Bahria Journal of Professional Psychology, January 2019, Vol. 18, No. 1, 11-34

Development of a Test of Nonverbal Intelligence for Youth

***Muhammad Imtiaz Chaudhry and Dr. Shazia Khalid**
Department of Psychology**,** Preston University Islamabad, Pakistan

Present study was designed to develop a test of nonverbal intelligence for youth in group settings for large scale testing and screening of human resource induction. Items were developed to minimize the effect of education and language proficiency. Items of the proposed test were constructed from figurative material and drawings of geometric symbols, shapes, designs and patterns. The theoretical foundation of the test was derived from the Cattell's (1963)concept of *Fluid Intelligence*. The test was divided into five subscales i.e. matrices, odd one out, similarity, series and analogies, and consisting ability areas (abstract reasoning, analogical reasoning, working memory, processing and perceptual speed, inductive & deductive reasoning). Research was conducted in three phases. In the first phase of the main study 200 items were developed and presented to subject matter experts (SMEs) for their opinion (qualitative analysis/ content validity). In the second phase, two studies were conducted. Firstly try out testing was conducted on a sample of *N*=100 students of Army Public College, Malir to assess the feasibility of test format, clarity and comprehension of items and test instructions. Item strength and distracter analysis were also carried out. Second study was conducted for item analysis on a randomly selected sample of *N*=300 intermediate level college students of Faisalabad. Item analysis was carried out to estimate the difficulty level, discrimination index and internal consistency (quantitative analysis). Items with difficulty level range between 0.30-0.70, discrimination index greater than 0.30 and item to total correlation values greater than 0.30 were retained. After data analysis 80 items were found fulfilling the set criterion for fair item selection. Items were arranged in relative order of difficulty and the final draft of the test was prepared. In the third phase of the main study, test taking time was estimated for each subtest and full test. The time limit for the test was decided to be 30 minutes. Current research provides a valid and reliable tool for the measurement of non-verbal intelligence in youth in a group setting.

Keywords: Intelligence, Non-verbal, test development, Psychometric, item analysis, difficulty level, discrimination level and internal consistency.

---

*Correspondence concerning this article should be addressed to Muhammad Imtiaz Chaudhry, Department of Psychology**,** Preston University Islamabad, Pakistan Email: mimtiazs5@hotmail.com

Nature of human intelligence has been discussed and debated for thousands of years, from the time of Plato and Aristotle (Reisberg, 2013). It holds a vital position because it gives understanding about individual capabilities, provides insight into why some individuals are better suