components, Performance components and Knowledge-acquisition. Meta components are involved in planning what an individual is going to do, then monitor one's actions, steps required to solve a problem and finally evaluate solutions to different problems. Whereas, Performance components manages instructions of Meta components. One of the most interesting features of performance components are those found in inductive reasoning. This reasoning is required in series completion tasks and analogies (e.g. A is to B as Y is to ?). Sternberg (1990, believes that identifying these performance components can provide insight into the nature of 'g'. The last component Knowledge-acquisition, is involved in "learning how to do something in the first place (Sternberg, 1994, p. 221)

Experiential subtheory

Sternberg has examined the relationship between one's experience and intelligence, in experiential intelligence. He recognized two factors important in intelligent behavior. First, is the ability to deal effectively with novel situations and second, is the ability to handle familiar tasks automatically without any effort.

Contextual subtheory

While illustrating contextual intelligence, Sternberg has defined intelligence as a culturally distinct concept. However, dissimilarity exists in behaviors considered intelligent in different cultures. Sternberg (1985) claims that common intelligence tests emphasis too much on performance rather than individual's prior knowledge and as such fails to assess knowledge acquiring skills. According to him, intelligent behavior is directed towards three behavioral goals i.e., adaptation to an environment, Shaping of an

environment and selection of an environment. Sternberg also believes that, all the three components are essential for developing an intelligent test.

A brief review of different theories of intelligence based on three different approaches makes obvious that these theoretical framework were developed in an era (mainly in the early period of twentieth century) wherein a lot of importance was given to the testing and measurement of intelligence.

The work of Charles Spearman's (1904, 1923) gained wide popularity. In particular, statistical procedure introduced by him "factor analysis" attained the status of most wanted formula with researcher interested in examining correlations among tests measuring precise mental abilities. His findings also identified an important factor common with all cognitive abilities, which is called general mental ability 'g'. Thurstone (1938, 1955) presented a different view of structure of intellect, consisting of seven distinct factors called primary mental abilities. These factors were: verbal comprehension, word fluency, number, space, associative memory, perceptual speed and inductive reasoning. Thurstone acknowledged the presence of 'g' as a factor of higher order. Cattell (1971, 1987) proposed a theory based on factor analysis which suggests that intelligence consists of two major factors, fluid intelligence (Gf) and crystallized intelligence (Gc). Fluid intelligence has been identified as relatively culture free whereas crystallized intelligence is highly culture dependent. Horn (1985) later extended the work of Cattell and also proposed that 'g' is composed of fluid intelligence (Gf) and crystallized intelligence (Gc).

To understand difference between Gf and Gc is important for a test developer. Crystallized ability can be measured with the help of items demanding knowledge and skills that is expected of individuals of a particular age and culture. Whereas, Fluid ability is dependent on an individual's innate potential that is dependent on psychological and biological factors than on formal schooling. As children grow older and gain a variety of experiences in life particularly at school, these abilities become more and more distinguished. It is important to note that most of the modern intelligence tests are aimed to assess the 'g' factors i.e., fluid and crystallized ability, reasoning ability, the ability to perceive relations and draw out correlations, which are used in managing different tasks (Gregory, 2004; Riaz, 2008).

Importance of Verbal Factor in Testing

A study of the available literature on Intelligence Tests, reveals that almost all intelligence tests such as WAIS & WISC, makes an attempt to measure a subject's current intellectual capacities. Main focus of intelligence test is not to assess, subjects' cognitive abilities; nor his / her educational, vocational or other competencies rather the results helps in determining the important factor that is called intelligent behavior.

Intelligence has been commonly defined as a distinctive ability, such as the ability to reason abstractly, to learn, or to adapt. Each of these abilities distinguishes various human behaviors which can be declared as intelligent in one way or another. Therefore, intelligence has been recognized as a factor that is multifaceted. Therefore, intelligence tests never identify a particular ability in an individual rather it attempts to measure an overall competency or global capacity, which enables an individual to comprehend the world and to deal effectively with its challenges.

Another important approach mentioned by Gregory (1996), to understand a construct like intelligence is to study its popular meaning. Although, such approach apparently appears to be not scientific but is very useful in understanding the concept of intelligence. In our daily life "Words" have a common meaning that helps us in understanding daily matters. Sternberg, Conway, Ketron, and Bernstein (1981) conducted a series of studies while using this approach to investigate concept of intelligence as held by American adults. In these studies both laypersons and experts (mainly academic psychologists) were asked to rate the behavior which in their opinion is important for an intelligent person.

The results of studies revealed that there was a similarity between the views of experts and that of laypersons regarding the origin of intelligence. In order of importance, expert considered verbal intelligence, problem-solving ability, and practical intelligence as crucial to intelligence. Whereas, layperson regarded practical problem-solving ability, verbal ability, and social competence as important components of intelligence. However, it is pertinent to note that experts placed more emphasis upon verbal ability than any other factor while determining an individuals' intelligence.

Thus, it can be concluded that verbal ability alongwith problem solving capacity carries much significance in the measurement of intelligence. Therefore, both factors are being considered as an important aspect of most of the intelligence tests being used presently. Wechsler scales of measurement of intelligence can be taken as an example that includes these factors in their test blue print.

Importance of Theory While Developing Intelligence Test

The contents of any intelligence test largely depend on the author understanding of intelligence. Therefore, while developing a new test, the developer has to decide at the onset about the theory on which his/her test will be based and also to identify areas of intelligent behavior that will be assessed by a particular approach.

It has been noted that in Pakistan, most of the organizations providing testing services are using abundantly Western tests by translating them in Urdu and bringing minor changes/adaptation. However, no serious effort has been made so far to develop and standardize indigenous tests of intelligence (Riaz, 2008). All intelligence tests to a certain degree reflect the culture in which they were developed and subsequently be used. We can also say that all intelligence test are culture loaded. Culture loading is defined as the extent to which a test incorporates the vocabulary, concepts, traditions, knowledge and feelings associated with a particular culture (Benson, 2003).

Therefore, it is also considered relevant to study intelligence in social context:

Intelligence in Cultural Background

It is generally understood that human intelligence exists as a shadow rather than a result of social and cultural circumstances. An individual who is familiar with a particular social milieu will definitely experience difficulties when tested in an environment not familiar to him/her. Reasoning is not only deep-rooted in the mental image through physical perception but even more so is due to social background. According to Vygotsky (1998), intelligence is a special human characteristic and is arranged in accordance with the principle of social thoughts. Intelligence in isolation to social context is meaningless and superfluous. Sternberg (1985) has underlined that intelligence is unyielding and goal oriented behavior that consists of two general skills; the ability to deal with original tasks and to learn from experience. Intelligence depends on attaining skills for getting

information and to solve problems. Intelligence cannot be understood outside a socio-cultural context. What may be relevant in one culture may not be in the other culture.

After understanding relationship between intelligence and culture it is also important to analyze relationship of intelligence with education as the researcher intends to develop an intelligence test for population involved in acquiring education at various levels. Whether, a person with higher level of intelligence is likely to achieve high grades or other wise. Such type of questions can best be answered by understanding role of intelligence in education.

Relationship between Intelligence and education

Gottfredson (1997) was of the view that, "Intelligence as measure by IQ tests is the single most effective predictor known of individual performance at school and on the job. It also predicts many other aspects of well-being,, including a persons chances of divorcing, dropping out of high school, being unemployed or having illegitimate children".

Intelligence tests are predictors of educational achievement (Cattel & Butchere, 1968; Lavin 1965, 1965; Tylor, 1965, Jensen, 1980). The average correlation between general mental tests and educational achievement measures is found to be about, .50. Intelligence and achievement measures are closely related but not equivalent psychologically. Furthermore, General ability and achievement are thought to differentiate with age through childhood and adolescence. The best evidence that supports this claim is supported by Anastasi, 1970; that ability and achievement differentiate as a result of education.

In school years, correlation between intelligence measure taken at one point in time and achievement measures taken at a later point in time tend to be higher than when the measures are taken in the reversed time – order (Crano, 1974; Crano, Denny, & Campbell, 1972). The theory of fluid and crystallized intelligence (Cattell, 1963, 1971; Horn 1976, 1978) suggests that, fluid ability is used by the individual in learning experiences that includes formal education, to produce verbal-crystallized ability. The growth curve for fluid ability appears earlier than the crystallized ability curve during the childhood period and later on.

The more general tests, of both fluid and crystallized abilities are highly correlated with marginal skills and abilities (Guttman, 1965-1969, Jensen, 1970; Snow, 1980b). Although, fluid and crystallized abilities are at times difficult to distinguish, but verbal crystallized ability relates more to achievement in relatively familiar instructional setting (e.g., lecture, recitation reading, discussion, etc), whereas fluid analytic ability relates to achievement where problem solving or adaptation to unfamiliar instruction methods and material is involved (Snow, 1981).

Therefore, it can be concluded that relationship between intellectual measures and educational achievement are also dependent on environmental variables. Furthermore, it is generally assumed that there is a strong relationship between intelligence and academic performance i.e., the higher the intelligence the better is academic performance. Therefore, it is considered relevant to analyze this assumption in a pragmatic manner (Anastasi, 1988, Anastasi & Urbina, 2012; Deary & Johnson, 2010).

Intelligence and Academic Performance

Intelligence concentrates mainly on the individual's cognitive processes when analyzed from academic perspective. In addition, it is also believed that an individual who is capable of meeting the challenges of the academic career can also meet the challenges of the practical world. Nevertheless, real world challenges also require certain strategies, styles, and skills that may not necessarily be dependent on academic performance. Consequently, it is possible that an illiterate person to be intelligent thus demonstrates in practical life all the characteristics of an intelligent behavior, like motor mechanic or plumber etc.

It is also commonly believed that intelligence is a natural capacity that helps in the performance of various cognitive abilities. It is an innate capacity for making achievement in areas such as abstract reasoning, verbal analysis, creative expression, quantification, Visio-spatial organizations, and so on. Nevertheless, environmental factors also contribute considerably in the development of these capabilities. Consequently, it is logical to understand that being gifted in one area of intellectual performance means being equally proficient in other area.

On the contrary, we normally observe that most of us are not equally gifted in all aspects of intellectual behavior. Furthermore, even if we have similar opportunities to develop different abilities even then we do not develop these abilities uniformly. The fact remains that most of us can perform only on those complex levels for which we have the potentials. Thus we do not function equally in all the tasks that are set to assess intelligence. The concept of generality of intelligence across tasks and across situations is not substantiated; it is more of a theoretical assumption rather than a practical reality.

Sternberg (1983) believes in the existence of multiple intelligence. According to him an individual possessing a lower level of IQ, may exhibit a higher level of cognitive abilities particularly in the non-academic setting. Whereas, individuals with a higher level of IQ may display less effective cognitive abilities in non-academic matters. The main reason for it is environmental factor which helps in utilizing different levels of the intelligence. Hence, IQ scores do not predict social behavior, but only a level of ability (Anastasi, 1988 & Anastasi & Urbina, 2012).

Brief appraisal of various psychological issues associated with intelligence testing revealed that study of intelligence in isolation to cultural background is worthless. Intelligence cannot be studied outside a socio-cultural background. What is relevant in one culture may not be relevant in the other. Furthermore, concept of generalization of intelligence across various tasks and situations is not supported and is considered as theoretical assumption rather than a practical reality. Sternberg (1983), believes in the existence of multiple intelligence. He believed that the individual possessing a lower IQ may exhibit a higher level of cognitive abilities in a non-academic setting and one with higher IQ may display lesser cognitive abilities in non-academic matters such as motor – mechanics, electrician etc.

Experts also hypothesized that intelligence tests are predictors of educational achievements. A positive correlation is found between general intelligence tests and educational achievement tests. General ability and achievement also differ with age particularly during childhood and adolescence. In addition, correlation between intelligence measures taken during school years and achievement measures taken at a later point in time tend to be higher than when the measures are taken in the reversed order. It has also been argued that verbal crystallized ability relates more to achievement, such as familiar instructional setting (e.g. lecture, recitation reading, discussion, etc.). On the other

hand fluid analytic ability relates to achievement where problem solving or adaptation to unfamiliar instruction methods and material is involved (Snow, 1981).

Intelligence Testing without considering cultural background has always been a point of concern for the experts since 1910 (Anastasi, 1982). Rapid industrialization forced excessive usage of Intelligence Testing particularly in personnel selection. However, such employment of intelligence testing has also earned a lot of criticism primarily due ignoring role of cultural differences in testing. For example a test developed for American culture may not be applicable for an Asian or African culture.

It is important to realize that each culture support and encourage certain behavior that commensurate with local norms and values. Therefore, when an individual makes an attempt on a test that was developed for another culture, any difference found in performance is likely to become cultural disadvantage.

Therefore, the researcher found it relevant to develop better understanding on the role of cultural background in testing particularly limitations in cross-cultural testing. For that reason, in ensuing paragraphs we will examine some limitations that are being experienced in cross-cultural testing:

Intelligence Testing Across Cultures

According to Vernon (1979) "access to education, freedom from physical disabilities, familiarity with test contents, freedom from associated test anxiety along with skills involved for making assessment on a test are considered important while making comparison between cultures". It is important to realize here that even though we are aware of factors (as mentioned by Vernon 1979) that plays important role but still cross cultural comparison particularly between advanced and third world countries, is difficult to achieve in reality. Therefore, comparison made between different cultures with the help of intelligence test is uncertain. Nevertheless, accurate measurement within cultures is possible to achieve if tests are developed properly while keeping cultural preferences in view.

In addition, lack of different perceptual experience, limited linguistic motivation, lack of interest in formal education, limited exposure to traveling, television or books,

little schooling and emphasis on rote learning in school are the most important environmental factor which Vernon (1979) considered to effect intelligence test scores when cross cultural comparison is attempted. Whereas, if a test is developed for a particular culture it is ought to have meaningful items, familiar task and comprehensible instructions which facilitates accurate inferences. If examiner understand the nature of the test and its purpose, it is then possible to test intelligence with some degree of accuracy.

Vernon (1969-79) also discussed numerous problems that are likely to influence negatively tests results when measuring intelligence cross-culturally. These problems affect the test scores in such a way that accurate measurement of intelligence is difficult to achieve. These problems include unfamiliarity with the test situation, lack of motivation, anxiety, excitement and suspicion of the tester when the psychologist is of different race. In some cultures, there may be difficulties with particular types of items or materials. Lack of test sophistication is another factor that adversely affects the test scores.

It is also important to note that every culture and its subcultures are recognized by some peculiar abilities and behavior of its inhabitants whereas; some other behavioral characteristics are discouraged by them. For example if a test is developed keeping in view characteristics of an African culture then a person belonging to that culture is likely to stand out in such a test. On the other hand, same test will yield different results if individuals of different cultural background are tested on it such as Pakistani youth. Therefore, it is considered important that test developed for a specific culture be administered on individuals of that cultural background and practice of using foreign tests on local population be discouraged. According to Bruner (in Gillham, 1975) maintains that: The culture-free test is the intelligence-free test, for intelligence is a cultural concept. Similarly, Gillham (1975) argues that any attempts to 'define' intelligence which doesn't involve identifying 'specially valued cultural attainments' must fail. The concept of intelligence derives its meaning only within a particular cultural and social context. Paunonen & Ashton (1998) found significant differences between the African, Asian, and Western groups on the one hand and the Minorities and Western Reference groups on the other hand. According to Anastasi & Urbina, (2012), it is important for a test examiner to have knowledge of other cultures so that he/she can understand cultural effects on the behavioral development of an individual in order to understand probable effects of such differences on the examinee's response on tests. Variations in the person's self-concept, general world view, self introspection and habits of solving problems individually or in a

group are examples of differential testing behavior being observed when a test is administered on individuals not belonging to same culture for which the test has been developed.

The researcher of present study is interested in developing an indigenous intelligence test primarily consisting of verbal items as verbal ability alongwith problem solving capacity, as it carries much significance in the measurement of intelligence. Furthermore, due this very reason, both these factors are found abundantly in most of the intelligence tests being used presently. Wechsler scales of measurement of intelligence can be taken as an example where similar factors are found in abundance.

It is also interesting to mention here that a glimpse on type of intelligence tests being employed in developed societies for measurement of intelligence reveals that Group Intelligence Tests are the most favored one. The list of such group tests is mentioned in the Mental Measurements Yearbook (Mitchell, 1985; Conoley & Kramer, 1989, 1992) and the Test Critiques series (Keyser & Sweetland, 1984-1988). Therefore, keeping in view the importance and popularity presently enjoyed by group intelligence tests it was considered much relevant to have a brief account of the nature and functioning of group intelligence tests.

Group Tests of Intelligence

The group tests of intelligence are based on the Spearman's 'g' factor primarily as they have the ability to demonstrate an individual's performance over a wide variety of cognitive tasks. This view in most of the group intelligence tests has also influenced selection of contents, scoring strategy and interpretation of results. It is pertinent to mention here that authors of such tests constructed items found capable of measuring variety of abilities, such as vocabulary, analogies, arithmetical reasoning, etc. The purpose of including variety of items is due to the fact that it facilitates a comprehensive sampling of the general ability. Scores on the several parts were added to yield an overall measure of mental ability, expressed as a mental age and an IQ.

Group tests make greater demands in understanding written and spoken language than do individual tests and thus may place a subject at a disadvantage who does not belong to a foreign language background or is deficient in reading skills. It is also true

that group tests are always time limited and may thus have an element of speediness. However, almost all the group tests have been designed to minimize the effect of speed on performance.

In addition to Group Tests for measuring Intelligence there are Multilevel Group Intelligence Tests available. A short appraisal of the same is as under:

Multilevel Group Intelligence Tests

The rationale underlying the construction of multilevel intelligence test is to compare intellectual growth over several years. The Stanford-Binet and Wechsler tests are individually administered tests, but more extensively used tests are the group administered tests such as the Otis-Lenon, Kuhlmann-Anderson Lorge-Thorndike, Hannon-Nalson, and California Test of Mental Maturity.

The Otis-Lennon Mental Ability Tests (Otis and Lennon, 1967) are a revision of the earlier tests in the Otis series. The Otis Self Administering Tests of Mental Ability and the Otis Quick-Scoring Mental Ability Test. Like its predecessors, the Otis-Lennon is composed of a variety of items developed to measure general mental ability.

The Kuhlmann-Anderson Tests (1960-63) and the Kuhlmann-Finch Scholastic Aptitude Tests (1953-56) are both modern adaptation of intelligence tests devised by Fredrick Kuhlmann many years ago.

The Multi- Level Edition of the Cognitive Ability Tests designed for grades 3-12 consists of three parallel batteries – Verbal, Quantitative, and Nonverbal. The Verbal Battery contains vocabulary, sentence completion, verbal classification, and verbal analogies subtests. The Quantitative Battery includes subtests of quantitative relations, number series and equation building assess designed to the ability to work with number and other quantitative symbols.

The Henmon-Nelson Test of Mental Ability revised by (Nelson, Lamki and French (1973) cover four grade levels: grade 3-6 6-9 and 9-12 (Form 1) and Kindergarten through grade 2 (Primary Battery). A college level edition of the tests is also available. Omnibus

format including items on scrabbled words, verbal analogies, verbal classification, verbal inference, number series, arithmetic reasoning, figure analogies and following directions.

Group Tests of high level intelligence for general reasoning ability, discrimination among selected high ability group was developed by Hehle (1989). Test consisted of two sub-tests, verbal and numerical plus diagrammatic.

Canadian Cognitive Ability Test was developed by Costantino, (1985) for assessing the development of cognitive abilities related to verbal, quantitative and non-verbal reasoning and problem solving.

Verbal Reasoning Test for measuring the combination of mental abilities was developed by Hess, (1985) that correlates highly with achievement in scholastic tasks.

MD5 Mental Ability Test was developed by Deniels, (1989) to assess mental ability quickly and over a wide range of educational and ability levels, such as, in staff selection, placement and counseling. It consists of 57 items of vocabulary and mathematical skill.

Schaie-Thurston Adult Ability Test was developed by Schaie, (1985) to measure the mental abilities of the adults age 22 and above. The test consists of many items of Thurston Primary Mental Ability Test Form 11-17. The test consists of 7 scales, Recognition, Vocabulary, Figure rotation, Letter series, Number addition, Word fluency, Object rotation and Word series.

Thorndike, Hagen, and Sattler (1987) revised Stanford-Binet Intelligence Scale, (Fourth edition) with a purpose to measure cognitive abilities that provides analyses of the pattern as well as the overall level of individual's cognitive development. This test was revised for ages 2 to adults.

Structure of Intellect Learning Abilities Test was developed by Meeker, Meeker and Roid (1985) to assess a wide variety of cognitive abilities or factors of intelligence in children and adults.

Kaufman & Kaufman (1993) developed an Adolescent and Adult Intelligence Test (KAIT). The test is developed for individuals between age range 11 to 85 years and is administered individually. The test provides Fluid, Crystallized and Composite IQs having a standard score mean of 100 and Standard Deviation 15. The test has six subtests, three each to measure Fluid and Crystallized intelligence.

After making a brief review of the nature and functioning of some popular Group Intelligence Tests and Multilevel Group Intelligence Tests it reveals that both types of group tests are being used excessively to measure diverse attributes of human intelligence with the help of different strategies incorporated in these tests. The measuring strategies included items of vocabulary, analogies, arithmetical reasoning etc. In group tests emphasis is more on understanding written and spoken language as compared to individual tests.

Multilevel intelligence tests are also extensively being used to compare intellectual growth over several years. Tests like Otis-Lenon, Kuhlmann-Anderson Lorge-Thorndike, Hannon-Nalson and California Test of Mental Maturity are some commonly used tests. These tests were developed to measure various fragments of human intelligence. These tests differ to each other in their format and structure of items but primary rational behind construction of each test in the measurement of human Intelligence.

After a brief analysis of Group Intelligence Testing currently being used by developed countries, it is also considered relevant to have a critical review of research work being carried out in Pakistan in the field of Testing and Measurement. The same is mentioned in the subsequent paragraphs.

Psychometric Research in Pakistan

In Pakistan also utilization of psychometrics has attained significance particularly in organization involved in personnel selection and in clinical setting. In this regard services of professional psychologists are being acquired excessively. Test development programs are also being designed to meet the demands and in this regard,. Achievement Tests, Personality Evaluation Tests, Aptitude testing, and Intelligence Testing are being utilized the most.

A brief review of the work carried out by local Psychologists in the field of Testing and Measurement is as under:

An abbreviated version of Wallach-Kogan CreatiPVITy Test, AH5 Verbal Part and Standard Progressive Matrices was developed by Ansari (1976). It was found that while the intelligence is equally related to the achievement test of Lower and Higher Cognitive objectives, the CreatiPVITy measures show significantly higher correlations with achievement test of Higher Cognitive Objectives. Effects of bilingualism in the performance of Pakistani school girls in the tests of verbal intelligence and reasoning was assessed by Hasan (1981). The results indicated that bilingualism was significantly related to poor performance in verbal intelligence and reasoning. A computer model for statistical analysis of progressive matrices was developed by Zoofashan (1982). The objective of the project was to develop a computer model for a psychologist with no or a little 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.

Validity of Catell's Culture Fair Test of Intelligence (CFTT) for Pakistani children was studied by Ain (1985). It was found that while CFIT scores are unrelated with age, they strongly related with grade, achievement test and teacher rating. Retest reliability was also found to be significantly high. Thirteen Piagetian tasks were tried out on a sample of 360 primary school children from all over Pakistan by Israr (1985). The findings show that grade 1 children are at early concrete-operational stage and grade 3 children are at mid concrete operational stage while grade 5 children are at the late concrete operational stage.

Raven's Standard Progressive Matrices Test was administered to 300 students by Ismail and Mahmood (1986) to study the effect of sex and social class. A significant difference was found between the performances of three different social classes. No significant difference was found between males and females. The effects of immediate and delayed knowledge of results on subsequent performance in verbal learning were studied by Kausar and Sheikh (1986). California Psychological Inventory (CPI) was adapted and translated in Urdu by Ahmed (1987). After administration of Urdu and English version on bilingual subject, the results showed sufficient similarity between the two versions. The overall positive psychometric evaluation posed sufficient confidence on CPI and suggests further use of test in Pakistan. Relationship between ideational fluency

and intelligence among academically gifted and average students was investigated by Khan (1987).

Raven's Standard Progressive Matrices for urban and rural school children was validated by Ansari and Iftikhar (1988). The results showed that RSPM is a useful test of intellectual performance for the urban school children but their utility for the rural children is not so certain. Imam and Munaf (1988) administered Raven's Standard Progressive Matrices on 66 grade 5 female students. A significant difference in intellectual performance was found among the first, second and third born children. The differences in some personality variables and intelligence, which could be attributed to the difference in birth orders, were investigated by Najam, Andrabbi, Malik and Ghaznavi (1990). With regard to performance on the test, the youngest scored highest on the test of mental ability.

A Group Verbal Intelligence Test in Urdu for High School students was developed by Hussain (1992) with an aim to facilitate the students of those high schools where the medium of instruction is Urdu. The test comprised of two subtests: Vocabulary Test and Numerical Reasoning Test. Validity indices for Vocabulary Test and Numerical Reasoning Test were .82 and .58 respectively. The reliability indices for Vocabulary Test and Numerical Reasoning Test were .86 and .82 respectively. The reliability index of the total test was .88. Percentile norms were computed for grade 10 students of Rawalpindi City. Syed (1993) developed a non verbal test of intelligence for Pakistani urban primary school children. This test comprises of two subtests: Block Design Test and Picture Completion Test. The validity indices of the two tests were .89 and .89 respectively and were significant at .01 level. Naheed (1993) has also developed a verbal intelligence test in Urdu for Pakistan Urban Primary School Children between the age range of 5 to 11 years. This test consisted of two subjects: Vocabulary and Arithmetic. The validity indices for Vocabulary, Arithmetic and total test were .79, .56, and .85 respectively. The reliability indices for Vocabulary, Arithmetic and total test space were .84, 82 and .86 respectively.

Gardezi (1994) developed a non verbal intelligence test for adolescents of grade 10 in the age range of 15 to 17. The test comprised of four subjects viz Series, Analogies, Classification and Matrices. The reliability index was .82 for KR-20, and .77 for split half categories respectively. The construct validity of the test was computed by correlating it with a Group Verbal Intelligence Test in Urdu (GPVITU) developed by Hussain (1992),

which was .76. Percentile norms were developed separately for boys and girls of grade 10. Similarly, Gardezi (2001) has also developed a non verbal intelligence test for Pakistani youth of grade 12. The test comprised of five subtests viz., Series, Matrices, Analogies, Oddman out and Similarities. The estimated indices of reliability were .89 for KR-20, .85 for split half and .90 for test-retest respectively. Significant difference in the mean scores of three grade/age groups on the subtest and the full test, except similarities subtest were found. To determine the construct validity of the test, factor analysis and the convergent validity and discriminant validation approaches were adopted. Three types of norms Percentiles, T scores and the Deviation IQ were developed by administrating the test on sample of 1000 subjects' representative of Pakistani youth.

Group Verbal Intelligence Test in Urdu (SPVITU) for Pakistani youth having age range of 17 to 20 years and educational qualification of Grade 12 was constructed by Hussain (2001). The test comprises of four subtests viz., Vocabulary test, Verbal Reasoning test, Numerical Reasoning test and Information test. The reliability of the SPVITU was determined by Kuder Richardson method, Split-half method, and Test-retest method. The result suggests high reliability of the test both in term of internal consistency and temporal stability of the results. The validity of the test was determined through Construct validity, Concurrent validity. All the four indices established the evidence of high validity of the test. Percentile norms, Z-scores and T-scores were computed for grade 12 boys only.

Translation and Adaptation of Teacher Stress Inventory was done by Hanif and Pervez (2003). A Social Anxiety Scale and Social Confidence Scale was developed by Khalique, Khan, Janangir and Iqbal (2003). Bukhari and Kamal (2003) developed an Indigenous Emotional State Scale for Dermatological Patients. The scale measures the intensity of emotional disturbance among patients with psychosomatic skin diseases. Nausheen and Kamal (2003) developed and validated Familial Social Support Scale for Breast Cancer Patients which measures social support of the family as perceived by the women breast cancer patients.

Malik and Ismail (2005) developed a Social Support Scale in Urdu language, which could be better suited to Pakistani environment due to cultural and linguistic affinity than Western available scales. Likewise, a Problem Checklist for Pakistani Adolescents was developed by Rohail (2005). Translation, Adaptation and Validation of Children's

Action Tendency Scale was carried out by Zahid and Pervez (2009). Indigenous Depression Scale for Adolescent Schoolgirls was developed by Naz and Siddiqui (2010). Jabeen and Kausar (2010), developed an Indigenous Obsessive Compulsive Disorder Scale for Pakistan. Piri-Muridi scale was validated by Hassan and Kamal (2010).

Muazzan and Khalid (2011) developed and validated Disordered Eating Behavior Scale. Whereas, Loona and Kamal (2011) translated and adapted Disruptive Behavior Disorder Rating Scale. A research on Gender and Specific Learning Difficulties: A Double Disadvantage, was carried out by Irshad (2011) to examine specific learning difficulties among children and its implication for Psychological Functioning in terms of gender differences in psychological reactions to disability. Riaz, Yasien & Khanam (2011) translated and adapted Perceived Social Self Efficacy Scale in Urdu and estimated reliability of Urdu version of Perceived Social Self-Efficacy.

While analyzing work of local psychologists in the field of testing and measurement it is noticeable that most of the work has been dedicated to adaptation and translation of Intelligence Tests being commonly used in the developed world. In this regard translation work on CPI, adaptation & validation of CFIT, Wallach-Kogan CreatiPVITy Test, AH5 Verbal Part and Standard Progressive Matrices are significant. In addition, significant efforts were undertaken to study the effects of bilingualism in school girls on the test of verbal intelligence and reasoning, administration of Raven's Standard Progressive Matrices to study the effect of sex and social class, Piagetian tasks were tried out to determine concrete and mid concrete operational stages.

Kausar & Sheikh (1986) has determined the effect on immediate and delayed knowledge of results in verbal learning. Similarly, Khan (1987) studied relationship between ideational fluency and intelligence were studied among academically gifted and average students. In other study, difference in some personality variables and intelligence was analyzed.

Naheed (1993) developed group verbal tests in Urdu for high school and Primary school students. Non verbal intelligence tests for urban primary school children and grade 10 adolescents with certain subtests was constructed by Syed (1993).

Malik and Ismail (2005) developed a Social Support Scale in Urdu language, whereas Translation, Adaptation and Validation of Children's Action Tendency Scale was carried out by Zahid and Pervez (2009). A rare effort of developing an Indigenous Obsessive Compulsive Disorder Scale for Pakistan was done by Jabeen and Kausar (2010) and Piri-Muridi scale was also validated by Hassan and Kamal (2010).

Muazzan and Khalid (2011), developed and validated Disordered Eating Behavior Scale and Loona and Kamal (2011) translated and adapted Disruptive Behavior Disorder Rating Scale. Whereas, Riaz, Yasien & Khanam (2011) translated and adapted Perceived Social Self Efficacy Scale in Urdu .

As highlighted earlier in Pakistan, most of organizations and institutions providing testing services still use Western tests with minor modification/adaptation and translation but no serious effort has been made as yet to develop standardized indigenous tests of intelligence (Riaz, 2008). This dearth of research work in Pakistan is indicative of the fact that after 2001 e.g., Hussain, & Gardezi, (2001) no serious effort has been attempted to develop an indigenous intelligence test. This scarcity of research encouraged the researcher to undertake a research work aimed in developing an indigenous intelligence test for adolescents.

Brief appraisal of the research work carried out by Pakistani Psychologists reveal that a lot is desired in the field of testing and measurement particularly in developing indigenous assessment tools. For this very reason it is considered relevant to refresh our minds with the concept of indigenous development and the concept of validation.

Indigenous Development and Validation

Earlier mentioned brief review on Intelligence Testing transpires, that practically, every intelligence test attempts to measure the intellectual growth of an individual after he or she has completed certain level of educational curriculum and to predict how well the student is prepared for the next level in the educational hierarchy.

Furthermore, one of the most essential tasks of intelligence tests is to measure the individual differences. In addition, evaluation of intellectual levels, categorization of the children with reference to their abilities, the identification of the intellectually retarded on

the one hand and the gifted on the other, diagnosis of academic failures, and the selection of the applicants for different jobs are also areas where intelligence testing is being employed abundantly. Personnel selection programs in Pakistan also utilize Intelligence testing excessively. An extensive application of intelligence testing is found in selection and recruitment process of the Armed Forces of Pakistan, in Public Service Commission exams and also during the admission process of various professional institutes where facility of specialized education is afforded. Furthermore, in all such organizations/institutes English enjoys the status of an official language. Therefore, a need was felt to develop a group intelligence test in English that can be used as screening tool and also is capable to examine individual differences in Pakistani youth having age range between 17 to 25 years and hails from all four provinces of Pakistan including Azad Jammu Kashmir (AJK) and Federally Administered Tribal areas (FATA) areas..

Furthermore, as identified by Gottfredson (1997), "intelligence as measured by IQ tests is the single most effective predictor known to measure individual performance at school and on the job" p. 2-8. It also predicts many other aspects of well-being, including a person's chances of divorcing, dropping out of high school, being unemployed. According to Sternberg (1998), conventional IQ test will correlate 0.4 – 0.6 with school grades. However, a test that predicts performance with a correlation of 0.5, still accounts for only 25 per cent of the variation between the performances of different individuals. This leaves 75 per cent of the variation unexplained, so there must be more to school performance than IQ.

Language, in almost every learning situation has played an important role because being an essential medium for acquiring new information. Furthermore, a verbal test of intelligence repeats this process more strongly than a nonverbal procedure. In addition, the examinees feels more comfortable with verbal items as they are more familiar with verbal contents because of years of studies in school settings as compared to artificial tasks that are instilled in most of the nonverbal approaches. The results of research work carried out by Ahmed (1987) to adopt and translate California Psychological Inventory (CPI) revealed sufficient similarity between the Urdu and English version of the test on bilingual subject.

A test that is fair to all cultures is not likely to yield a meaningful ability measure. Bruners' diary (as cited by Gillham, 1975) the culture-free test is the intelligence-free test,

for intelligence is a cultural concept. Similarly, Gillham (1975) argues that any attempts to 'define' intelligence which doesn't involve identifying 'specially valued cultural attainments' must fail. The concept of intelligence derives its meaning only within a particular cultural and social context.

Another premise to give preference to verbal test over non-verbal is due to fact that, substitution of pictures or symbols for words increases the problems of test adaptation. Probably as a function of the education methods, words are a more effective medium for testing than are symbols or pictures. The research study conducted by Ansari and Iftikhar (1988) to validity of Raven's Standard Progressive Matrices on children of urban and rural school revealed that RSPM is a useful test for measuring intellectual performance but their utility for the rural children is not so certain.

Another reason for developing verbal test is that the adapted verbal ability tests turned out to be highly accurate in selection procedures. Although its development costs is substantial, the cost-effectiveness of the Verbal Analogies and Reading Comprehension tests are perhaps the highest of the I – D test series.

Typical intelligence tests designed for use with adolescents or adults, measure largely verbal abilities and to a smaller degree, numerical and other abstract symbols (Anastasi, 1990). As intelligence is not a single unitary ability, the term is commonly used to cover the combination of abilities required for survival and for advancement within a particular culture (Anastasi, 1986).

Keeping the above mentioned supportive literature for verbal intelligence test in view, the researcher decided to develop an objective type of Verbal intelligence test comprising of various subscales having verbal items. In this regard subscales of Vocabulary, Analogy, Arithmetic, Information, Comprehension and Similarity were identified for the intended verbal intelligence test. It is also important to remember here that, Individual tests of intelligence are used mainly as diagnostic tests in clinical settings, while group tests are used mainly for educational selection and research. This general understanding regarding group and individual intelligence tests strengthen the decision of the researcher for developing a group verbal intelligence test.

A brief review of the available literature on the proposed subscales has been deliberated upon in the following section.

Vocabulary Test

Vocabulary in all forms has long been recognized as the most accurate, stable and general measure of mental abilities or intelligence. Within the WAIS-R, Vocabulary has the highest correlation with Verbal IQ across all age groups included in the standardization sample. Many brief forms of the WAIS-R make use of Vocabulary as one of the best single measure of intelligence. House and Lewin believe that WAIS-R Vocabulary subtest is an excellent measure of the verbal ability (see Newmark, 1985 p. 45).

The ability to define words is not only one of the best measures of intelligence, but it is also the most stable and least deteriorating aspect of intelligence (Rapaport et al., 1968). Vocabulary tests are included in nearly all individual tests devised to measure verbal intelligence. As vocabulary scores provide an estimate of general verbal intelligence that is relatively stable and independent of deterioration, it can be used to evaluate the base line intelligence, that is, what a person's intellectual capacity probably was prior to an emotional illness, brain injury or trauma.

Another premise for including vocabulary test in IIT is that it is one of the most preferred estimate of intelligence when only a brief test is desired (Jenson, 1980). According to Cattell (1971), vocabulary test loads highly on fluid ability at an early age but gradually shifts over to crystallized ability. Vocabulary scores are not just dependent on educational experience, they do demand reasoning and mental processing of past.

Analogy

Most of the intelligence tests do include this type of items. One of the major reasons for making this type of items a preferred choice for most of the intelligence test developers is due to the fact that difficult level of such items can easily be maneuvered (Kaline, 1986). Furthermore, analogy type items makes it possible to present tangible relationship among objects and not mere ambiguity that has to be correlated. Thus, this type of items is considered suitable for measuring intelligence in individuals belonging from all age group. Analogy Test is a measure of ability to understand concepts framed in the words. It is aimed at the evaluation of the subject's ability to abstract or generalize and to think constructively. The analogy items are appropriate for the measurement of reasoning ability. The particular type of analogy item recommended and constructed for

the test is a single ended analogy in which the last term of the second pair are missing whereas both terms of first pair are provided. The examinee is asked to choose the missing term best completes the analogy from among four terms given as the options. This style of items is especially useful to provide a measure of reasoning. Another advantage of the analogy is that the content of the items may be varied in as many ways as desired. The words used in these items may come from history, geography, literature or any other content area. The items assess the subject's knowledge and ability to abstract and generalize relationships inherent in that knowledge.

The items for analogy test can be best drawn from a wide range of disciplines (i.e., history, geography, literature, science, etc.). The test also attempts to measure the subject's understanding of various concepts and verbal relations.

Arithmetic

This test is concerned with the ability to invent solutions to problems. Although, simple algebraic problems involve numbers, the main ability being measured in the test is not that of numerical calculation. The subject will be required to add, subtract, multiple, divide, or perform any two or more to these functions without any special aid. The skills needed to solve most of items of arithmetic test are not beyond those taught in the grade school or what an average adult would learn by himself during day-to-day interactions with various persons or situation. In order to solve the problem, the subject must concentrate, analyzed the nature of the problem, grape the principle by which each problem can be solved, remember the practical results until the problem is solved and finally the answer reported. The general reasoning factor also appears in the items concerning serial completion, in which the subject is required to supply the response in the patterned series of letter or digits. There is an element of discovery in all tests that measure the factor of general ability, hence Arithmetic test is also loaded on g-factor, measuring fluid and crystallized intelligence.

Information

The information subtest involves both intellective and non-intellective components including the ability to comprehend instructions, follow directions and provide a response. This subtest consists of items seeking a wide variety of information that an average adult

with average opportunity may be able to acquire for himself. In addition, items also included factual inquiry questions sampling the subject's general fund of knowledge. The item content will be selected keeping in view their ability to represent background information which is a result of a developmental exposure to the Pakistani culture and as such will bring within the common knowledge of most citizens. Information is believed to reflect the acquired knowledge that correlates with formal education. However, a pattern of general reading on diverse topics will also yield a high information score regardless of formal education.

Information along with Vocabulary is usually seen as one to the best measure of general ability among WAIS-R subtests and has the second highest correlation with the verbal IQ and the full scale IQ's. It consistently loads highly on the first verbal factor identified in most common factor and principal component factor analyses.

Comprehension

The Comprehension subtest is a collection of mix items that require explanation rather than mere factual knowledge. The easy questions stress common sense, whereas the more difficult questions require an understanding of social and cultural conventions (Gregory, 1996). Items that will be included in the test are such that an average adult may have an opportunity to answer them by himself or heard being discussed by others.

Comprehension would appear to be, in part, a measure of "social intelligence" in that many items tap the examinee understanding of social and cultural conventions. Sipps, Berry, and Lynch (1987) found that Comprehension scores were moderately related to measures of social intelligence on the California Personality Inventory. Of course, a high score signifies only that the examinee is knowledgeable about social and cultural conventions; choosing right action may or may not flow from this knowledge. It has also been proved through research that the comprehension items are among those which suffer least from practice effect (Wechsler, 1958). According to Matarazzo (1992), success on comprehension test seemingly depends on the possession of a certain amount of practical information and a general ability to evaluate one's past experience.

The questions included are of a sort that an average adult know the answers of these questions due exposure with them in his past life. The question involves no unusual

words, so that individuals of even limited education generally have little difficulty in understanding their content.

The Comprehension Test holds up well with age and when it begins to fall off, drops less than most of the other tests. It correlates best with Information and Vocabulary and least well with Digit Span and Object Assembly.

Similarity

In this subtest, the examinee is asked questions of the type "In what way are shirts and socks alike?" The Similarities subtest evaluates the examinee's ability to distinguish important from unimportant resemblances in objects, facts, and ideas. Indirectly, these questions assess the assimilation of the concept of likeness (Gregory, 1996). The examinee must also possess the ability to judge when a likeness is important rather than trivial. For example, "shirts" and "socks" are alike in that both begin with the letter "s" but this is not the essential similarity between these two items. The important similarity is that shirts and socks are both examples of a concept, namely, "clothes". As this example illustrates, Similarities can be thought of as a test of verbal concept formation.

According to Wechsler 1958, co-relational studies show that a well constructed similarities test is one of the most reliable measures of intellectual ability. Experience has also shown that while a certain degree of verbal comprehension is necessary for even minimal performance, however sheer word knowledge contributes as a minor factor only. It is possible to increase the difficulty of test items without including unfamiliar words. Wechsler also concluded that by answering questions of similarity test the subject supposedly manifests his ability to perceive the analogies (common elements) between two concepts.

The Similarity Test has several merits. It is easy to give and appears to have an interest appeal for the average adult. It is the kind of test which has been recognized by all investigators as containing a great amount of g. Over and above this, the test has certain qualitative features, the most important of which is the light that the type of response sheds upon the logical character of the subject's thinking processes. There is an obvious difference both as to maturity and as to level of thinking between the individual who says

that a banana and an orange are alike because they have a skin, and the individual who says that they are both fruit.

In Pakistan, most of organizations and institutions providing testing services still use Western tests with minor modification/adaptation and translation but no serious effort has been made as yet to develop standardized indigenous tests of intelligence (Riaz, 2008). All tests of intelligence, to a greater or lesser degree, reflect the culture in which they were devised and will be used. In other words intelligence tests differ in the extent to which they are culture loaded. One of the important aspects of an indigenous intelligence test is that how much a test incorporates the vocabulary, concepts, traditions, knowledge, and feelings associated with a particular culture.

The dearth of research work in Pakistan (Riaz, 2008), particularly in developing indigenous intelligence tests, encouraged the researcher to undertake a research work aimed in developing an indigenous Intelligence Test for adolescents of this country. The intended test will include items (as a measuring tool) culture friendly to Pakistani youth. Nevertheless, reasonable efforts have been put in by various organizations but much of the work has been limited in its scope therefore a lot is desired while developing indigenous tools of measurement in particular intelligence.

Keeping above mentioned literature review in view, the researcher decided to develop an indigenous intelligence test consisting of verbal items as verbal ability alongwith problem solving capacity carries much significance in the measurement of intelligence. Wechsler scales of measurement of intelligence can be taken as an example where such factors have been included in the test blue print.

After critically analyzing literature available on verbal intelligence testing, the researcher was in a position to start with the second stage of test development i.e. to develop Research Design for the proposed test. The same has been identified in the subsequent chapter.

RESEARCH DESIGN

The researcher decided to develop an objective verbal intelligence test in English language for Pakistani youth who are interested to apply in various professions i.e., in Armed Forces of Pakistan or appear in Public Service Commission exams or try to seek admission in professional institutes for obtaining specialized education. The new test comprises of subscales Vocabulary, Analogy, Arithmetic, Information, Comprehension and Similarity. The test will be used as a screening tool and a measure of general intelligence '*g*' of Pakistani youth.

Statement of the problem

To construct and standardize an indigenous group verbal intelligence test in simple English language to be used as a screening tool and to measure general intelligence level of youth in adolescence & early adulthood having age range from 17 to 25 years (Gregory 2004).

Objectives

The proposed test (IIT) was developed to meet following objectives:

1. To be used as a screening tool.
2. To measure general intelligence level of adolescents hailing from all the four provinces of the country (including AJK and FATA areas).
3. IIT will also help to identify individuals with higher intellectual endowments.

The Theoretical Approach

Psychometric approach was adopted for the development of the indigenous test as this method is based on assumption that intelligence can best be measured by gauging individual differences in performance on tests of intellectual abilities. Due this very reason

intelligence tests developed on the basis of this approach yields a single score or IQ. Thus suggesting that intelligence is a global trait, Spearman (1927), Burt (1971), Thurstone (1935, 1938, 1947), Guilford (1967, 1985), Vernon (1950, 1971), Raymond Cattell (1941, 1971. The items were specific for Pakistani culture.

Target population

To measure general intelligence level of adolescents having age range between 17 to 25 years and minimum educational qualification of intermediate (both Humanity and Science group).

Type of Test Material

The IIT consists of verbal items prepared separately for all the six sub-scales Vocabulary, Analogy, Arithmetic, Information, Comprehension and Similarity selected for the test. The researcher followed Psychometric approach for the development of test and its items as this method is based on assumption that intelligence can best be measured by gauging individual differences in performance on tests of intellectual abilities. Due this very reason intelligence tests developed on the basis of this approach yields a single score or IQ. Thus, suggesting that intelligence is a global trait. The format of the test items is similar to those tests items developed keeping in view Psychometric approach such as, tests developed by Spearman (1927), Burt (1971), Thurstone (1935, 1938, 1947), Guilford (1967, 1985), Vernon (1950, 1971), Raymond Cattell (1941, 1971. The items are specific for Pakistani culture.

One of the reasons for selecting verbal format of Intelligence test is due to its capability to measure largely verbal abilities and to a smaller degree; they also deal with numerical and other abstract symbols (Anastasi, 1990). As intelligence is not a single unitary ability, the term is commonly used to cover the combination of abilities required for survival and for advancement within a particular culture (Anastasi, 1986, 2012).

Another premise to give preference to verbal test over non-verbal is due to fact that, substitution of pictures or symbols for words increases the problems of test adaptation. Probably as a function of the education methods, words are a more effective medium for testing than are symbols or pictures. The research study conducted by Ansari

and Iftikhar (1988) to validity of Raven's Standard Progressive Matrices on children of urban and rural school revealed that RSPM is a useful test for measuring intellectual performance but their utility for the rural children is not so certain.

In addition language, in almost every learning situation has played an important role because of being an essential medium for acquiring new information. Furthermore, a verbal test of intelligence repeats this process more strongly than a nonverbal procedure. In addition, the examinees feels more comfortable with verbal items as they are more familiar with verbal contents because of years of studies in school settings as compared to artificial tasks that comprise most to the nonverbal approaches.

Another reason for developing verbal test is that the adapted verbal ability tests turned out to be highly accurate in selection procedures. Although its development costs is substantial, the cost-effectiveness of the Verbal Analogies and Reading Comprehension tests are perhaps the highest of the I – D test series.

Administration Procedure

The IIT is a group administered test. Individual tests of intelligence are used mainly as diagnostic tests in clinical settings, while group tests are used mainly for educational selection and research. This general understanding regarding group and individual intelligence tests strengthen the decision of the researcher for developing a group verbal intelligence test.

Standardization sample

The IIT was administered on a sample of 1700 subjects selected from all four provinces of Pakistan including AJK and FATA to establish reliability, validity and norms for the test.

The sample population included adolescents having age range between 17 to 25 years and minimum educational qualification of intermediate (both Humanity and Science group.

Stratified sampling technique was used to draw a sample of the population as this technique make possible representation of different subgroups in the population proportionately (Gregory, 2004; Riaz, 2008).

Norms Development

Percentile norms were developed for the Indigenous Verbal group Intelligence test as all standardized tests provide some form of within-group norms. With such norms, the individual's performance is evaluated in terms of performance in the comparable standardization group.

Reliability

Reliability of the test was established with the help of Kuder Richardson Reliability (KR-20), Split-half Reliability methods.

Validity

To determine validity of the test, constructed validity, concurrent validity, internal consistency methods were used to determine the validity of the present study.

Item Analysis

To select suitable items for the test, each item was subjected to following two types of item-analysis:

Qualitative item-analysis

The analysis was carried out by getting judgment from experts of the field. A pool of approximately 200 items was constructed with the help of reading material with which target population is generally exposed (e.g., course syllabi books identified by respective Provincial Educational Boards and Universities) and presented to professional psychologist working in various National and International educational institutes and in Armed Forces of Pakistan with a request to rate each item on a five point scale. Vocabulary subscale had 40 items, arithmetic 29 items, analogy 49 items, information 33

items, comprehension 33 items and similarity 16 items. The rating point 01 is indicative of most suitable item whereas 05 reflects least suitable item.

Quantitative approach

Data obtained from the first experimental try-out was subjected to appropriate statistical analysis with the help of ITEMAN software and SPSS. Items that are either too easy or too hard were eliminated at this stage and the most suitable items were retained. Selected item pool was tested on students of Psychology, studying in various institutes of the country. Sufficient time was allowed to students so as to attempt each item. The responses of students helped in improving doubtful items alongwith distracters. After making necessary correction items were arranged in an appropriate order of difficulty.

Try-outs

After the required modifications, the draft test was ready for experimental try-outs and as such was administered on a large representative sample of target population to establish Reliability, validity and to prepare norms for the indigenous test.

Scoring of the test

It was also decided to undertake computerized scoring for the test with the help of Optical Mark Reader (OMR) technology. The examinee was supposed to mark their responses with the help of a lead pencils by shading appropriate option on the computerized (OMR) answer sheet. Scoring was managed by scanning each sheet through computerized software OMR.

METHOD

Layout of the Test

The first step in the construction of any Intelligence Test is to present a test "blue print". The details of the layout for proposed Verbal Intelligence Test are mentioned below. The test blue print was prepared with the help of guidelines provided by Riaz, (2008):

Conceptual Foundation of the test

The researcher followed Psychometric approach for the development of test and its items as this method is based on assumption that intelligence can best be measured by gauging individual differences in performance on tests of intellectual abilities. The format of the test and its items was similar to those tests developed keeping in view Psychometric approach such as, tests developed by Spearman (1927), Burt (1971), Thurstone (1935, 1938, 1947), Guilford (1967, 1985), Vernon (1950, 1971), Raymond Cattell (1941, 1971. The items were specific to Pakistani culture.

Cattell (1971, 1987) has identified two major cognitive abilities i.e. fluid intelligence and crystallized intelligence. According to Cattell, fluid intelligence (Gf) consists of nonverbal, relatively culture-free, and independent of specific instruction (such as memory for digits). On the other hand, crystallized intelligence (Gc) consists of acquired skills and knowledge that are dependent on exposure to a specific culture as well as formal and informal education (for example vocabulary). Retrieval of information and application of general knowledge are considered as components of crystallized intelligence. Fluid abilities (Gf) drive the individual's ability to think and act quickly, solve novel problems, and utilize short-term memories. Fluid intelligence is also associated with physiology of an individual, and is thus considered relatively independent of education and acculturation (Horn, 1968).

Crystallized abilities (Gc) on the other hand, stem from learning and are reflected in tests of knowledge, general information, use of language (vocabulary) and a wide

variety of acquired skills (Horn & Cattell, 1967). Personality factors like motivation coupled with educational and cultural opportunities plays important role in its development.

Target Population

The target population for the proposed Intelligence Test was male and female students with minimum qualification of Intermediate level. The sample was selected from all the four provinces of Pakistan (including AJK and FATA) having age range between 17 – 25 years of age. Maximum qualification of the target population was Post Graduate level. The students with medium of instructions both in English and Urdu were eligible for the test.

Test Content

The proposed test was designed on the format of a Verbal test and in objective format. The objective format was selected primarily due to the reason that this type of test is free from personal bias of the examiner. On the other hand, essay type test when scored by two different examiners, often arrives at quite different results. Since, such difference in scoring is attributable to the person who is examining and grading the performance; therefore, they are called personal errors. It is important to note that, as personal bias of the examiner vary from time to time and from paper to paper therefore, it is essential to reduce it to a minimum level so as to ensure the reliability of the test. Furthermore, a test cannot be reliable unless it can be scored with a reasonable amount of objectivity.

Structure of Items

Multiple-choice items format was selected for the test because this format is considered most flexible and probably the most effective of the objective type items. A multiple choice item can be used to assess all those educational objectives that can be measured by a paper-and-pencil test. Multiple-choice items not only can assess knowledge, but also comprehension, application, interpretation, analysis or synthesis to arrive at the answer keyed correct (Gregory, 2004; Riaz, 2008).

First test draft

It was decided to construct items having information with which target population is most familiar. In this regard, books and literature on different subjects selected to be studied was syllabi books identified by respective Provincial Educational Boards and Universities. With this strategy, items constructed were comprehendible to students of the four provinces studying in English and Urdu medium government educational institutes. After laborious efforts items were constructed and presented to five psychologists having vast experience in the field Testing and Measurement for their expert advice on the format and contents of each item while utilizing content analysis procedure. They were also requested to grade each item on the basis of its perceptual complexity and face validity, on a five point scale. In this regard they are to grade each item on a scale of 1 to 5 wherein, 1 refers to the poorly constructed or least appropriate item and 5 for the most suitable or most appropriate item for the measurement of intelligence.

Item Evaluation through judgment method

To ensure the qualitative of test items, it was planned to put drafted items of the sub-scales to intensive item analysis. In the first phase it was planned to seek help of expert psychologists particularly those having extensive experience in the field of Testing and Measurement. Selected experts were requested to carry out content analysis of each item for its appropriateness or otherwise on a five point scale provided on a paper sheet. Furthermore, they were briefed that on five point scale, 01 stands for least appropriate item whereas 05 represents most appropriate item. Consequently, keeping expert opinion in view, items of all the sub-scales of Intelligence test identified with vague and difficult words were reworded and their order was rearranged as recommended to augment overall understanding and face validity of the test.

Scoring

Optical Mark Reader (OMR) format answer sheet was used for taking the test. The examinee marked their responses with the help of lead pencils by shading appropriate option on the computerized based (OMR) answer sheet. Scoring was managed by scanning each sheet through a scanner controlled with the help of computerized software OMR.

Stratified random sampling technique was selected to collect data for the main study. Government sector Colleges and Universities were selected from various cities of the respective province so that target population is true representative of the respective province. This technique provided an opportunity to choose the standardization sample from a population consisting of various subgroups. After selection of educational institute class rooms of the students studying in Intermediate level, Graduation level and Post Graduation level were visited with the help respective faculty and test instructions were imparted on each class with a request to attempt the test. With the help of this strategy each defined population can be classified in subgroups who share some common characteristics deemed important for test score.

Norms Development

Finally, it was decided to formulate Percentile norms for the main study. Percentile scores are expressed in terms of the percentage of persons in the standardization sample, whose score on the test falls below a certain raw score. Thus, percentiles represent ranking of examinees with regard to their relative standing within a group. Lower percentiles will indicate poor performance of examinees. Percentile scores are easy to compute and can be readily understood, even by technically untrained test users. In addition, percentile scores are universally applicable and as such are easily adapted to a wide range of tests. They can also be used effectively with persons of varying age levels (Anastasi & Urbina, 2012; Riaz, 2008).

Administration Procedure

The test was administered in a group form because, a group intelligence test allows for the quick and efficient testing of dozens or hundred of examinees at the same time. This strategy was selected due to advantages inherited by such testing. The most important advantage of group testing is that it employees a multiple-choice format. As a result the multiple-choice format can be quickly and objectively scored by an optical scanning device attached to a computer. Computer scanning also eliminates examiner errors and halo effects that may occur in the scoring of individual tests. Furthermore, Psychometricians gain easy access to item analysis and test data bank, so computer scoring promotes quick development and revision of group tests.

In the administration of group tests examiner plays a minimal role and is restricted just to reading instructions and enforcing time limits. In addition, minimal opportunity for one-on-one interaction between the examinee and the examiner is possible. Secondly, due limited role of examiner in group tests, each examinee will attempt each item in the same order. As the group tests is arranged in increasing order-of-difficulty, so examinees encounter easy items first and then proceed to higher levels of difficulty. Thus, maintaining motivation level of relatively weak examinee.

Group tests are generally standardized on large samples of subjects instead of limited number of carefully selected subjects as in case of individual tests. Hence group tests results can be easily generalized for the whole population for which it has been designed. Finally, a group test can still be administered to a single individual because the mode of administration (minimal examiner involvement) and the method of scoring (purely objective) make it possible to use such test in both setting.

Reliability of the IIT

The following methods were selected to determine the reliability of the Intelligence Test and its sub-tests:

1. Kuder Richardson Reliability (KR-20)
2. Split-half Reliability

Validity of the IIT

There are many types of evidences that can be used in to determine the validity of a test. However, construct validity, concurrent validity, grade differentiation and area differentiation were selected as a source for determining the validity of the present research.

Practical consideration

It was also decided that the length of the test should be such that it could be completed in 30 minutes time. It was also planned that test will be administered in all the four provinces with the help of trained professional Psychologists. As the researcher

himself is serving in Pak Navy as professional Psychologist therefore help of colleague Psychologists from the three Services posted at various Selection Centers located in different cities of the country were acquired for test administration. In addition, help of professional psychologists posted in the Inter Service Selection Boards (ISSB) situated at Kohat, Gujranwala, Karachi and Quetta was sorted to obtain data from these places. Efforts were made to reach Government sector Colleges and Universities located in different cities of each province so that target population can be considered true representative of the respective area. After selection of educational institute students studying in Intermediate level, Graduation level and Post Graduation level of respective college and universities were approached with the help of respective faculty member. Test instructions were imparted on the students to make them understand very purpose of the test and were requested to attempt the test in prescribed time limit.

Additionally, it is important to note that these Psychologists are professionally trained in administration of Intelligence and Personality tests therefore; no requirement of specific training arouse.

Scoring of the test was planned to be conducted through computer software. For this purpose answer sheets as per requirement of OMR technology were planned to be printed for the test. Interpretation of the test scores were conducted with the help of SPSS software. Standardized instructions were printed on each booklet and the copy of same was provided to each Psychologist nominated for test administration.

Item analysis

The computer based software ITEMAN (Crocker, & Algina, 1986) was used to carry out all the analysis of the data for determining the psychometric properties of the items. Item analysis was the major objective of the experimental tryout. Each item was analyzed to find out its effectiveness, considering the role it could play in developing the final version of the test. A good test item is one that is marked correctly by most of the high scorer on a particular test and answered incorrectly by most of the low scorers on a particular test, (Anastasi & Urbina, 2012; Gregory, 2004; Riaz 2008).

For item analyses of the Intelligence Test, two indices were identified to be calculated:

1. Index of item difficulty: To see the percentages of subjects who gave correct answer to the item.

2. Index of item discrimination: To see if the test discriminates between high and low achievers on the sub-test.

Item difficulty (p)

The item difficulty for a single test item is defined as the proportion of examinees in a sample who answered a particular item correctly. Theoretically, the value of an item-difficulty index can range between 0 (if no one chooses the correct answer) to 1 (if everyone got the answer correct). As suggested by Anastasi & Urbina, 2012; Riaz 2008, for a test having multiple-choice items with four alternatives the difficulty level of an item closer to .50 is recommended. Furthermore, an item with item-difficulty index closer to .50 helps in extracting maximum information about individual differences. Therefore, for most tests, the optimal average item difficulty recommended by experts is approximately .50 with individual items on the test ranging in difficulty from .30 to .70. However, final decision regarding item difficulty level rests on the user keeping in view specific testing goals (Gregory, 1996).

Item discrimination level (d)

An effective test item is one which discriminates between high scorers and low scorers on the entire test. Therefore, a test item can be considered ideal which is answered correctly by most of the high scorer on a particular test and answered incorrectly by most of the low scorers on a particular test (Anastasi & Urbina, 2012; Gregory, 1996, 2004; Riaz, 2008). It has also been suggested if the difficulty level of an item is closer to .50 then such an item is likely to have higher discrimination values.

If the total test scorers are normally distributed then in order to compute item discrimination the test constructor will select examinees from upper (U) (high achievers) and lower (L) (low achievers) groups. In a normal distribution of scores, 27% is considered the number to form the extreme groups. The item-discrimination index is based on a comparison of performance on a particular test item with performance of examinees

in the upper and lower regions of a distribution of continuous test scores. In other words, *d* is a measure of the difference between the proportions of high scorers and low scorers answering an item correctly. A high value of *d* shows that the number of high scorers answering the item correctly is higher, whereas a negative *d* value on a particular item suggests that low-scoring examinees are more likely to answer the item correctly than the high-scores.

Ebel (1965) has also recommended that an item having discrimination index equal or higher to .40 is discriminating very well and as such does not require any amendment or change in the item.

Keeping in view recommended limits for both item difficulty and discrimination level in view, researcher decided to keep average difficulty-level and discrimination-level of test items closer to .50 or equal or higher to .40 respectively.

Estimation of Average Time limit of the IIT

When suitable items were identified with the help of item analysis, then selected items for each subscale were arranged in ascending order of difficulty level, so as to get the final form of the sub-tests. In order to determine the time limit for the complete test, an additional study was carried out where complete test was administered to a sample of 50 candidates. The candidates were informed about the purpose of the test and were also told to complete the test as quickly as possible.

Development of norms

Nearly all standardized tests provide some form of within-group norms. With such norms, the individual's performance is evaluated in terms of the performance of the most nearly comparable standardization group, for example, comparing a child's raw score with that of children of the same chronological age or in the same college grade. The key in this process is to obtain samples representing a cross section of the target population. For this purpose Percentile norms were developed for the Indigenous Verbal Intelligence test.

Objectives of the Study

The construction of the proposed test commenced by identifying following objectives:

1. To develop a valid and indigenous Intelligence Test.
2. To establish the reliability of the test.
3. To ascertain the validity of the test.
4. To develop norms for Pakistani population with the help of the test.

The first and foremost important step in the construction of an intelligence test is to generated appropriate item pool for the test. Item writing has usually been considered as an art rather than a scientific endeavor. It is also important to realize that the quality of the test depends on the psychometric properties of the items. Therefore, the test constructor has to ensure that at every stage of test development items must in every way form a representative sample of the total domain for which the test is being devised.

Item Generation

Items of the test being the scoring unit demand a lot of efforts and attention in its construction. It was decided to include multiple-choice form of items as it is free from many of the weaknesses inherent in other forms of items. Multiple-choice item are considered as the most flexible and probably the most effective of the objective item types (Anastasi & Urbina, 2012; Riaz 2008). Such type of items permits a wide sampling of content and objectives in a short period of response time. They can elicit responses measuring all levels of cognitive skills from knowledge to evaluation. Since, there is only one correct response to a multiple-choice item, the difficulty and subjectivity of scoring that plague other form of test items are avoided.

Now we will discuss steps taken to develop items pool for six sub-scales of the Verbal Intelligence test.

Vocabulary Test

As intelligence is usually considered a multidimensional construct, consisting of several verbal abilities, the content of intelligence test will cover various dimensions or it may be measuring a single dimension such as vocabulary which has often been regarded by psychologists as a single effective indicator of intelligence. The source of items of Vocabulary Test was general English reading books recommended by respective Educational Board of respective province and as such sample population has exposure to the same. While constructing the test items, simple English words (Common noun) were selected from the identified English books randomly. The initially generated item pool consisted of 41 items for Vocabulary Test, which included only common nouns and adjectives.

Arithmetic

To generate the items of Arithmetic, the basic concepts and principles of arithmetic, such as addition, subtraction, multiplication, division, percentage etc, were selected from syllabi books recommended by respective provincial education boards and taught in government sector colleges, with which a student of intermediate level is well familiar. In the construction of items, main emphasis was laid upon perceptual clarity, comprehension and application rather than mere reproduction of factual knowledge. In other words, the items included in the Arithmetic Test evaluate ability of individuals to reason with numbers and deal with quantitative material in an intelligent way. The items initially generated for Arithmetic Test were 28 in number.

Analogy

The Analogy Test is a measure of ability to understand concepts framed in words. This test is designed to evaluate an individual's ability to think, extract and generalize rather than focusing on comprehension of vocabulary. The item format used in this test is an efficient single-ended analogy in which last term of second pair is missing. The examinee is asked to choose from four words, the word best completes the analogy. The correct word consisted of missing part of the last pair. This type of format is a useful measure of logical reasoning as it is slightly complex without being tricky. The words used in these items require general awareness about various professions, geography,

science and related context areas. This test aims at assessment of an individual's ability to abstract and generalize relationships inherent in their knowledge. A total of 65 initial item pool was generated for this test.

Information Test

Items for Information Test were selected from literature of general information covering different concepts, purely from the Pakistani culture. Efforts were made to select the items of information, which seem to be the part of common pool of experience of the population so that the generated items should load on the g factor. The initial pool of items generated for the test was 30.

Comprehension

Such items for Comprehension Test were selected that primarily involved common sense. Efforts were made to include such concepts in items that can tap the examinee's understanding of social and cultural principles. In addition, items included situations with which a Pakistani youth encounters on daily basis and as such involved certain amount of practical information and general ability to evaluate past experience. The initial pool of items generated for the test was 20.

Similarity

Items of Similarity subtest were selected from objects/things with which a Pakistani youth is exposed with on daily basis. Efforts were made to select items covering different conception, purely from the Pakistani culture. Efforts were also made to select the items, which seems to form a part of general pool of experience of the population so that the generated items should load on the g factor. The initial pool of items generated for the test was 21.

The next important step in line was to construct distracter for each item. In a multiple-choice test, there is only one correct or preferred answer to each item, the rest of the options or answers are called distracters. Efforts were made to prepare near to perfect items. A perfect multiple-choice item has two characteristics as mentioned by Riaz, (2008):

1. All the respondents who know the correct answer to that question will always choose the right option.

2. The examinees who do not know the answer will make a random choice among the given options. In other words, some examinees will get the answer correct by mere guessing. Furthermore, each distracter will be selected by an almost equal number of respondents.

Selection of Distracters

In an attempt to find the distracters of the items pertaining to six subtests of Intelligence Test, all the items of Vocabulary Test, Analogy Test, Arithmetic Test, Comprehension, Information and Similarities were given to students of Islamabad College for Boys G-10 (n=25) and to students of Bharia College Islamabad (n=25). This procedure was used to obtain good distracters from the wrong answers actually given by the students. Three plausible responses were selected as the distracters for each item of the six sub-tests. Correct answers to items were taken from Oxford English Dictionary, books of relevant subjects being taught at government sector colleges of all the provinces. The researcher also consulted some popular books on general information for instance, Current Affairs, Who is Who? And What is What?. An effort was made to change the serial position of the correct answer to each item randomly with a purpose to eliminate of guessing by the examinee.

An effort was also made to construct the items in such a manner that people who know the answer to the item always chooses the correct answer and the people who do not know the answer, chose randomly among the possible distracters. This means that if an examinee resorts to guess work then every distracter including the correct answer has equal chance of being selected as an answer.

After finalizing the format of the items of the six sub-tests, instructions were written for the whole test. Instructions were simple and brief whereas, use of examples was made to clarify the instructions.

Initial Editing and arrangement of the items

Once the items were ready in their initial form, they were presented to five Senior Psychologists for their expert advice and opinion. They were requested to grade each item on the basis of its perceptual complexity and face validity, on five point scale. There were requested to grade each item on a scale of 1 to 5, as scale 1 refers to the poorly constructed or least appropriate item and 5 refers to the properly constructed or most appropriate item for the measurement of intelligence. In the light of their views and grading, items of Vocabulary Test were reduced from 41 to 37, the items of Analogy were reduced from 65 to 48, items of Arithmetic Test were reduced from 28 to 25, items of Information Test were reduced from 30 to 28, items of Comprehension were reduced from 35 to 30 and items of Similarity Test were reduced from 21 to 15. The items of all the sub-tests were arranged in an approximate ascending order of difficulty.

Tryout Stages

After the set of items for each sub-test has been written, reviewed by subject experts and revised on the basis of their suggestions, the sub-tests were ready for experimental tryout on a sample representative of the target population with whom the final form of the test will be used. The various purposes of the try out were as follows:

1. To identify weak or defective items and to reveal needed improvements. More specifically, to identify ambiguous items, indeterminate items, nonfunctioning or implausible distracters, overly difficult and too easy items.

2. To determine internal consistency of each sub-scale.

3. To determine the difficulty of each individual item.

5. To determine the discriminating power of each individual item.

6. To provide data needed to determine the final selection of items.

7. To provide data needed to determine the time limits for the test.

First Tryout

The first experimental tryout for the proposed test was carried keeping in view following purposes:

1. To carryout Item analysis to determine the following:

   a. To study internal consistency of items of each sub-scale.

   b. To determine item difficulty.

   c. To determine item discrimination.

2. To undertake revision of sub-tests in the light of the above mentioned information.

Sample

The sample consisted of 200 students of Islamabad College for Boys G-10, Bahria College Islamabad, candidates appearing for PN Cadet Course at Peshawar, Rawalpindi and Karachi. This sample was selected simply because these candidates constituted a representative group of different areas of Pakistan fulfilling the predetermined criteria of having intermediate qualification and falling within age range of 17 – 25 years. Stratified sampling technique was used for the selection of candidates to have a balanced and more equally distributed group with reference to their various socioeconomic backgrounds.

Test Administration

First draft of the proposed six sub-tests was administered in five batches at each selected place, as mentioned above thus making a total of 200 students/candidates. Efforts were made to reduce test anxiety by presenting the test as a research instrument rather than an intelligence test. The candidates were seated in a comfortable atmosphere to minimize test anxiety. They were also assured of the confidentiality of the results. At the beginning of each administration, the instructions printed on each booklet, were clearly read out to the subjects. The candidates were instructed to attempt all the questions and not to leave

any question, as there was no time limit for taking the test. They were encouraged to ask questions, if any, regarding the instructions, examples and recording of their responses. After responding to certain queries of the students/candidates test was started.

Scoring and Analysis of Data

The computer based software ITEMAN was used to carry out all the analysis of the data for determining the psychometric properties of the items.

Item Analysis

Item analysis was the major objective of this experimental tryout. Each item was analyzed to find out its effectiveness, considering the role it could play in future revision and for developing the final version of the test.

For item analyses of the subscale of Intelligence Test, three indices were calculated:

1. Index of item difficulty: To see the percentages of subjects who gave correct answer to the item.

2. Index of item discrimination: To see if the test discriminates between high and low achievers on the sub-test.

3. Internal Consistency: To determine the homogeneity of each sub-scale, indices of internal consistency based on Pearson product-moment correlation were computed. Consequently, it was decided to retain only those items that were internally consistent ($p < .01$).

The ITEMAN software was used for carrying out Item Analysis. The sample was divided into two groups designated as high scorers and low scorers on the basis of the medians of the whole test. One group consists of high-scoring subjects falling in the upper bracket and the second low-scoring subjects falling in the lower bracket on the test. Comparisons were made for each item between high and low scores in terms of their true and false responses.

Time of the Test

When the statistically effective items were selected for every sub-test, they were arranged in ascending order of difficulty level, so as to get the final form of the sub-tests. In order to determine the time limit for the complete test, a separate study was carried out, in which the test was administered to a sample of 50 candidates appearing before Inter Service Selection Board. The candidates were informed about the purpose of the test and were told to complete the test as quickly as possible. With the help of stopwatch, the time taken by the first candidate (completing the test) was noted down, followed by the time taken by the last candidate to complete the test. The average time taken by these 50 candidates was calculated in order to decide the time limit of the test (Table 13).

Items are arranged in order to their difficulty, so that the examinees begin with relatively easier items and proceed to items of increasing difficulty. Such an arrangement tends to be helpful to motivate the examinees to work on the test with confidence (Riaz. 2008). It also helps the respondents to work on the test items suitable to their ability level without wasting their time on items that are beyond their ability level.

Main Study

The purpose of the main study was to determine the psychometric characteristics, that is, reliability, validity and norms of Intelligence Test and its sub-tests.

Sample

The sample for the main study consisted of 1669 candidates selected from various educational institutes of the four provinces of the country. The sample was selected by employing stratified sampling technique from various cities of four provinces of Pakistan. The sampling plan used in this study is given at Annexure-A.

Administration of the Test

The procedure of the test administration in this study was the same as was used in tryout studies, except that a time limit was allocated to the Intelligence Test. The time limit was also communicated to examinee at the beginning of the test, and was instructed to do their best to complete the test within the allotted time limit.

Reliability of the Test

The following methods were used to establish the reliability of the Intelligence Test and its sub-tests:

1. Kuder Richardson Reliability (KR-20)
2. Split-half Reliability

Kuder Richardson Reliability (KR-20)

Kuder Richardson Formula 20 (KR-20) for estimating reliability is generally assumed as the best technique to find out inter-item consistency of the intelligence tests. This method is based on the numbers of the items in the test and the average inter-correlation among the test items. Responses of the items are coded as 1 for right response and 0 for incorrect response. The results are presented in table 17.

Split-half Reliability

To determine the Split-half reliability, Intelligence Test was administered to the sample of main study. The items of the tests were divided into two equal half based on odd and even numbers. Since split-half reliability is determined by correlating the scores of two half tests, an estimate of the reliability of the full test was made by applying the Spearman-Brown formula (Table 16).

Validity of the Test

When constructing a psychological test, the most important question is to what extent the interpretation of the scores will be appropriate, meaningful and useful for the intended application of the results. Validity, one of the psychometric characteristic of the test, is the answer to this question. Validity refers to the appropriateness of the interpretations made from test scores and other evaluation results with regard to the particular use for a given group of individuals and not to the instrument itself.

Although there are many types of evidences that can be used in the process of test validation, constructed validity, concurrent validity, internal consistency, gender differentiation and area differentiation were selected for the present research. Construct validity involves the demonstration of the psychological characteristics of the variable measured by the test. Concurrent validity involves prediction of the future performance and estimation of the present performance of the candidates on the test scores with reference to some valued measure as the criterion. Internal consistency was determined by correlating the scores of subscales with each other and with Intelligence Test. Gender differentiation involves the differences in the performance of Ss on the test among the male and female students. Whereas, area differentiation involves the differences in the performance on the tests among the four provinces of Pakistan.

Construct Validity

The construct validity of a test is the extent to which the test may be said to measure a theoretical construct or trait. The goal of construct validity is to determine whether test scores provide a good measure of a specific construct. The process of construct explication provides a definition of the test construct in terms of concrete

behaviors. There are few methods of assessing to construct validity. The most basic method is to correlate scores on the test with scores on a number of other tests.

For this purpose of validity, internal consistency method was used wherein scores of all the sub-tests were correlated with each other and also with Intelligence Test. Obtained correlation coefficients (table 18) were another measure of construct validity.

Concurrent Validity

To determine the concurrent validity of an intelligence test, the criterion frequently employed is some index of academic achievement. Criterion selected to estimate the concurrent validity of Intelligence Test was the final examination marks which each candidate has taken in their last annual examination. Marks obtained in the last examination were converted into percentages for making the statistical calculations easy. Results are shown in table 19.

Other Statistical Analysis

To study the significance of differences between mean and Intelligence Test scores of various groups, t- test for independent groups was applied for the following comparisons:

1. Urban versus rural students
2. Economic status wise difference
3. Province wise difference
4. Male versus female students

Norms Development

Nearly all standardized tests provide some form of within-group norms. With such norms, the standardization group performance is evaluated, for example, comparing a child's raw score with that of children of same chronological age or in the same college grade. The key in this process is to obtain samples representing a cross section of the

target population. For this purpose Percentile norms were developed for the test. In addition, province wise norms were also developed to gauge true performance of candidates hailing from different provinces.

Procedure

Once the final version of the test was ready for future administration, a study was conducted for developing norms IIT for Pakistani youth having age range between 17 to 25 years. Development of Norms was as follows:

Sample

The sample comprised of 1669 students of grade 12 to 16 from various colleges selected from the six provinces of the country i.e. Punjab, Sindh, Balochistan, Khyber Pakhtoonkhwa, Federal Area and AJK as shown in the annexure - A. Total process of administration was completed in whole one year. From Punjab province 639 students of various colleges were randomly selected, 543 from Sindh province, 267 from Balochistan, 234 from Khyber Pakhtoonkhwa, 23 from federally administered areas and 16 from AJK. Each group has completed the test in stipulated time of 30 minutes. The students marked their responses by shading (black) appropriate column on the computerized based (OMR) answer sheet. Scoring was accomplished by scanning each sheet through a scanner controlled with the help of computerized software OMR.

The minimum and maximum possible scores on IIT and its sub-tests are as follows:

Test	Range of scores
Vocabulary	0 – 12
Arithmetic	0 - 12
Analogy	0 – 14
Information	0 - 11
Comprehension	0 – 12
Similarity	0 – 10
IIT	0 – 71

Percentiles

Percentiles scores are expressed in terms of the percentage of persons in the standardization sample who fall below a given raw score. They indicate the individual's relative position in the standardization sample. For example, if 20 percent persons give correct answer to 10 items in a test then the raw score of 10 correspond to the 20^{th} percentile. Percentile can be expressed as ranks in a group of 100. The 50^{th} percentile corresponds to the median. Percentile above 50 represents above average performance, those below 50 signify below average performance. Percentiles scores are easy to compute and comprehend. They are useful in explaining test results to individuals who have little background in the statistics of testing. They are universally applicable and can be used equally well with individuals of all ages, and for any type of test. However, one of the limitations of percentile scores is that they under emphasize the difference between scores lying towards the extremes of distribution. The percentile norms of IIT and its sub-tests are given in tables 27 – 38.

RESULTS

First Tryout

In the process of assessing the items, first consideration was given to the internal consistency of the items, calculated by item total correlation method. The difficulty levels and discrimination powers of only those items were determined that bear a significant correlation with the total score ($p < .001$). It was decided to retain the items with difficulty level ranging from .30 to .70 and with discrimination power equal or greater than .40.

Table 1

Indices showing discriminatory power of the items of each sub-test (First Draft)

	Number of Items					
Range	Vocabulary	Arithmetic	Analogy	Information	Comprehension	Similarity
.91 – 1.0	0	0	0	0	0	0
.81 - .90	0	0	0	0	0	1
.71 - .80	0	1	2	4	5	0
.61 - .70	2	5	5	4	6	3
.51 - .60	7	6	11	7	7	0
.40 – 50	7	7	8	5	1	1
.31 - .39	9	5	20	4	6	5
.21 - .30	7	0	0	1	2	3
.11 - .20	3	0	0	1	0	1
.01 - .10	2	2	0	3	3	1

Table 1 shows that discrimination power of items comprising Vocabulary Test ranged from .01 to .70. It is also evident from the table that 16 items out of 37 (43%) fall within the desired range (.40 and above), whereas 21 items (56%) fall below .40. Discrimination power of Arithmetic Test items ranged from .01 to .70. Table further reveals that 18 items out of 25 (72%) fall within the desired range (.40 and above), whereas 14 items (56%) fall below .40. Discrimination power of items belonging to Analogy Test ranged from .31 to .80. It is also evident from the table that 26 items out of 46 (56%) fall within the desired range (.40 and above), whereas 11 items (22%) fall below .40. Discrimination power of items included in Information Test ranged from .01 to .80. It is also evident from the table that 20 items out of 29 (68%) fall within the desired range (.40 and above), whereas 9 items (31%) fall below .40. Discrimination power of items comprising Comprehension Test ranged from .01 to .80. It is also evident from the table that 19 items out of 30 (63%) fall within the desired range (.40 and above), whereas 11 items (36%) fall below .40. Discrimination power of items included in Similarity Test ranged from .31 to .90. It is also evident from the table that 10 items out of 15 (66%) fall within the desired range (.40 and above), whereas 5 items (33%) fall below .30.

Table 2

Indices showing the difficulty level of the items of each sub-test (First Draft)

	Number of Items					
Range	Vocabulary	Arithmetic	Analogy	Information	Comprehension	Similarity
.91 – 1.0	0	0	0	0	0	0
.81 - .90	6	2	0	1	2	2
.71 - .80	6	4	13	7	2	3
.61 - .70	2	5	7	7	8	1
.51 - .60	6	3	10	6	6	3
.40 – 50	9	4	12	3	6	4
.31 - .39	2	4	3	2	2	2
.21 - .30	5	1	1	2	3	0
.11 - .20	1	2	0	1	1	0
.01 - .10	0	0	0	0	0	0

Table 2 shows that difficulty level of Vocabulary Test ranged from .11 to .90. As evident from table, 17 items out of 37 (45%) fall within the desired range (.40 to .70). 12 items were found too easy and 8 were too difficult. Difficulty level of Arithmetic Test ranged from .11 to .90. Results show that 12 out of 25 (48%) fall within the desired range (.40 to .70), 6 items were too easy (p=.71 to .90) and 7 items were below .39 level of difficulty. For Analogy Test difficulty level ranged from .21 to .80. As evident from the table, 29 items out of 46 (63%) fall within the desired range (.40 to .70). Out of the remaining 13 items were too easy (p=.71 to 90) and 4 was too difficult (p=.39 to .25) and hence were discarded. Difficulty level of Information Test ranged from .11 to .90. As evident from table, 13 items out of 29 (44%) fall within the desired range (.40 to .70). Out of the remaining 13 items, 8 were too easy (p=.71 to 90) and 5 were too difficult (p=.11 to .39) and hence were discarded. Difficulty level of Comprehension Test ranged from .01 to .90. As evident from table, 20 items out of 30 (46%) fall within the desired range (.40 to .70). Out of the remaining 10 items, 4 were too easy (p=.71 to 90) and 6 were too difficult (p=.11 to .39) and hence were discarded. Difficulty level of Similarity Test items ranged from .31 to .90. As evident from table, 10 items out of 15 (66%) fall within the desired range (.40 to .70). Remaining 5 items were too easy (p=.75 to 90) hence were discarded.

Table 3

The number of items discarded from the test during the experimental try-out

	Number of Items		
Sub –test	**Total**	**Discarded***	**Selected**
Vocabulary	37	25	12
Arithmetic	25	13	12
Analogy	46	32	14
Information	29	19	11
Comprehension	30	18	12
Similarity	15	5	10

*Vocabulary = 1, 2, 3, 4, 5, 6, 7, 9, 10, 12, 15, 16, 20, 22, 24, 25, 26, 29, 30, 31, 33, 34, 35, 36, 37.

*Arithmetic = 1, 5, 7, 8, 12, 15, 19, 20, 21, 22, 23, 24, 25.

*Analogy = 2, 3, 4, 5, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32, 33, 36, 37, 38, 39, 44.

*Information = 1, 2, 3, 6, 7, 8, 10, 13, 14, 17, 18, 20, 23, 24, 26, 27, 28, 29, 30, 31.

*Comprehension = 1, 2, 3, 4, 8, 9, 10, 11, 13, 16, 17, 19, 23, 24, 27, 28, 29, 30.

*Similarity = 3, 6, 9, 13, 17.

On the basis of the selection criteria for the effective items, all those items proved ineffective, obscure or complex, were discarded from the item pool as shown in table 3. As evident from table, the percentages of items discarded from Vocabulary Test, Arithmetic Test, Analogy Test, Information Test, Comprehension Test and Similarity Test were 67 percent, 52 percent, 69 percent, 63 percent, 60 percent and 33 percent respectively.

Table 3 depicts that 37 items of Vocabulary Test were reduced to 12, 25 items of Arithmetic Test were reduced to 12, 46 items of Analogy Test were reduced to 14, 30 items of Information Test were reduced to 11, 30 items of Comprehension Test were reduced to 12 and 15 items of Similarity Test were reduced to 10. All the selected items were rearranged in each sub-test according to their ascending order of difficulty level

(tables 4 - 9) and their distribution with reference to difficulty level and discrimination power are shown in tables 10 - 11.

Table 4

Three psychometric indices of the items of Vocabulary Test (Final Draft)

Items	Difficulty level	Discrimination level
1	.71	.56
2	.70	.56
3	.63	.55
4	.60	.60
5	.59	.59
6	.56	.47
7	.54	.64
8	.49	.65
9	.49	.41
10	.49	.57
11	.48	.41
12	.43	.48

Results of table 4 reveal that difficulty level indices for Vocabulary Test ranges from .43 to .71 and discrimination level from .41 to .65. As both difficulty and discrimination levels of each item is within acceptable range therefore is indicative of a good items.

Table 5

Three psychometric indices of the items of Arithmetic Test (Final Draft)

Items	Difficulty level	Discrimination level
13	.70	.57
14	.68	.62
15	.65	.67
16	.64	.64
17	.56	.57
18	.53	.49
19	.51	.50
20	.44	.49
21	.42	.62
22	.39	.53
23	.39	.48
24	.36	.51

Results of table 5 reveal that difficulty level indices for Arithmetic Test ranges from .36 to .70 and discrimination level from .48 to .67. As both difficulty and discrimination levels of each item is within acceptable range therefore is indicative of a good items.

Table 6

Three psychometric indices of the items of Analogy Test (Final Draft)

Items	Difficulty level	Discrimination level
25	.63	.68
26	.63	.68
27	.61	.59
28	.61	.45
29	.60	.55
30	.59	.80
31	.57	.47
32	.56	.48
33	.56	.67
34	.53	.52
35	.52	.75
36	.49	.50
37	.47	.58
38	.42	.43

Results of table 6 show that difficulty level indices for Analogy Test range from .42 to .63 and discrimination level from .43 to .68. As both difficulty and discrimination levels of each item is within acceptable range therefore is indicative of a good items.

Table 7

Three psychometric indices of the items of Information Test (Final Draft)

Items	Difficulty level	Discrimination level
39	.70	.78
40	.68	.74
41	.64	.58
42	.64	.72
43	.59	.55
44	.58	.46
45	.57	.53
46	.54	.50
47	.53	.60
48	.50	.72
49	.47	.36

Results of table 7 reveal that difficulty level indices for Information Test ranges from .47 to .70 and discrimination level from .36 to .78. As both difficulty and discrimination levels of each item is within acceptable range therefore is indicative of a good items.

Table 8

Three psychometric indices of the items of Comprehension Test (Final Draft)

Items	Difficulty level	Discrimination level
50	.66	.76
51	.65	.57
52	.61	.39
53	.61	.59
54	.60	.57
55	.58	.76
56	.52	.65
57	.49	.57
58	.49	.51
59	.47	.62
60	.45	.48
61	.45	.64

Results of table 8 demonstrate that difficulty level indices for Comprehension Test ranges from .45 to .66 and discrimination level from .39 to .76. As both difficulty and discrimination levels of each item is within acceptable range therefore is indicative of a good items.

Table 9

Three psychometric indices of the items of Similarity Test (Final Draft)

Items	Difficulty level	Discrimination level
62	.60	.61
63	.60	.82
64	.58	.25
65	.54	.35
66	.50	.50
67	.50	.50
68	.50	.39
69	.49	.39
70	.46	.64
71	.33	.34

Results of table 9 reveal that difficulty level indices for Similarity Test ranges from .33 to .60 and discrimination level from .34 to .82. As both difficulty and discrimination levels of each item is within acceptable range therefore is indicative of a good items.

Table 10

Table showing the discrimination power of the items of each sub-test (Final draft)

	Number of Items					
Range	**Vocabulary**	**Arithmetic**	**Analogy**	**Information**	**Comprehension**	**Similarity**
.91 – 1.0	0	0	0	0	0	0
.81 - .90	0	0	0	0	0	0
.71 - .80	0	0	0	0	0	0
.61 - .70	3	4	4	4	4	0
.51 - .60	4	3	7	5	3	4
.41 – 50	5	2	3	2	5	5
.31 - .40	0	3	0	0	0	1
.21 - .30	0	0	0	0	0	0
.11 - .20	0	0	0	0	0	0
.01 - .10	0	0	0	0	0	0

Results presented in table 10 show that 71 items fall within the desired range of discrimination (.3 and above). Discrimination power of 03 items of Vocabulary sub-scale was from .61 to .70, 04 were measured between .51 to .60 and discrimination power of 05 items ranged between .41 to .50. Whereas, discrimination power of 04 items of Arithmetic sub-scale was found between .61 to .70, 03 items ranged from .51 to .60, 02 items ranged from .41 to .50, and 03 items were from .31 to .40. 04 items of Analogy sub-scale demonstrated discrimination power between .61 to .70, 07 items were from .51 to .60 and 03 items were from .41 to .50. 04 items of Information sub-scale displayed discrimination level from .61 to .70, 05 items were from .51 to .60 and 02 items have discrimination power from .41 to .50. As far as Comprehension sub-scale is concern discrimination index of 04 items fall between .61 to .70, 03 items were measured between .51 to .60 and 05 items were found with .41 to .50 discrimination power. For Similarity sub-scale, discrimination power of 04 items was measured between .51 to .60, range of 05 items was found between .41 to .50 and 01 item was adjudged between .31 to .40.

Table 11

Table showing the difficulty level of the items of each sub-test (Final draft)

	Number of Items					
Range	Vocabulary	Arithmetic	Analogy	Information	Comprehension	Similarity
.91 – 1.0	0	0	0	0	0	0
.81 - .90	0	0	0	0	0	0
.71 - .80	0	0	0	0	0	0
.61 - .70	3	4	4	4	4	0
.51 - .60	4	3	7	5	3	4
.41 – 50	5	2	3	2	4	5
.31 - .40	0	3	0	0	0	1
.21 - .30	0	0	0	0	0	0
.11 - .20	0	0	0	0	0	0
.01 - .10	0	0	0	0	0	0

Results presented in table 11 shows that 71 items fall within the desired range of discrimination (.3 to .7). Difficulty level of 03 items of Vocabulary sub-scale was from .61 to .70, 04 were between .51 to .60 and 05 were from .41 to .50. Whereas, difficulty level of 04 items of Arithmetic sub-scale fall between .61 to 70, 03 items ranged from .51 to .60, 02 items were from .41 to .50, and 03 items fall between .31 to .40. 04 items of Analogy sub-scale demonstrated difficulty level from .61 to .70, 07 items were from .51 to .60 and 03 items were from .41 to .50. 04 items of Information sub-scale displayed difficulty level between .61 to .70, 05 items were from .51 to .60 and 02 items were from .41 to .50. For Comprehension sub-scale difficulty level of 04 items was between .61 to .70, 03 items ranged between .51 to .60 and 05 items were from .41 to .50. Discrimination power of 04 items of Similarity sub-scale were measured from .51 to .60, range of 05 items were found between .41 to .50 and 01 item was measured from .31 to .40.

Average time needed for IIT

The time limits for newly developed test were determined by computing the average time taken by 80% of the subjects who completed each sub-test. As shown in table 12 the average time required by majority of subjects is 30 minutes which was taken as time limit for the newly developed test.

Table 12

Average time needed for IIT

Test	Items	Time in minutes
IIT	71	30

Reliability of the Test

Reliability is one of the major index of the efficiency of any measure. The extent, to which one can depend upon a test, is very much determined by the reliability of the test. Following methods were used to establish test reliability:

1. Split – half Method
2. Kuder Richardson Method (KR – 20)

Split – half Reliability

To determine the split-half reliability, the test was divided into two halves based on odd-even items. Pearson product-moment coefficient of correlation was computed for the two half scores. The split-half reliability of the test is .79 and .77 respectively for Part –I and Part-II.

Table 14

Kuder Richardson Reliability

The inter-item consistency of the test was estimated by applying the Kuder Richardson Formula-20. Table 14 shows the KR-20 Reliability estimates for the subtests and the full test.

Reliability coefficients of IIT and its sub – tests (KR-20)

Test	Total No. of Items	KR – 20
Vocabulary	12	.632*
Arithmetic	12	.654*
Analogy	14	.564*
Information	11	.494*
Comprehension	12	.503*
Similarity	10	.659*
IIT	71	.695*

***p<.001

Table 14 shows that all reliability coefficients are high and as such demonstrates internal consistency of IIT and its sub-tests. These indices also indicate that items of IIT and its sub-tests are highly homogenous. These results suggest that the newly developed test is a reliable instrument for measuring intelligence of adolescence with age range between 17 years to 25 years and with minimum academic qualification of Intermediate Pakistan.

Validity of the Tests

Validity is an essential characteristic of any test. It refers to the degree or extent to which a test measures what it purports to measure. The validity of the test was determined by the following methods:

1. Construct Validation.
2. Concurrent Validation

Intercorrelaion among the Subscales and Indigenous Intelligence Test (IIT)

Table 15

Subscale	1	2	3	4	5	6
Vocabulary	1					
Arithmetic	.375***					
Analogy	.549***	.390***				
Information	.283***	.297***	.373***			
Comprehension	.401***	.302***	.413***	.363***		
Similarity	.275***	.152***	..275***	.233***	.447***	
IIT (Total)	.732***	.648***	.762***	.585***	.711*	.584***

***p<.001

Results presented in Table 15 show significant correlations among all the subtests as well as the full test (p<.001). These findings demonstrate internal consistency of the test suggesting that all the subtests measure the general ability and similar cognitive functions. An inspection of Table 15 reveals that correlation between each one of the subtests and the full test is much higher than the intercorrelation among the subtests. Statistically significant results also exhibits construct validity of the test. The significant correlation among sub-scales of IIT is also indicative of convergent validity of the test as it correlates highly with the variable (g). In addition, it does not correlate to any other variable with which it should not (discriminant validity).

Concurrent Validity

Table 16

Concurrent Validity of IIT and sub-tests

Test	College Marks
Vocabulary	.302***
Arithmetic	.223***
Analogy	.267***
Information	.086**
Comprehension	.167***
Similarity	.131***
IIT	.304***

**p<.01; ***p<.001

Statistically significant correlations between the scores of Ss on IIT and marks obtained in the last examination (table 16) are an evidence of the concurrent validity of the test. The significant correlation of the test scores presents a positive correlation between general intelligence and educational achievement.

Table 18

Item total Correlation for subs-test Arithmetic (N=1669)

Items	*r*
13.	.452**
14.	.476**
15.	.483**
16.	.523**
17.	.441**
18.	.472**
19.	.530**
20.	.463**
21.	.526**
22.	.514**
23.	.497**
24	.429**

*p<.05; **p<.01; ***p<.0.01

Results (Table 18) show item-total correlation of sub-scale Arithmetic with IIT. Each item is highly significant measure of the Arithmetic sub-test. Statistically significant results also exhibit the evidence of construct validity of the sub-scale.

.

Table 19

Item to total Correlation for sub-test Analogy (N=1669).

Items	*r*
25	.382**
26	.367**
27	.405**
28	.462**
29	.268**
30	.513**
31	.216**
32	.475**
33	.378**
34	.415**
35	.474**
36	.407**
37	.470**
38	.523**

*p<.05; **p<.01; ***p<.0.01

Results (Table 19) show item-total correlation of sub-scale Analogy with IIT. Each item is highly significant measure of the Analogy sub-test. Statistically significant results also exhibit the evidence of construct validity of the sub-scale.

Table 20

Item to total Correlation for sub-test Information (N=1669)

Items	*r*
39	.408**
40	.420**
41	.370**
42	.452**
43	.467**
44	.422**
45	.422**
46	.437**
47	.483**
48	.463**
49	.414**

*p<.05; **p<.01; ***p<.0.01

Results (Table 20) show item-total correlation of sub-scale Information with IIT. Each item is highly significant measure of the Information sub-test. Statistically significant results also exhibit the evidence of construct validity of the sub-scale.

Table 21

Item to total Correlation for sub-test Comprehension (N=1669)

Items	*r*
50	.436**
51	.415**
52	.444**
53	.412**
54	.374**
55	.994**
56	.454**
57	.466**
58	.498**
59	.489**
60	.519**
61	487**

*p<.05; **p<.01; ***p<.0.01

Results (Table 21) show item-total correlation of sub-scale Comprehension with IIT. Each item is highly significant measure of the Comprehension sub-test. Statistically significant results also exhibit the evidence of construct validity of the sub-scale.

Table 22

Item to total Correlation for sub-test Similarity (N=1669)

Items	*r*
62	.566**
63	.494**
64	.535**
65	.427**
66	.517**
67	.536**
68	.611**
69	.515**
70	.678**
71	.428**

*p<.05; **p<.01; ***p<.0.01

Results (Table 22) show item-total correlation of sub-scale Similarity with IIT. Each item is highly significant measure of the Similarity sub-test. Statistically significant results also exhibit the evidence of construct validity of the sub-scale.

Other Statistical Analysis

To study the significance of differences between mean IIT scores of various groups, t-test for independent groups and ANOVA for within groups was applied for the following comparison:

1. Male versus female students.
2. Urban versus rural students.
3. Province wise students.
4. Income group wise.

Difference in test scores of the male and female students

To determine the significance of differences between the mean IIT scores of male and female students, t-test for independent groups was applied. The sample consists of 157 female and 1484 male students of various colleges from four provinces, AJK and federally administered areas. Results of the t-test are given in table 26.

Difference in test scores of the male and female students

Table 23

Mean, Standard Deviation and t-values showing significance of gender differences on IIT and its sub-tests

IIT/Sub-scales	Male (n=1484)		Female (n=157)		t-value	Cohan's *d*
	Mean	SD	Mean	SD		
Vocabulary	7.98	2.45	9.01	2.16	5.028***	0.43
Arithmetic	8.04	2.60	6.72	2.63	5.982***	0.51
Analogy	9.12	2.56	9.57	2.58	2.142*	0.18
Information	9.23	1.65	8.56	1.98	4.728**	0.40
Comprehension	8.12	2.09	8.22	2.23	.576	0.05
Similarity	6.69	2.29	6.62	2.50	.387	0.03
IIT	49.20	9.24	48.73	9.23	.601	0.05

*p < .05; **p< .01; ***p< .01; df = 1639

Table 23 shows the mean difference between male and female students on the IIT score and its sub-scales. The figure shows that there are significant mean differences on vocabulary, arithmetic, analogy and information and non-significance difference between male and female on Comprehension, Similarity and overall IIT scores.

Difference in test scores of the urban and rural students

Table 24

Mean, Standard Deviation and t-values showing significance of urban and rural differences on IIT and its sub-tests

IIT/Sub-scales	Urban (n=1159)		Rural (n=439)		t-value	Cohan's *d*
	Mean	SD	Mean	0.43		
Vocabulary	8.26	2.44	7.65	0.51	4.485***	0.25
Arithmetic	7.96	2.58	7.89	0.18	.487	0.03
Analogy	9.33	2.53	8.77	0.40	3.894***	0.22
Information	9.12	1.69	9.35	0.05	2.467*	0.14
Comprehension	8.22	2.17	7.97	0.03	2.092*	0.12
Similarity	6.71	2.36	6.60	0.05	.870	0.05
IIT	49.62	9.34	48.25	**Cohan's *d***	2.680**	0.15

*p < .05; **p< .01; ***p< .01; df = 1639

Table 24 shows the mean difference between urban and rural students on the IIT score and its sub-scales. The figure shows that there are significant mean differences between urban and rural students on vocabulary, arithmetic, analogy information and overall IIT scores. Whereas, no-significant difference has been observed between scores of urban and rural students on Comprehension and Similarity sub-scales.

Table 25

Province wise differences in scores on Indigenous Intelligence Test (IIT) and its subscale

IIT & Subscales	Punjab		Sindh		Balochistan		Khyber Pakhtonkhwa		Federal Area		AJK			
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	*F*	*P*
Vocabulary	8.53	2.28	7.82	2.49	M	SD	7.72	2.60	8.60	2.53	7.73	2.64	12.161	.000
Arithmetic	8.36	2.49	7.95	2.55	7.34	2.49	7.76	2.81	8.14	2.87	7.88	2.77	19.704	.000
Analogy	9.70	2.39	8.93	2.58	6.56	2.59	8.97	2.3o	9.72	2.69	9.84	1.75	19.208	.000
Information	9.21	1.73	9.17	1.52	8.04	2.83	9.11	1.70	9.16	1.85	10.07	1.16	2.795	.016
Comprehension	8.27	2.18	8.06	1.98	8.92	1.81	8.44	1.90	8.38	1.98	8.30	1.91	5.906	.000
Similarity	6.69	2.41	6.56	2.36	7.52	2.35	6.79	2.07	7.23	2.07	7.46	1.50	1.642	.146
Total	50.78	8.70	48.51	9.11	6.56	2.27	48.81	9.00	51.25	10.24	51.30	8.53	17.449	.000

F (5, 1444) = 12.161; p < .001

As evident from Table 25, all the F values are highly significant except for Similarity subscale. The difference in Mean scores of each subscale is indicative of differences in performance of population from different provinces and as such different cultural and environmental background. However, insignificant *F* value for similarity subscale is indicative of no difference in the performance of groups on this subscale.

Table 26

Difference in test scores of the income group wise

Income group wise differences in scores on Indigenous Intelligence Test (IIT) and its subscales

IIT & Subscales	Vocabulary		Arithmetic		Analogy		Information		Comprehension		Similarity	
Income Range	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
< 15000	7.38	2.54	7.75	2.66	8.64	2.53	8.94	1.82	7.76	2.28	6.08	2.49
15000-24999	7.94	2.32	7.68	2.59	8.92	2.58	9.18	1.64	7.96	2.04	6.67	2.15
25000-49999	8.40	2.44	8.05	2.64	9.47	2.59	9.27	1.63	8.32	2.11	6.86	2.37
50000-74999	8.42	2.49	8.27	2.65	9.57	2.66	9.01	1.97	8.51	2.11	7.41	1.83
75000-98999	8.48	2.30	8.45	2.30	10.43	2.19	9.45	1.52	8.22	2.00	7.22	2.60
>99000	9.19	2.11	8.57	2.24	9.91	2.45	9.44	1.34	8.76	1.97	6.60	2.59
Total	8.05	2.46	8.57	2.24	9.15	2.59	9.15	2.59	8.11	2.14	6.66	2.34

The result of Table 26 shows the mean difference between various income group students on the IIT score and subscales. Mean scores of IIT and all the subscales show progressive increase with increase in income range which demonstrates validity of IIT. The test scores do provide empirical evidence that IIT can differentiate well between various income groups.

Determining the Classificatory indices of the test

Percentile norms were computed for the normative sample of 1669 candidates representing all the provinces of Pakistan that also included FATA and AJK. Furthermore, province wise norms were also calculated so as to gauge true performance of candidates hailing from different provinces. The detail of which is given in Annexure A.

Percentiles

The maximum attainable score on the Vocabulary test was 12 whereas the minimum possible score was zero. The highest score obtained by the students on the test was 12 while the lowest score was 4. The maximum attainable score on the Arithmetic test was 12 whereas the minimum possible score was zero. The highest score obtained by the students on the test was 12 while the lowest score was 3. The maximum attainable score on the Analogy test was 14 whereas the minimum possible score was zero. The highest score obtained by the students on the test was 14 while the lowest score was 4. The maximum attainable score on the Information test was 11 whereas the minimum possible score was zero. The highest score obtained by the students on the test was 11 while the lowest score was 6. The maximum attainable score on the Comprehension test was 11 whereas the minimum possible score was zero. The highest score obtained by the students on the test was 11 while the lowest score was 4. The maximum attainable score on the Similarity test was 10 whereas the minimum possible score was zero. The highest score obtained by the students on the test was 10 while the lowest score was 1. The maximum attainable score on the total test was 71 whereas the minimum possible score was zero. The highest score obtained by the students on the total test was 65 while the lowest score was 32. Scores of Ss normative sample of the present study were converted into percentiles. Separate percentiles were computed for total test scores and for the scores of each sub-test. Percentiles norms of IIT and its sub-test are shown in tables 27 – 33. Whereas, maximum score attained by a candidate on province Punjab scale is 63 and lowest were 34. Similarly, highest score attained by a student on province Sindh scale is 62 and lowest were 33. Highest score attained by a student on province Balochistan scale is 60 and lowest were 28.On Khyber Pakhtoonkhwa scale the highest and lowest scores were 62 and 31. In Federal Area scale both the scores were 63.2 and 31. AJK population managed to score 63 and 31.4 as highest and lowest score. Province wise norms are presented in tables from 34 – 38.

Table 27

Percentile score of Vocabulary test (N = 1696)

Percentiles	Raw Scores
5	4
15	5
25	6
35	7
45	8
55	9
70	10
85	11
95	12

The maximum attainable score on the Vocabulary test was 12 whereas the minimum possible score was zero. The highest score obtained by the students on the test was 12 while the lowest score was 4.

Table 28

Percentile score of Arithmetic test (N = 1696)

Percentiles	Raw Scores
5	3
10	4
15	5
25	6
30	7
45	8
55	9
70	10
85	11
95	12

The maximum attainable score on the Arithmetic test was 12 whereas the minimum possible score was zero. The highest score obtained by the students on the test was 12 while the lowest score was 3.

Table 29

Percentile score of Analogy test (N = 1696)

Percentiles	Raw Scores
5	4
10	5
15	6
20	7
25	8
40	9
50	10
70	11
85	12
95	13
99	14

The maximum attainable score on the Analogy test was 14 whereas the minimum possible score was zero. The highest score obtained by the students on the test was 14 while the lowest score was 4.

Table 30

Percentile score of Information test (N = 1696)

Percentiles	Raw Scores
5	6
10	7
15	8
30	9
50	10
80	11

The maximum attainable score on the Information test was 11 whereas the minimum possible score was zero. The highest score obtained by the students on the test was 11 while the lowest score was 6.

Table 31

Percentile score of Comprehension test (N = 1696)

Percentiles	Raw Scores
5	4
10	5
15	6
20	7
35	8
50	9
75	10
95	11

The maximum attainable score on the Comprehension test was 11 whereas the minimum possible score was zero. The highest score obtained by the students on the test was 11 while the lowest score was 4.

Table 32

Percentile score of Similarity test (N = 1696)

Percentiles	Raw Scores
5	1
10	3
15	5
25	6
40	7
60	8
85	9
99	10

The maximum attainable score on the Similarity test was 10 whereas the minimum possible score was zero. The highest score obtained by the students on the test was 10 while the lowest score was 1.

Table 33

Percentile score of IIT (N = 1696)

Percentiles	Raw Scores
5	32
10	37
15	39
20	41
25	43
30	45
35	46
40	47
45	49
50	50
55	51
60	52
65	54
70	55
75	56
80	57
85	59
90	60
95	62
99	65

The maximum attainable score on the total test was 71 whereas the minimum possible score was zero. The highest score obtained by the students on the total test was 65 while the lowest score was 32.

Table 34

Percentile score of Province Punjab IIT (N = 712)

Percentiles	Raw Scores
5	34
10	38
25	45
50	52
75	57
90	61
95	63

The maximum attainable score on the province Punjab scale was 71 whereas; the minimum possible score was zero. The highest score obtained by the students on the scale was 63 while the lowest score was 34.

Table 35

Percentile score of Province Sindh IIT (N = 458)

Percentiles	Raw Scores
5	33
10	36
25	43
50	49
75	55
90	60
95	62

The maximum attainable score on the province Sindh scale was 71 whereas; the minimum possible score was zero. The highest score obtained by the students on the scale was 62 while the lowest score was 33.

Table 36

Percentile score of Province Balochistan IIT (N = 273)

Percentiles	Raw Scores
5	28
10	32
25	39
50	46
75	52
90	56
95	60

The maximum attainable score on the province Balochistan scale was 71 whereas; the minimum possible score was zero. The highest score obtained by the students on the scale was 60 while the lowest score was 28.

Table 37

Percentile score of Province Khyber Pakhtoonkhwa IIT (N = 126)

Percentiles	Raw Scores
5	31
10	38
25	44
50	49
75	55
90	59
95	62

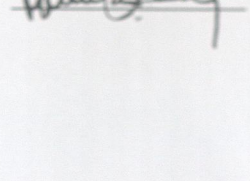
The maximum attainable score on the province Khyber Pakhtoonkhwa scale was 71 whereas; the minimum possible score was zero. The highest score obtained by the students on the scale was 62 while the lowest score was 31.

Declaration

I, MAHMOOD SADIQ, do hereby solemnly declare that the work submitted in this thesis is my own, and has not been presented previously to any other institution or university for a degree.

This work has carried out and completed at the Department of Psychology, Peshawar University.

Researcher:



Appendix-F

1456

ANSWER SHEET
INTELLIGENCE TEST

Time Limit: 30 Mins

Name: SOFIA IQBAL /23 College/University: Quaid-i-Azam University

Marks Obtained % F.A/F.Sc/B.A/B.Sc/M.Sc: 72%

SEX
- MALE ○
- FEMALE ●

- URBAN ●
- RURAL ○

PROVINCE
- PUNJAB ○
- NWFP ○
- SINDH ○
- BLOCHISTAN ○
- AJK ○
- FEDERAL AREA ●

Age: 24

Marks %: 72

Monthly Average Income Group
- Less Than 15000 ○
- 15000 to 25000 ○
- 25000 to 50000 ●
- 50000 to 75000 ○
- 75000 to 99000 ○
- More than 99000 ○

EXAMPLES
- Q.1 (A) (B) (C) ●
- Q.2 (A) ● (C) (D)
- Q.3 ● (B) (C) (D)
- Q.4 (A) (B) (C) (D)
- Q.5 (A) (B) (C) (D)
- Q.6 (A) (B) (C) (D)

INTELLIGENCE TEST — ANSWERS

Q.1	A	Q.15	D	Q.29	D	Q.43	C	Q.57	C
Q.2	C	Q.16	D	Q.30	A	Q.44	B	Q.58	D
Q.3	A	Q.17	A	Q.31	C	Q.45	D	Q.59	C
Q.4	B	Q.18	C	Q.32	C	Q.46	C	Q.60	B
Q.5	D	Q.19	B	Q.33	B	Q.47	A	Q.61	B
Q.6	B	Q.20	A	Q.34	D	Q.48	D	Q.62	A
Q.7	B	Q.21	A	Q.35	D	Q.49	D	Q.63	D
Q.8	C	Q.22	A	Q.36	C	Q.50	B	Q.64	B
Q.9	B	Q.23	B	Q.37	A	Q.51	B	Q.65	B
Q.10	B	Q.24	B	Q.38	D	Q.52	B	Q.66	D
Q.11	B	Q.25	B	Q.39	D	Q.53	B	Q.67	D
Q.12	B	Q.26	B	Q.40	C	Q.54	B	Q.68	C
Q.13	A	Q.27	D	Q.41	B	Q.55	C	Q.69	D
Q.14	B	Q.28	B	Q.42	C	Q.56	C	Q.70	A
								Q.71	A

LIST OF EDUCATIONAL INSTITUTES FOR DATA COLLECTION

PUNJAB PROVINCE

1. Government G C University, Lahore
2. Government Forman Christian College, Lahore
3. Punjab University, Lahore.
4. Fazaia College, Lahore
5. Bahria College, Islamabad.
6. Islamabad Boys School, F/10

KHYBER PAKHTON KHWA PROVINCE

1. University of Peshawar
2. Islamia College Peshawar
3. Fazaia College, Kohat
4. Women Degree College, Kohat

SINDH PROVINCE

1. Karachi University
2. Bahria College NORE I, Karachi
3. APS & College, Malir Cantt, Karachi
4. Urdu University, Karachi

BALOCHISTAN PROVINCE

1. F G Degree College Quetta Cantt
2. PAF Degree College Sumungli Base Quetta
3. Tameer e Nau College Quetta
4. Government Science College Quetta

Table 38

Percentile score of Federal Area IIT (N = 55)

Percentiles	Raw Scores
5	31
10	38
25	46
50	53
75	58
90	62
95	63.2

The maximum attainable score on the Federal Area scale was 71 whereas; the minimum possible score was zero. The highest score obtained by the students on the scale was 63.2 while the lowest score was 31.

Table 39

Percentile score of AJK Area IIT (N = 26)

Percentiles	Raw Scores
5	31.4
10	36.1
25	47.5
50	51.5
75	58.25
90	63
95	63

The maximum attainable score on the AJK scale was 71 whereas; the minimum possible score was zero. The highest score obtained by the students on the scale was 63 while the lowest score was 31.4.

DISCUSSION

An Indigenous Verbal Intelligence Test in English for local population (IIT) has been developed, validated and standardized. As discussed earlier, the test is intended to be used as a screening tool for candidates applying in Armed Forces, Central Superior Services and Public Service Commission examinations where English carries a status of an official language. The IIT was developed in English to be used as a screening tool and a measure of general intelligence for adolescence having age range between 17 to 25 years with minimum academic qualification of Intermediate and are interested to apply in Armed Forces of Pakistan or want to appear in CSS and other such competitive examinations. Efforts have been submitted to make IIT and its sub-tests psychometrically valid and reliable.

The IIT consists of six sub-tests: Vocabulary test, Arithmetic test, Analogy test, Information test, Comprehension test and Similarity test. The total number of items of the test is 71 and time limit is 30 minutes.

The item analysis (Tables 1 -11) was carried out to select the best items. The items having discrimination power within the range (.40 to .80) were considered highly effective as they seem to discriminate between high and low achievers on the test. Items showing discrimination power less than .40 were discarded from the item pool. As far as the difficulty levels of the items are concerned, the items falling within the range (.40 to .70) were retained, since they are neither very easy nor very difficult (Gregory, 1996, 2004; Riaz, 2008).

A detailed review of the statistical information presented in Table 1 reveals that discrimination power of items comprising Vocabulary Test ranged from .01 to .70. It is also evident from the table that 16 items out of 37 (43%) fall within the desired range (.40 and above), whereas 21 items (56%) fall below .40. Table 2 shows that difficulty level of Vocabulary Test ranged from .11 to .90. As evident from table, 17 items out of 37 (45%) fall within the desired range (.40 to .70). Out of the remaining 20 items, 12 were too easy (p=.71 to 90) and 8 were too difficult (p=.11 to .25) and hence were discarded. Results

reveal that 17 items (45%) fall within the range of .40 to .70 of difficulty level while 16 items (45%) fall within the range of .40 and above of discrimination power.

The discrimination power of Arithmetic Test items ranged from .01 to .70. Table further reveals that 18 items out of 25 (72%) fall within the desired range (.40 and above), whereas 14 items (56%) fall below .40. Table 2 shows that difficulty level of Arithmetic Test ranged from .11 to .90. Results show that 12 out of 25 (48%) fall within the desired range (.40 to .70), 6 items were too easy (p=.71 to .90) and 7 items were below .39 level of difficulty. Results reveal that 12 items (48%) fall in the range of .40 to .70 of difficulty level while 18 items (72%) fall within the range of .41 and above of discrimination power (see Table 1).

The items belonging to Analogy Test reveals discrimination power from .31 to .80. It is also evident from the table that 26 items out of 46 (56%) fall within the desired range (.40 and above), whereas 11 items (22%) fall below .40. Table 2 shows that difficulty level of Analogy Test ranged from .21 to .80. As evident from table, 29 items out of 46 (58%) fall within the desired range (.40 to .70). Out of the remaining items, 13 were too easy (p=.71 to 90) and 4 were too difficult (p=.21 to .39) and hence were discarded. Results reveal that 29 items (63%) fall within the range of .40 to .70 of difficulty level while 26 items (56%) fall within the range of .40 and above of discrimination power (see Table 1).

Information Test items displayed discrimination power of from .01 to .80. It is also evident from the table that 20 items out of 29 (68%) fall within the desired range (.40 and above), whereas 9 items (31%) fall below .40. Table 2 shows that difficulty level of Information Test ranged from .11 to .90. As evident from table, 16 items out of 29 (55%) fall within the desired range (.40 to .70). Out of the remaining 13 items, 8 were too easy (p=.71 to 90) and 5 were too difficult (p=.11 to .39) and hence were discarded. Results reveal that 16 items (55%) fall within the range of .40 to .70 of difficulty level while 20 items (68%) fall within the range of .40 and above of discrimination power (see Table 1).

The subscale Comprehension items demonstrated discrimination power from .01 to .80. It is also evident from the table that 19 items out of 30 (63%) fall within the desired range (.40 and above), whereas 11 items (36%) fall below .40. Table 2 shows that difficulty level of Comprehension Test ranged from .01 to .90. As evident from table, 20 items out of 30 (66%) fall within the desired range (.40 to .70). Out of the remaining 10

items, 4 were too easy (p=.71 to 90) and 6 were too difficult (p=.11 to .39) and hence were discarded. Results reveal that 20 items (66%) fall within the range of .40 to .70 of difficulty level while 19 items (63%) fall within the range of .40 and above of discrimination power (see Table 1).

The discrimination power of items included in Similarity Test ranged from .31 to .90. It is also evident from the table that 10 items out of 15 (66%) fall within the desired range (.40 and above), whereas 5 items (33%) fall below .30. Table 2 shows that difficulty level of Similarity Test items ranged from .40 to .90. As evident from table, 10 items out of 15 (66%) fall within the desired range (.40 to .70). Remaining 5 items were too easy (p=.75 to 90) hence were discarded. Results reveal that 10 items (66%) fall within the range of .40 to .70 of difficulty level while 10 items (66%) fall within the range of .40 and above of discrimination power (see Table 1).

The analysis of results mentioned at table 3 determines that when item analysis was run on first try-out, 25 items of Vocabulary sub-scale were discarded, 13 from Arithmetic sub-scale were found below the standard, 32 of Analogy were not within unacceptable range, 19 from Information sub-scale were not meeting the criteria, 18 of Comprehension sub-scale were found below the standard and 5 from Similarity sub-scale were gauged as weak items and as such discarded.

Results presented in tables 4 to 9 makes obvious the psychometric indices of each item of all the sub-scales (final draft). The difficulty level of Vocabulary sub-scale items ranges from .43 to .71 whereas; discrimination level ranged from .41 to .65. Difficulty level of Arithmetic sub-scale items ranged from .36 to .70 and discrimination level from .49 to .67. Ranges of indices for discrimination level and difficulty level were .42 to .63 and .43 to .68 respectively for Analogy sub-scale items. Discrimination level of items for Information sub-scale ranged from .47 to .70 and difficulty level from .36 to .78. Items discrimination range for Comprehension sub-scale was measured from .39 to .76, and difficulty level ranged from .45 to .66. Finally, indices indicating item difficulty level and discrimination level ranged from .33 to .60, and .25 to .82 respectively for Similarity sub-scale.

Results presented in Table 10 displays that 71 items fall within the desired range of discrimination (.40 and above) as recommended by Anastasi & Urbina (2012); Ebel (1965). Discrimination power of 03 items of Vocabulary sub-scale was from .61 to .70, 04

were between .51 to .60 and 05 were from .41 to .50. Whereas, discrimination power of 04 items of Arithmetic sub-scale were from .61 to .70, 03 were from .51 to .60, 02 were from .41 to .50, and 03 were from .31 to .40. 05 items of Analogy sub-scale demonstrated discrimination power from .61 to .70, 11 items were from .51 to .60 and 08 were from .41 to .50. 04 items of Information sub-scale displayed discrimination level between .61 to .70, 05 items were from .51 to .60 and 02 items were from .41 to .50. As far as Comprehension sub-scale is concern 04 items fall between .61 to .70, 03 items were from .51 to .60 and 05 items were from .41 to .50. Finally, for Similarity sub-scale discrimination power of 04 items was measured between .51 to .60, range of 05 items was found between .41 to .50 and 01 item was adjudged from .31 to .40. The above indices indicating discrimination for each item of every sub-scale were found in consonant with the laid down ranges as recommended by various experts of the field (Ebel, 1965; Anastasi, 1980; Anastasi & Urbina 2012; Gregory 1996, 2004; Riaz 2008).

Results presented in table 11 shows that 71 items fall within the desired range of difficulty i.e., .4 to .7 (Riaz, 2008). Difficulty level of 03 items of Vocabulary sub-scale was from .61 to .70, 04 were between .51 to .60 and 05 were from .41 to .50. Whereas, difficulty level of 04 items of Arithmetic sub-scale were from .61 to .70, 03 were from .51 to .60, 02 were from .41 to .50, and 03 were from .31 to .40. 07 items of Analogy sub-scale demonstrated difficulty level from .61 to .70, 10 items were from .51 to .60 and 12 were from .40 to .50. 04 items of Information sub-scale displayed difficulty level between .61 to .70, 06 items were from .51 to .60 and 03 items were from .41 to .50. As far as Comprehension sub-scale is concerned difficulty level of 04 items fell between .61 to .70, 06 items were from .51 to .60 and 06 items were from .41 to .50. Finally, for Similarity sub-scale discrimination power of 04 items was measured from .51 to .60, range of 04 items were found between .41 to .50 and 02 item was adjudged from .31 to .40. The above indices showing difficulty level of each item in every sub-scale were found in harmony with the recommended ranges mentioned by various experts (Ebel, 1965; Anastasi, 1980; Anastasi & Urbina 2012; Gregory 1996, 2004; Riaz 2008).

The items in each sub-test give a reasonable distribution of ranges of discrimination power and difficulty level. Most of the items possess sound psychometric bases. Item discrimination analysis of the items of six sub-tests of IIT establishes the validity of each item in measuring intelligence, besides ascertaining their discriminatory power (Tables 10 – 11).

The average time needed to complete the whole test is mentioned in Table 12. The time calculation for the test is based on the average time consumed by a sample of 50 students. While administering the test on the sample, clear instructions were communicated to complete the test as early as possible and must raise the hand when an examinee completes the test. Start time for the test was common for each subject whereas finishing time for each subject was different and was noted separately. Finally, time consumed by the subject who has completed the test first and the subject completing the test in the last were taken out and average of both extremes was measured as the time limit for the test, which was 30 minutes.

The reliability coefficients calculated for IIT and its sub-test by two applied methods i.e. Split-half reliability and Kuder Richardson reliability method, were significant. These results suggest a higher level of reliability both in terms of internal consistency and temporal stability of results. The reliability index reveals that the test is a reliable instrument for measuring general intelligence (see tables 13-14). It is important to note that split-half reliability of each sub-scale is relatively low (due lesser number of item in each subscale) when compared with reliability indices of total IIT. This phenomenon goes with the assertion that the length of a test (number of items in a test) is a major factor that influences the test reliability (Riaz, 2008).

The coefficient of correlation between IIT and its sub-tests (table 15) are also highly significant ($P<.001$). The results demonstrate the internal consistency of IIT and its subtests which is indicative of the fact that the sub-tests are taping some universal or general factor and similar mental functions. This positive correlation indicates that all the verbal items included in the sub-test are loaded on general intelligence factor *g* and crystallized *gf*. The correlations of six sub-test with the total score also provide evidence of the consistency of the test across individual components. According to Anastasi (1990), these correlations, if moderately high can be cited as evidence that the new intelligence test measures approximately the same area of behavior that has been measured by other established Intelligence tests. In addition, significant correlation among sub-scales of IIT is also indicative of convergent validity of the test due its high correlation with the variable (*g*). Furthermore, it does not correlate to any other variable with which it should not (discriminant validity)

The correlation between college marks and scores of IIT and its sub-tests are statistically significant ($p<.001$) (see Table 16). These results support the study by Crano,

Denny and Campbell (1972), that correlation of intelligence tests taken at one point of time and achievement measure taken at a later point of time tend to be higher than the two measures taken in reverse order of times. These results are in accordance with the theoretical assumption of Murphy and Davidshofer (1998), that, theoretical correlation could range in absolute value from 0.0 to 1.0, whereas in practice, most validity coefficients tend to be fairly small. A good, carefully chosen test is not likely to show a correlation greater than .50 with an important criterion and in fact, validity coefficients greater than .30 to .40 are commonly considered high (Riaz, 2008) in applied settings. Schmidt, Hunter, & Pearlman (1981) in one of the studies found validity coefficients of .19, .24 and .21 for job grades with verbal ability, quantitative ability and reasoning ability respectively. Guion (1991) noted the validity of measurement is not always necessary to guaranteed validity for decision. Although both sets of scores seem to be going in the same direction, it should not be inferred that any one who does good on a traditional college examination, does equally good on intelligence test. However, one can conclude that educational achievement has nothing to do with the understanding of the text knowledge; rather one can say that school achievement is not dependent upon only intelligence. Modern researches on general intelligence have proved validity generalization aimed at forecasting educational outcomes, occupational training and work performance. Fore some benchmarks, general cognitive ability co-varies .70 to .80 with academic achievements, and .40 to .70 with military training assignments, if intelligence test is administered before the two mentioned events. (Brody, 1992; Gottfredson, 1997; and Jensen, 1998).

The validity coefficient of the IIT and its subtests (Tables 15-22) indicate significant correlation between IIT and its sub-scale. These findings do establish the construct and concurrent validity of the test and demonstrate the validity of IIT as a valid measure of general intelligence.

In view of the statistical evidence presented in the above paragraphs, it can be concluded that items included in IIT has demonstrated their validity significantly and when presented in the form of a test can prove an effective measure of intelligence, as it can differentiate adequately between individuals varying in intellectual abilities. The scale can also be used as a screening tool for students who apply in Armed Forces, Public Service commission and similar organization in high number.

Generally there exists no significant gender difference in general intelligence; however, differences were observed when the factors forming intelligence are broken into different parts. The data of present research (Table 23) show that females did better than males on Vocabulary test. These findings go with the research (Wechsler, 1938) that females tend to be superior to males in rote memory, vocabulary and verbal fluency.

In another study, Feingold (1972) reported that males varied than females in quantitative reasoning, spatial visualization, spelling and general knowledge. Because these sex differences in variability were coupled with corresponding sex difference in means.

The difference between males and females on the Information Test (Table 23) may be attributed to the gender specific socialization practices of our society. Boys performed better on Information Test which may be attributed to the fact that they are generally more exposed to the daily happenings of common life and are more acquainted with current affairs.

The items of Vocabulary, Analogy and Comprehension Test are likely to favor the females as compared to males while Arithmetic and Information Test may favor the experiences of males than females (see Table 23).

The differences in the intelligence scores of the urban and rural background as shown in Table 24 may be attributed to their differences in social and educational background. Students belonging to rural area mostly belong to the lower socioeconomic strata and are afforded with limited educational facilities. Generally it has been observed that in rural area students are offered lesser opportunities particularly in co-curricular activities and other social / educational facilities that are considered necessary for personality grooming. Whereas, in urban areas students are provided with better opportunities as such conducive environments that facilitates development of better personality.

The student of rural background did better on Information test. This difference of performance may be attributed to the fact that rural background students are generally more exposed to the daily happenings and as such are more familiar with the current issues in the society as compare to students from urban areas that are more involved in indoor activities such as computer chatting, video games, more time they spend in front of

television, mobile messaging and so an. However, students from urban areas did better on the remaining sub-tests of IIT. The difference in result may be attributed to the fact that better educational facilities provide opportunities for development of better intellectual capacities and as such they performed better on Vocabulary, Arithmetic reasoning, understanding of Analogy, Comprehension and Similarity as compared to Urdu medium students (see table 24).

The result of Table 26 shows the mean difference between various income group students on the IIT score and subscales. Mean scores of IIT and all the subscales show progressive increase with increase in income range which demonstrates validity of IIT. The test scores do provide empirical evidence that IIT can differentiate well between various income groups.

Results also demonstrate performance of students from Punjab and Federal Area better on Vocabulary and Arithmetic scores as compared to students of other provinces. Whereas, students from Punjab, Federal area and AJK have higher analogy scores as compared to students from other provinces. Furthermore, students from AJK have performed better on Information test as compared to students form other provinces. Students from NWFP, Federal area, AJK and Punjab have shown better performance on Comprehension test as compared to students from other provinces. Students from AJK and Federal area have higher Similarity scores as compared to students from other provinces.

Performance of students from different economic background, results (Table 26) reveals that students from higher economic background have performed relatively better as compared to students from lower economic background. The result goes with the assertion better economic status allows better environments for learning and overall grooming.

To interpret the percentile norms, Tables 27-38, it is useful to describe a person as superior if his score lies at or above the 95th percentile for the normative group, above average when score lies at or above 75th percentile, average when his score lies between 25th – 75th percentiles, below average, when his score lies at or above 10th percentiles and defective when his scores lies at or above 5th percentile.

Province wise percentiles norms of IIT and its subtest, whereas results are shown from Table 34 to 38. Province wise norms are considered necessary to gauge true

potentials of candidates hailing from various provinces and as such diversified cultural background. Maximum score attained by a candidate on province Punjab scale is 63 and lowest were 34. Similarly, highest score attained by a student on province Sindh scale is 62 and lowest were 33. On Khyber Pakhtoonkhwa scale the highest and lowest scores were 62 and 31. In Federal Area scale both the scores were 63.2 and 31. AJK population managed to score 63 and 31.4 as highest and lowest score. Province wise norms are presented in tables from 34 – 38. Vernon (1969, 79) in his studies have asserted that such differences in culture and environments are to be taken care of while testing intelligence.

Limitations

In the beginning the researcher has experienced some difficulties due limited availability of local working models wherein indigenous Verbal Intelligence scale in English language has been developed. The developed test is intended to be used as a screening tool for candidates applying in Armed Forces, Central Superior Services and Public Service Commission examinations where English carries a status of an official language. Some indigenous work has been undertaken by local scholars for developing an intelligence scale. However, majority of the work has been carried out in Urdu language and some studies were found where a non verbal scale of intelligence has been developed for children below 17 years.

Keeping in view limited research work carried out locally for development of indigenous intelligence scale, the present study can be considered as a rare effort based on the scientific principles of test development.

Due prevalent fragile law and order situation excess to female colleges particularly in the province of Balochistan and Khyber Pakhtoonkhwa was difficult. Resultantly, male subjects outnumbered female participants during the process of data collection.

Due prevailing law & order situation excess to students studying in Government educational institutes of Federally Administered Areas was minimal therefore in any future effort to improve psychometric indices of IIT more population from these areas be included.

IIT be administered on the population studying in private sector educational institutions to compare and measure general intelligence in both on scientific lines.

IIT be also translated in Urdu language so that youth not conversant with English language can be provided with an opportunity to perform on the test as per their true potentials.

Future Recommendations

1. To provide additional support to the validity of the test, the sample size of the study can be enlarged by including more female subjects from both urban and rural areas belonging to different social, cultural and educational backgrounds.

2. In order to increase the reliability of the test, items found not discriminating well among high and low achiever may be re-written and some new items can also be included. Moreover, some other sub-tests of verbal ability can also be added in IIT.

3. Research study can also be carried out to validate test norms for population with lesser ages. Similarly studies can be conducted to develop separate norms for all the four provinces of Pakistan.

4. Cross-cultural comparisons can be carried out between the countries of South-Asia, like Bangladesh and India, where English language is spoken, read and understood.

5. To make IIT even better a tool to measure general intelligence '*g*' more subscales be included in any future study venture.

Implications

1. IIT being a tool to measure general intelligence '*g*' can be administered as a screening tool on a population interested to apply in Armed Forces, Public Service Commission and similar organizations where English as a language carries official status.

2. IIT can be used as a screening tool during admission process of educational institutes imparting professional education in English language for students of Intermediate and higher grades.

3. IIT can also be utilized during selection process to identify students with relatively better understanding of English language and interested to pursue higher studies in foreign countries.

REFERENCES

Ahmad, I. (1987). Initial psychometric evaluation of Urdu version of California Psychological Inventory (CPI). *Pakistan Journal of Psychological Research, 1* (1 -2), 3 – 16.

Ain, Q. (1985). A validation study of Cattell's Culture Fair Intelligence Test on Pakistani children. New Directions in Pakistani Psychology. *Proceedings of the 5th Conference of Pakistan Psychological Associations.* Z. A. Ansari; P. N. Tariq; R. M. Zaman; S. Chowdhri and S. Watanabe (Editors). Islamabad: 155 – 161.

Anastasi, A. (1967). Psychology, Psychologists, and psychological testing. *American Psychologist, 22,* 297-306.

Anastasi, A. (1980). Abilities and the measurement of achievement. *New Directions for Testing and Measurement, 5,* 1-10.

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

Anastasi, A., & Urbina, S. (2012). *Psychological testing* (7th ed.). PHI Learning Private Limited New Delhi-110001.

Ansari, Z. A. (1976). Creativity intelligence and academic achievement. *Pakistan Journal of Psychology, 9* (3-4) 17 – 26.

Ansari, Z. A., & Iftikhar, M. (1988). Validity of Raven's Standard Progressive Matrices for urban and rural school children in Pakistan, Part-1: Basic facts. *Psychology Quarterly, 19,* 14 – 27,

Benson. E. (2003). Intelligence across cultures. American Psychological Association. *Monitor* staff, *34*(2), 56. Retrieved from http://www.apa.org/monitor (feb 03) intelligent.html.

Berry, J. W., Poortinga, Y. A., Segall, M. H., & Dasen, P. R. (1992). *Cross Cultural Psychology: Research and applications.* New York: Cambridge University Press.

Binet, A. (1890). Perceptions d' enfants. *Revue Philosophique, 30,* 582 - 611.

Binet, A., & Simon, T. (1905). Mehodes nouvelles pourle diagnostic du neveau intellectual des anormaux. *L' Annee Psycholigique, 11,* 191-244.

Binet, A., & Simon, T. (1909). Intelligence des imbecilies. *L'Annee Psychologique, 15,* 1-147.

Bjorklund, D. F. (2004). *Children's Thinking: Cognitive Development and Individual Differences.* Stamford, CT: Wadsworth Publishing.

Bondy, M. (1974). Psychiatric antecedents of psychological testing (before binet). *Journal of the History of the Behavioral Sciences, 10,* 180-94.

Boring, E. G. (1923, June). Intelligence as the tests test it. *New Republic, 35,* 35-37.

Brislin, R. W. (1993). *Understanding Culture's influence on behavior.* Orlando, FL: Harcourt Brace Jovanovich.

Brody, Nathan (1992). *Intelligence Second Edition.* San Diego: Academic Press.

Brody, N. (2003). "Construct validation of the Sternberg Triarchic Abilities Test: Comment and reanalysis". *Intelligence* **31**: 319–329.

Bukhari. S., & Kamal. A. (2003). Development of an Indigenous Emotional State Scale for Dermatological Patients. *Pakistan Journal of Psychological Research,18,* (1-2), 59-79.

Burt, C. (1931). "The Differentiation Of Intellectual Ability". *The British Journal of Educational Psychology.*

Burt, C. (1971). The *gifted child.* New York: Wiley.

Campbell, D. P. (1965). A cross-sectional and longitudinal study of scholastic abilities over twenty-five years. *Journal of Counseling Psychology. 12,* 55-61.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, England: Cambridge University Press.

Cattell, J. M. (1885). Uber die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben. *Philosophische Studien 2,*635 – 650. Translated in A. T. Poffenberger (Ed.), James Mckeen Cattall: Man of science. Vol I: Psychological research. Lancaster, Pa Science Press, 1947.

Cattell, J. M. (1886a). Psychometrische Untersuchungen. *Philosophische Stdien, 3,* 452 – 492.

Cattell, J. M. (1886b). The time taken up by cerebral operations. Mind, 11, 222, 242, 377 – 392, 524 – 538.

Cattell J. M. (1890). Mental tests and measurements. *Mind, 15,* 373 – 381.

Cattell J. M., & Farrand, L (1896). Physical and mental measurement of the students of Columbia Univesity. *Psychological Review, 3,* 618 – 648.

Cattell, J. B., & Maxwell, S. E. (1979). Individual differences in cognitive abilities. *Annual Review of Psychology, 30,* 603 – 640.

Cattell, R. B. (1940). A culture free intelligence tests, Part I. *Journal of Educational Psychology, 31,* 161-179.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38,* 592.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 4-,* 153 – 193.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence; A critical experiment. *Journal of Education Psychology, 54,* 1 – 22.

Cattell R. B. (1965). *Higher order factor structures and reticular vs. hierarchical formulae for their interpretation.* In C. Banks & P. L. Broadhurst (Eds.) London: University of London Press.

Cattell, R. B., & Butcher, J. (1968). *The prediction of achievement and creativity.* Indianapolish Ind: Bob Merrill,.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action.* Boston: Houghton Mifflin.

Cattel, R. B. (1987). *Intelligence.* New York: Elsevier.

Ceci, S. J. (1994). Bioecological theory of intellectual development. In R. J. Sternberg (Ed.). *Encyclopedia of human intelligence.* New York: Macmillan.

Chaffee, J. W. (1985). *The thorny gates of learning in Sung China: A social history of examinations.* Cambridge University Press.

Conoley, J. C., & Kramer, J. J. (Eds.). (1989). *Tenth mental measurements yearbook.* Lincoln, NE: University of Nebraska Press.

Conoley, J. C., & Kramer, J. J. (Eds.). (1992). *Eleventh mental measurements yearbook.* Lincoln, NE: University of Nebraska Press.

Croker, L.,& Algina, J. (1986). *Psychometric methods.* (2nd ed.) New York: McGraw-Hill.

Das, J. P. (1994). Serial and parallel processing. In R. J. Sternberg (Ed.). *Encyclopedia of human intelligence.* New York: Macmillan.

Deary, J. I & Johnson, W. (2010). Intelligence and education: casual perceptions drive analytic processes and therefore conclusions. *International Journal of Epidemiology 39* (5), 1362 – 1369. Retrieved from http//ije.oxfordjournals.org/search.

Detterman, D. K. (1984). Understand cognitive components before postulating metacomponents, etc., part 2. *Behavioral and Brain Sciences, 7,* 289-290.

DuBois, P. H. (1970). *A history of psychology testing.* Boston: Allyn & Bacon.

Ebel,, R. L. (1965). *Measuring educational achievement.* Englewood Cliffs, J. J.: Prentice Hall.

Elliott, C. D. (1990). The nature and structure of children's abilities: Evidence from the Differential Ability Scales. *Journal of Psychoeducational Assessment, 8,* 376-390

Ertil, J. P., & Schafer, E. W. P. (1969). Brain response correlates of psychometric intelligence. *Nature, 223,* 421-422.

Eysenck, H. J. (1979). *The structure and measurement of intelligence,* New York: Springer-Verlag.

Eysenck, M. W. (1984). *A handbook of cognitive psychology.* London: Erlbaum.

Eysenck, H. J. (1986). Is intelligence? In: R. J. Sternberg & D. K. Detterman (Eds.). *What is intelligence? Contemporary viewpoints on its nature and definition.* Norwood, NJ: Ablex.

Feingold, A. (1992). Gender differences in mate Selection Preferences: *A test of the Parental investment model.* Psychological Bulletin, *112*, 125 – 139.

Franke, W. (1963). *The reform and abolition of the traditional Chinese examination system.* Cambridge: Harvard University Press.

Galton, F. (1833). Life and letters of Francis Galton by Karl Pearson, Vol. 1.

Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences.* New York: Appleton.

Galton, F. (1870). *Inquiries into human faculty and its development.* New York: Appleton.

Galton, F. (1879). Psychometric experiments. *Brain, 2,* 149-162.

Galton, F. (1888). *Natural inheritance.* London: Macmillan.

Gardezi, A. H. (1994). *Development of a non-verbal intelligence test for adolescents.* Unpublished M. Phil. dissertation. National Institute of Psychology, Quaid-i-Azam University, Islamabad, Pakistan.

Gardezi, A. H. (2001). *Development and Standardization of an indigenous non-verbal intelligence test.* Unpublished Ph.D dissertation, National Institute of Psychology, Quaid-i-Azam University, Islamabad, Pakistan.

Gardner, H. (1986). The waning of intelligence tests. In R.J. Sternberg & D. K. Detterman (Eds.). *What is intelligence? Contemporary viewpoints on its nature and definition.* Norwood, NJ: Ablex.

Gardner, H. (1993). *Multiple intelligence: The theory in practice.* New York: Basic Books.

Ginsburg, H., & Opper, S. (1988). *Piaget's theory of intellectual development* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Goodenough, F. L. (1949). *Mental testing: Its history, principles, and applications.* New York: Rinehart.

Gopnik, A. and Meltzoff, A. (1997). *Words, Thoughts and Theories.* Cambridge, MA: MIT Press.

Gottfredson, L. S. (1997). Intelligence and social policy. [Special issue] *Intelligence, 24,* 2 – 8.

Gregory, R. J. (1996). *Psychological Testing: History, principles and applications. (2nd Ed).* London: Allyn & Becon Inc.

Gregory, R. J. (2004). *Psychological Testing: History, principles and applications. (4th Ed).* India: Dorling Kindersley Pvt Ltd.

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

Guilford, J. P. (1985). The Structure-of-Intellect model. In: B. B. Wolman (Ed.). *Handbook of intelligence: Theories, measurements, and applications.* New York: Wiley.

Gutttman, L., & Levy, S. (1991). Two structural laws for intelligence tests. *Intelligence, 15,* 79-103.

Hassan, B., & Kamal, A. (2010). Development and Validation of the Piri-Muridi Scale. *Pakistan Journal of Psychological Research,25,* 1, 79-97.

Hanif. R., & Pervez. S. (2003). Translation and Adaptation of Teacher Stress Inventory. *Pakistan Journal of Psychological Research,18,* (1-2), 45-58.

Harman, H. H. (1975). *Modern factor analysis (3rd ed.).* Chicago: University of Chicago Press.

Harnquist, K. (1968). Relative changes in intelligence from 13 to 18. I Background and methodology. II. Results. *Scandinavian Journal of Psychology, 9,* 50-82.

Hasan, R. (1981). A comparative study of the Pakistani bilingual and monoglot school children's performance in verbal and non verbal test. *Pakistan Journal of Psychology, 12* (3-4), 3-12.

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

Horn, J. L. (1970). *Organization of date on life-span development of human abilities.* In L. R. Goulet & P. B. Baltes (Eds), *Life-span development psychology. Research and theory.* New York: Academic Press.

Horn, J. L. (1978). *Individual differences in cognitive development.* Final report. Betheoda, Maryland: National Science Foundation.

Horn, J. L. (1981). *Age difference in human abilities.* Final report. Bethesda, Maryland: National Institute on Aging.

Horn, J. L. (1982). *The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood.* In F. I. M. Craik & S, Trehub (Eds), Aging and cognitive process New York: Plenum Press.

Horn, J. L. (1985). Remodeling old theories of intelligence. In: B. B. Wolman (Ed.). *Handbook of intelligence: Theories, measurements, and applications.* New York: Wiley.

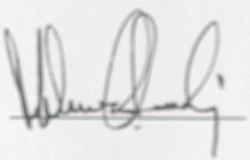
Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26,* 107-129.

Declaration

I, MAHMOOD SADIQ, do hereby solemnly declare that the work submitted in this thesis is my own, and has not been presented previously to any other institution or university for a degree.

This work has carried out and completed at the Department of Psychology, Peshawar University.

Researcher:



Humphreys, L. G. & Taber. T, (1970). Postdiction study of the graduate record examination and eight semesters of college grades. *Journal of Educational Measurement, 10*, 179-184.

Humphreys, L. G. (1971). Theory of intelligence. In R. Cancro (Ed.). *Intelligence: genetic and environmental influences.* New York: Grune & Stratton.

Humphreys, L. G. (1979). The construct of general intelligence. *Intelligence, 3*, 105-120.

Husen, T. (1951). The influence of schooling upon IQ. *Theoria, 17*, 61-88.

Hussain, S. S. (2001). *Development validation and standardization of a group verbal intelligence test in Urdu for adolescents.* Unpublished PhD dissertation, National Institute of Psychology, Quaid-i-Azam University, Islamabad, Pakistan.

Hynd, G. W., & Willis, W. G. (1985). Neurological foundations of intelligence. In B. B. Wolman (Ed.). *Handbook of intelligence: Theories, measurements, and applications.* New York: Wiley.

Imam, A., & Munsaf, S. (1988). Birth order and intelligence. *Pakistan Journal of Psychology, 19* (1 – 4), 27 – 35.

Irshad. E. (2011). Gender and Specific Learning Difficulties: A Double Disadvantage. *Pakistan Journal of Psychology, 42*, 1, 03-19.

Ismail.Z. & Mehmood, L. (1986). Social class and sex differences in Pakistan on the Raven's Progressive Matrices. *Pakistan Journal of Psychology, 17* (3-4), 23-25.

Israr, M. (1985). Cognitive development of Pakistan primary school children: Salient findings of a research project; New Directions in Pakistani psychology. *Proceedings of the 5th Conference of Pakistan Psychological Associations. Ansari, Z. A.; Tariq, P. N.; Zaman, R. M.; Chowdhri, S. and Watanabe, S. (Editors).* Islamabad: PPA, 67 – 5.

Jabeen, S., & Kausar R. (2010). Development of an Indigenous Obsessive Compulsive Disorder Scale for Pakistan. *Pakistan Journal of Psychological Research,25*, 1, 01-18.

Jackson, D. N., III., Vernon, P. A., & Jackson, D. N. (1993). Dynamic spatial performance and general intelligence. *Intelligence*, 17, 451-460.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Jensen, A. R. (1998). *The g factor.* New York: Praeger.

Kaufman, A. S., & Kaufman, N. L. (1993). *Manual for Kaufman Adolescent & Adult Intelligence Test (KAIT).* Circle Pines, MN: American Guidance Service.

Kausar, S., & Sheikh, H. (1986). Effect of Immediate vs delayed knowledge of results on subsequent performance in verbal learning. *Pakistan Journal of Psychology, 17* (1 – 2), 21 – 32.

Khalid, H. (1986). *Differences between primary students of government and private schools in cognitive development.* Unpublished M.Phil Thesis. Islamabad: National Institute of Psychology.

Khalique. N.; Khan. J. F. M.; Jahangir. A., & Iqbal. F. (2003). Development of Social Anxiety Scale and Social Confidence Scale: A Preliminary Report. *Pakistan Journal of Psychology, 34*, 45-58.

Khan, S. (1987). Relationship between ideational fluency and intelligence among academically gifted and average students. *Journal of Psychology, 18* (1 – 4), 17 – 28.

Keyser, D. J., & Sweetland, R. C. (Eds.). *Test Critiques* (Volumes I-VI). Kansas City, MO: Test Corporation of America.

Lai, T. C. (1970). *A scholar in imperial China.* Hong Kong: Kelly & Walsh.

Loona. M. I., & Kamal. A. (2011). Translation and Adaptation of Disruptive Behaviour Disorder Rating Scale. *Pakistan Journal of Psychological Research, 26*, 2, 149-165.

Lorge, I. (1945). Schooling makes a difference. *Teachers College Record, 46,* 483-492.

Luria, A. R. (1966a). *Human brain and psychological processes.* New York: Harper & Row.

Luria, A. R. (1966b). *Higher cortical functions in man.* New York: Basic.

Luria, A. R. (1973). *The working brain.* New York: Basic Books.

Malik. A.A., & Ismail. Z. (2005). Development of Social Support Scale. *Pakistan Journal of Psychology, 36*, 1, 03-30.

Matarazzo, J. D. (1992). Psychological testing and assessment in the 21st century. *American Psychologist, 47,* 1007 – 1018.

McGrew, K. & Flanagan, D. (1998). *The Intelligence Test Desk Reference: Gf-Gc cross-battery assessment*. Allyn & Bacon.

Meltzoff, A. N., & Moore, M. K. (1989). Limitation in newborn infants: Exploring the ranges of gestures imitated and the underlying mechanisms. *Developmental Psychology, 25,* 954 – 962.

Miles, T. R. (1957). Contributions to intelligence testing and the theory of intelligence. On defining intelligence, *British Journal of Educational Psychology, 27,* 153-165.

Mitchell, J. V. (Ed.). (1985). *Ninth mental measurements yearbook.* Lincoln, NE: University of Nebraska Press.

Murphy, K. R., & Davidshofer, C. O. (1998). *Psychological testing: Principles and application* (4th ed.). New Jersey: Prentice-Hall.

Naheed, G. (1993). *Development of verbal test of intelligence for Pakistani urban primary school children.* Unpublished M.Phil Thesis. Islamabad.: National Institute of Psychology.

Najam, N., Andrabi, N. & Ghazanavi, M. (1990). Relationship between intelligence, order of birth and some personality variables. *Journal of Behavioral Sciences, 1* (2), 21 – 32.

Nausheen. B., & Kamal. A. (2003). Familial Social Support Scale for Breast Cancer Patients: Development and Validation. *Pakistan Journal of Psychological Research, 18*, (3-4), 119-130.

Naz. M. A., & Siddiqui. S. (2010). Development of an Indigenous Depression Scale for Adolescent Schoolgirls. *Pakistan Journal of Psychological Research, 25*, 2, 115-130.

Neisser, U. (1979). The concept of intelligence. *Intelligence, 3*, 217-227.

Otis, A. S. (1918). An absolute point scale for the group measure of intelligence. *Journal of Educational Psychology, 9*, 238-261, 333-348.

Owens, A. W (1953). Slave Mutiny: *The revolt of the Schooner Amistad: Revolt of the Schooner Amistead:* An account of he Slave rebellion on the Schooner La Amistad ISBN 15747800 42.

Owens, W. A. (1966). Age and mental abilities: A second adult follow-up. *Journal of Educational Psychology, 57*, 311-325.

Paunonen, S. V & Ashton, M. C. (1998). Structured Assessment of Personality across Cultures. *Journal of Cross-Cultural Psychology 29* (1), 150-170. Retrieved from http//ije.oxfordjournals.org/search.

Pearson, K. (1914, 1924, 1930a, b). *The life, letters, and labours of Francis Galton* (Volumes I, II, III, IIIb). Cambridge: Cambridge University Press.

Pearson, K. (1982). *The grammar of science.* London: Walter Scott.

Pearson, K. (1984). Contributions to the mathematical theory of evolution. I: On the dissection of asymmetrical frequency curves. *Philosophical Transactions (Ser. A), 185*, 71 – 110.

Piaget, J. (1926). *Judgment and reasoning in the child.* New York: Harcourt, Brace & World.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International Universities Press.

Piaget, J. (1967). *Six psychological studies.* New York: Random House.

Piaget, J. (1972). *The psychology of intelligence.* Totowa New Jersey: Littlefield, Adams.

Piaget, J. (1980). *Adaptation and intelligence Organic Selection and Phenocopy.* Chicago: University of Chicago Press.

Pintner, R. (1921). Intelligence. In E. L. Thorndike (Ed.). Intelligence and Its Measurement: A Symposium. *Journal of Educational Psychology, 12,* 123-247, 295-216.

Riaz, M. N. (2008). *Test Construction Development and Standardization of Psychological Tests in Pakistan.* D.G. Administration, Higher Education Commission , Islamabad, Pakistan.

Riaz, Z., Yasien. S., & Khanam. S. J. (2011). Translation and Adaptation of Perceived Social Self Efficacy Scale. *Pakistan Journal of Psychology, 42,* 2, 79-92.

Ree, M. J., & Earles, J. A. (1991b). The stability of g across different methods of estimation. *Intelligence, 15,* 271-278.

Robinson, R. (1950). *Definition.* Oxford University Press.

Rogoff, B. (1984). What are the interrelations among the three subtheories of Sternberg's triarchic theory of intelligence? *Behavioral and Brain Sciences, 7,* 300-301.

Rohail. I. (2005). Development of a Problem Checklist for Pakistani Adolescents. *Pakistan Journal of Psychological Research,20,* (1-2), 65-79.

Schaie, K. W. (1983). The Seattle Longitudinal Study: a twenty-one year exploration of psychometric intelligence in adulthood. In K. W. Schaie (Ed.). *Longitudinal studies of adult psychological development.* New York: Guilford Press.

Schaie, K. W. (1985). *Manual for the Schaie-Thurstone Adult Mental Abilities (STAMAT).* Palo Alto, CA: Consulting Psychologists Press.

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

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

Spearman, C. (1927). *The abilities of man.* London: Macmillan.

Sternberg, R. J., Conway, B. E., Ketron, J. L., & Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology, 1,* 37-55.

Sternberg, R. J. (1984). *Handbook of intelligence.* New York: Cambridge University Press.

Sternberg, R. J. (1985a). *Beyond IQ: A theory of human intelligence.* Cambridge: Cambridge University Press.

Sternberg, R. J. (1985b). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Vernon, P. E. (1960). *Intelligence and attainment tests.* London: University of London Press.

Vernon, P. E. (1979). *Intelligence, heredity and environment.* New York: W. H. Freeman.

Vernon, P. A., & Mori, M. (1990). Physiological approaches to the assessment of intelligence. In: C. R. Reynolds & R. W. Kamphaus (Eds.). *Handbook of psychological and educational assessment of children: Intelligence and achievement.* New York: Guilford.

Vernon, P. A. (1994), Reaction time. In R. J. Sternberg (Ed.). *Encyclopedia of human intelligence.* New York: Macmillan.

Wechsler, D. (1938). *The measurement and appraisal of adult intelligence (4th ed.).* Baltimore: Williams & Wilkins.

Wechsler, D. (1939). *The measurement of adult intelligence. Baltimore,* MD: William & Wilkins.

Wigdor, A. K., & Garner, W. R. (1982a). *Ability testing: Uses, consequences, and controversies. Part I: Report of the committee.* Washington, DC: National Academy Press.

Wigdor, A. K., & Garner, W. R. (1982b). *Ability testing: Uses, consequences, and controversies. Part II: Documentation section.* Washington, DC: National Academy Press.

Yerkes, R. M. (1919). Report of the Psychology Committee of the National Research Council. *Psychological Review, 26,* 83-149.

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.

Zoofashan. T. (1982). *Computer model for statistical analysis of progressive matrices* Unpublished M.Sc. Dissertation, Islamabad: Department of Computer Science, Quaid-I-Azam University.

Appendix-A

Sampling plan adopted for the present study

Study	City	Sample
First Tryout	Islamabad	200
Time limit of the test	Karachi	50
Reliability	Total Sample	1669
Validity		
Norm		
	PUNJAB	712
	SINDH	458
	KHYBER PUKTONKHWA	126
	BALOCHISTAN	273
	FATA	55
	AJD	26

Appendix-B

Keys of the sub-scale of first try out

Answers	Items
	Vocabulary Test
a	2, 3, 4, 5, 8, 11, 12, 18, 22, 24, 26, 27, 29, 36
b	9, 14, 17, 23
c	1, 6, 10, 13, 15, 19, 20, 25, 30, 31, 33, 34, 37
d	7, 16, 21, 28, 32, 35
	Arithmetic Test
a	39, 44, 46, 48, 51, 56, 61
b	38, 42, 49, 55, 57, 60, 62
c	41, 45, 50, 53, 59
d	40, 43, 47, 52, 54, 58
	Analogy Test
a	64, 80, 86, 87, 91, 94, 97, 105, 109, 114, 121
b	66, 72, 74, 77, 84, 88, 89, 92, 99, 101, 104, 107, 113, 117, 119, 122
c	67, 69, 70, 78, 79, 81, 85, 93, 96, 98, 100, 103, 106, 108, 112, 115, 123
d	63, 65, 68, 71, 73,75, 76, 82, 83, 90, 95, 102, 110, 111, 116, 118, 120
	Information Test
a	130, 136, 142
b	125, 129, 131, 135, 137, 141, 146, 147, 148, 152, 153
c	126, 128, 132, 134, 138, 140, 143, 144, 149, 154
d	124, 127, 133, 139, 145, 150, 151
	Comprehension Test
a	156, 159, 164, 167
b	157, 160, 163, 168
c	162, 165
d	155, 158, 161, 166, 169
	Similarity Test
a	3, 14, 20
b	4, 5, 6, 8, 9, 13, 16, 17
c	1, 10, 11, 12, 15, 19, 21
d	2, 7, 18

Three indices obtained through item analysis of first experimental try out

Items	Difficulty level	Discrimination level
		Vocabulary Test
1.	.81	.28
2.	.87	.39
3.	.74	.40
4.	.83	.43
5.	.16	.43
6.	.83	.51
7.	.86	.37
8.	.63	.55
9.	.48	.21
10.	.54	.32
11.	.70	.56
12.	.79	.49
13.	.48	.41
14.	.49	.41
15.	.52	.12
16.	.75	.24
17.	.56	.47
18.	.43	.48
19.	.49	.65
20.	.77	.33
21.	.54	.64
22.	.44	.36
23.	.71	.56
24.	.38	.40
25.	.32	.11
26.	.50	.33
27.	.49	.57
28.	.60	.60
29.	.26	.20
30.	.44	.23

Cont....

31.	.25	.09
32.	.59	.59
33.	.27	.21
34.	.29	.27
35.	.27	.08
36.	.83	.37
37.	.46	.23
		Arithmetic Test
38.	..74	.38
39.	.70	.57
40.	.39	.53
41.	.81	.52
42.	.82	.47
43.	.64	.64
44.	.79	.52
45.	.72	.45
46.	.56	.57
47.	.68	.62
48.	.42	.62
49.	.71	.39
50.	.53	.49
51.	.65	.67
52.	.44	.49
53.	.51	.50
54.	.36	.51
55.	.39	.48
56.	.30	.34
57.	.43	.38
58.	.70	.66
59.	.32	.46
60.	.20	.01
61.	.48	.31
62.	.30	.31.

Cont....

Sternberg, R. J. (Ed.) (1985c). *Human abilities: An information – processing approach.* London: W. H. Freeman.

Sternberg, R. J. (1986). *Intelligence applied: Understanding and increasing your intellectual skills.* San Diego, CA: Harcourt Brace Jovanovich.

Sternberg, R. J., & Detterman, D. K. (Eds.). (1986). *What is intelligence applied? Contemporary viewpoints on its nature and definition.* Norwood, NJ: Ablex.

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence.* New York: Cambridge.

Sternberg, R. J. (1995). For whom the bell curve tallies: A review of the Bell Curve. *Psychological Science, 6,* 257-266.

Sternberg, R. J. (1997). Intelligence and life long learning. *American Psychologist, 52,* 1029-1139.

Syed, N. (1993). *Development of non verbal test of intelligence for Pakistani urban Primary school children.* Unpublished M.Phil Thesis. Islamabad: National Institute of Psychology.

Teng, S. (1942–43). *Chinese influence on the western examination system. Harvard Journal of Asiatic Studies, 7,* 267-312.

Terman, L. M. (1916). *The measurement of intelligence.* Boston: Houghon-Mifflin.

Thorndike, E. L. (Ed.). (1921). Intelligence and Its Measurement: A Symposium. *Journal of Educational Psychology, 12,* 123-147 and 195-216.

Thorndike, L. L. (1921). *The nature of intelligence.* New York: Harcourt.

Thurstone, L. L. (1921). Intelligence. In E.L. Thorndike (Ed.). Intelligence and Its Measurement: A Symposium. *Journal of Educational Psychology, 12,* 123-147 and 195-216.

Thurstone, L. L. (1925). A method of scaling psychological and educational test. *Journal of Educational Psychology, 16,* 433-451.

Thurstone, L. L. (1931). Multiple factor analysis. *Psycholgical Review, 38,* 406-427.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monograph,* No. 1, Chicago: University of Chicago Press.

Thurstone, L. L., Thurstone, T. G. (1941). Factorial studies in intelligence. *Psychometric Monographs, No. 2.* Chicago: University of Chicago Press.

Thurstone, L. L., & Thurstone, T. G. (1962). *SRA Primary Mental Abilities.* Chicago: Science Research Associations, 1949-1965.

Tuckman, B. W. (1975). *Measuring Educational Outcomes: Fundamentals of Testing.* New York: Harcourt, Brace Jovanovich, Inc.

		Analogy
63.	.61	.59
64.	.73	.34
65.	.74	.40
66.	.80	.50
67.	.46	.31
68.	.53	.52
69.	.70	.34
70.	.62	.56
71.	.44	.32
72.	.74	.38
73.	.71	.39
74.	.81	.52
75.	.48	.31
76.	.32	.46
77.	.54	.34
78.	.73	.34
79.	.33	.36
80.	.53	.39
81.	.54	.42
82.	.50	.35
83.	.73	.51
84.	.32	.46
85.	.43	.38
86.	.56	.54
87.	.30	.46
88.	.46	.35
89.	.54	.32
90.	.78	.68
91.	.75	.57
92.	.74	.65
93.	.50	.33
94.	.48	.31
95.	.72	.75

Cont....

96.	.66	.60
97.	.65	.57
98.	.46	.35
99.	.73	.68
100.	.40	.33
101.	.51	.33
102.	.56	.48
103.	.47	.58
104.	.42	.43
105.	.64	.77
106.	.40	.33
107.	.70	.69
108.	.53	.69
		Information
109.	.70	.64
110.	.81	.48
111.	.79	.47
112.	.70	.78
113.	.64	.72
114.	.70	.53
115.	.72	.64
116.	.74	.47
117.	.59	.55
118.	.72	.49
119.	.64	.58
120.	.68	.74
121.	.72	.65
122.	.72	.53
123.	.54	.50
124.	.57	.53
125.	.70	.66
126.	.72	.54
127.	.53	.60
128.	.23	.08

Cont....

129.	.47	.36
130.	.50	.72
131.	.58	.33
132.	.35	.37
133.	.58	.46
134.	.34	.12
135.	.49	.28
136.	.26	.05
137.	.61	.39
138.	.61	.59
		Comprehension Test
139.	.78	.72
140.	.70	.71
141.	.60	.57
142.	.65	.57
143.	.61	.59
144.	.49	.51
145.	.54	.36
146.	.66	.76
147.	.45	.64
148.	.81	.57
149.	.81	.54
150.	.41	.35
151.	.33	.36
152.	.52	.65
153.	.47	.62
154.	.45	.48
155.	.61	.39
156.	.12	.05
157.	.29	.00
158.	.40	.33
159.	.23	.24
160.	.28	.07
161.	.49	.57

Cont....

162.	.72	.69
163.	.53	.69
164.	.63	.68
165.	.56	.48
166.	.45	.48
167.	.61	.39
168.	.12	.05
		Similarity Test
169.	.58	.25
170.	.49	.39
171	.50	.50
172.	.61	.86
173.	.50	.39
174.	.54	.35
175.	.71	.68
176.	.86	.32
177.	.78	.16
178.	.53	.24
179.	.33	.24
180.	.38	.29
181.	.44	.64
182.	.80	.34
183.	.60	.69

Appendix-D

Keys of the sub-scale of IIT (Final Draft)

Answers	Items
	Vocabulary Test
a	2, 3, 10, 12
b	1, 6, 9
c	8, 11
d	4, 5, 7
	Arithmetic Test
a	13, 15, 17, 20, 21
b	23
c	18, 19,
d	14, 16, 22, 24
	Analogy Test
a	30, 33, 35, 37
b	32,
c	25, 28, 31, 34, 36, 38
d	26, 27, 29,
	Information Test
a	47
b	40, 44,
c	41, 42, 43, 46, 49
d	39, 45, 48
	Comprehension Test
a	55
b	50, 51, 52, 54, 57, 59
c	53,
d	56, 58, 60 61
	Similarity Test
a	63, 70
b	64, 65, 68, 71
c	62, 66, 69
d	67

INSTRUCTIONS

Read the following carefully:

1. This is a short test of observation and general understanding to find out how quickly and accurately you can answer a series of simple questions.

2. Some of the questions are easy and others are hard. However, no one is expected to do everything, just do the very best you can.

3. If you find any question difficult, leave it out and go on to the next. Return to it later if you have any time left.

4. When you finish one page only then move on to the next page.

5. The answers are to be written only on your answer sheet. Do not write anything or make any mark on this booklet.

6. In this test you will have very little writing to do. To all questions possible answers are given and these answers are mentioned with the help of a letter (A, B, C, D). All you have to do is to select the best answer and **shade** the corresponding letter (column) on the answer sheet against the number of the question with the help of a lead pencil.

EXAMPLES

Now we will do some example together so that you can understand the test.

Example 1. Solar energy is produced by?

	a. Water	b. Coal	c. Wood.	**d. Sun**

Example 2. Ice is to cold as tobacco is to?

a. Medicine	**b. Cigarette**	c. Smoke	d. Smell

Example 3. Which one of the following word is similar in meaning to the word HAZARD?

a. Danger	b. Tactic	c. Approach	d. Method

Example 4. National game of Pakistan is?

a. Football	b. Cricket	c. Volleyball	d. **Hockey**

Example 5. If an athlete covers 01 KM distance in five minutes then 06 KM he can cover in?

a. 20 Min.	b. **30 Min**	c. 35 Min	d. 25 Min

Example 6. Find the similarity among the following?

Internet – Telephone?			
a. Status Symbol	b. Requires computer	c. **Means of communication**	d. Liked by everyone

You will be given **30** minutes to do this test.
No questions are allowed during the test.

DO NOT OPEN THE BOOKLET UNTIL YOU ARE TOLD TO DO SO

VOCABULARY

1. ORGANISM
   a. Person b. Creature c. Birth d. Blood

2. CONSCIOUSNESS
   a. Awareness b. Aggression c. Nature d. Character

3. ADVANCE
   a. Progress b. Entrance c. Design d. Chart

4. INTERRUPTION
   a. Force b. Supply c. Power d. Break

5. MAGNITUDE
   a. Magnet b. Variety c. Product d. Size

6. AGREEMENT
   a. Confirm b. Contract c. Package d. Argument

7. ABUNDANCE
   a. Mass b. Bottom c. Load d. Plenty

8. NOURISHMENT
   a. Taste b. Tendency c. Diet d. Medicine

9. BASIC
   a. Need b. Important c. Support d. Desire

10. OFFICER
    a. Executive b. Office c. Educated d. Honor

11. DEFICIT
    a. Amount b. Full c. Shortage d. Profit

12. APOLOGY
    a. Regret b. Aggressive c. Control d. Oppose

ARITHMETIC

13. During stock-taking, a total of 342 washing soaps were found in three boxes. If 174 washing soaps are in the first box and 87 in the second box, how many will be in the third box?
    a. 81 soaps b. 342 soaps c. 67 soaps d. 87 soaps

14. If a man drills a hole 75 MM deep in a 120 MM thick cast-iron bar, how far is the drill from breaking through?
    a. 25 MM b. 75 MM c. 55 MM d. 45 MM

15. A girl works at a home for Rs 50 an hour. She worked 1½ hrs for 06 days.

How much money did she make?
a. Rs 450 b. Rs 425 c. Rs 440 d. Rs 400

16. If a car has averaged 28 miles per liter on a journey of 70 miles, how many liters of petrol did it use?
a. 3.5 litres b. 4.5 litres c. 5.5 litres d. 2.5 litres

17. A boy has Rs 100/ in pocket and wants to see a movie with his friend. If the tickets cost Rs 10.50 each and they both had cold drinks worth Rs. 12.75/ after the show, how much money was he left with?
a. Rs 66.25 b. Rs 66.10 c. Rs 66.20 d. Rs 66.30

18. If a spindle revolves at 600 revolutions per minute, how many times does it rotate in one second?
a. 36,000 revolutions b. 600 revolutions c. 10 revolutions d. 60 revolutions

19. A man has Rs 40,000/ which he wants to divide between his wife and daughter with a ration of 5 to 3. How much amount his daughter receives?
a. Rs 20,000/ b. Rs 10,000/ c. Rs 15,000/ d. Rs 25,00/

20. A shopkeeper has 1000 eggs, if 10% of the eggs are cracked how many eggs could he sell in the market?
a. 900 eggs b. 830 eggs c. 840 eggs d. 855 eggs

21. A factory has 250 workers out of which 170 are men. What percentage of the work force are women?
a. 32% b. 80% c. 28% d. 8%

22. If a cutter has 5 teeth in 01 cm and it is 02 meters long, how many teeth it has?
a. 100 teeth b. 2000 teeth c. 1500 teeth d. 1000 teeth

23. It takes an assembly line worker 15 seconds to perform a particular operation. At that rate how many such operations should be completed in an hour?
a. 04 operations b. 240 operations c. 24 operations d. 120 operations

24. If 10 men can dig 20 holes in 40 days. Then how many days' 20 men will take to dig 10 holes?
a. 25 days b. 20 days c. 15 days d. 10 days

ANOLOGY

25. Love is to hate as traitor is to?
a. Affection b. Partisanship c. Loyalty d. Conscience

26. If coward is to timid then courageous is to?
a. Name b. Support c. Appearance d. Soldier

27. Mason is to construction as tailor is to?

a. Architect b. Artist c. Brick d. Stitching

28. When justice is to court then ailment is to?
a. Corruption b. Criminal c. Hospital d. Zone

29. Culture is to civilization as man is to?
a. Class b. Neighbor c. VIP d. Mannerism

30. Detach means the same as?
a. Disconnect b. Length c. Piece d. Correct

31. Small is to size as volume is to?
a. Dark b. Heavy c. Mass d. Large

32. Contract means the opposite of?
a. Impact b. Expand c. Shrink d. Bridge

33. If yesterday is to Monday then Tuesday is to?
a. Today b. Tomorrow c. Past d. Don't Know

34. Elders are expected to be affectionate and children are expected to be?
a. Liking b. Argument c. Respectful d. Admiration

35. Barber is to hair-cut as laborer is to?
a. Construction b. Nut c. Screw d. Wrench

36. Vertical means the same as?
a. Flat b. Parallel c. Perpendicular d. Horizontal

37. Hold is to leave as against is to?
a. Support b. Remain c. Talk d. Run

38. Strong-willed means the same as?
a. Competent b. Brave c. Determined d. Moderate

INFORMATION

39. National Anthem of Pakistan was written by?
a. Allama Iqbal b. Faiz Ahmed Faiz c. Ahmed Fraz d. Hafeez Jhalandhri

40. Labor day is celebrated on which date?
a. 1st June b. 1st May c. 1st April d. 1st January

41. Capital city of Iraq is?
a. Damascus b. Tehran c. Baghdad d. Cairo

42. What is the numbers of Khulafa-e-Rashideen?

a. 02 b. 03 c. 04 d. 05

43. Who was the first Governor General of Pakistan?
a. Mountbatten b. Sir Syed Ahmed Khan c. Muhammad Ali Jinnah d. Liaqat Ali Khan

44. Mangla Dam is constructed on which river?
a. Sindh b. Jhelum c. Sutlaj d. Ravi

45. The color of holy rock "Hajra I Aswad" is?
a. White b. Red c. Green d. Black

46. Supreme Court of Pakistan is in which city?
a. Karachi b. Lahore c. Islamabad d. Rawalpindi

47. Noble prize was given to which Pakistani scientist?
a. Abdus Salam b. Munir Ahmed Khan c. Saleem u Zaman d. Abdul Qadeer

48. Which country exists in east of Pakistan?
a. Afghanistan b. Iran c. China d. India

49. The present constitution of Pakistan was formally approved in the year?
a. 1954 b. 1956 c. 1973 d. 1971

COMPREHENSION

50. Why doctor prescription is important to buy medicine?
a. To avoid smuggling. b. To avoid misuse of medicines. c. To bring down prices. d. To pay fee of doctor.

51. Velocity is also known as?
a. Loud b. Speed c. Distance d. Force

52. We wash clothes for one of the following reasons?
a. To look smart. b. To keep hygiene. c. To iron them. d. To dye them.

53. If by chance you find a lost National Identity card on the road side, what you will do?
a. Telephone the owner. b. Inform the Police. c. Post it in near Mail box. d. Inform any newspaper editor.

54. What you will do when you see a smoke and fire in the train?
a. Inform the driver. b. Pull the emergency breaks lever in the cabin. c. Jump out of the train. d. Try to extinguish fire.

55. A combination of various metals is called?

a. Element b. Mass c. Liquid d. Alloy

56. A and B are children of C. C is the father of A but B is not the son C. What is the relation of B to C?
a. Niece b. Sister c. Cousin d. Daughter

57. Which of the following instruments is used to measure atmospheric pressure?
a. Thermometer b. Barometer c. Radiator d. Voltmeter

58. What is the requirement of a legal marriage?
a. Permission of the parents. b. Inform the court. c. Inform the police. d. Written legal marriage contract.

59. If fourth of the month falls on Monday then what day will come on eighteenth of the month?
a. Tuesday b. Monday c. Wednesday d. Thursday

60. The democratic form of government is elected by?
a. Supreme court b. Senate c. National assembly d. The people

61. Why is the land generally expensive in cities than in rural areas?
a. Due to its importance. b. Due to the size of land. c. Due to quality of land. d. Due to high demand.

SIMILARITIES

62. Ship – Aero plane?
a. Have ample space b. Used by rich people only c. Modes of transport d. Requires passport

63. Death – Life?
a. Reality b. Gloomy c. Years d. Faith

64. Fasting – Prayer?
a. Pious b. Religious practices c. Holy Month d. Beliefs

65. East – West?
a. Opposite Direction b. Sun rises and Sets c. Two ends of sky d. Far away from each other

66. Oxygen – Water?
a. Normal b. Matter c. Life d. Earth

67. Cat - Lion?
a. Live in Jungle b. Hunt their food c. Pet animals d. From the same family

68. Nose – Ear?

a. Part of face b. Sense organs c. Beautiful body parts d. Opposite to each other

69. Wheat – Rice?

a. We eat them b. We cook them c. Major crops of Pakistan d. Common food

70. College – University?

a. Places of Knowledge b. Playground c. Principal d. Cities

71. Butter – Milk?

a. Gift of Nature b. Balanced diet c. Soft drinks d. Found in villages

Declaration

I, MAHMOOD SADIQ, do hereby solemnly declare that the work submitted in this thesis is my own, and has not been presented previously to any other institution or university for a degree.

This work has carried out and completed at the Department of Psychology, Peshawar University.

Researcher:

Declaration

I, MAHMOOD SADIQ, do hereby solemnly declare that the work submitted in this thesis is my own, and has not been presented previously to any other institution or university for a degree.

This work has carried out and completed at the Department of Psychology, Peshawar University.

Researcher: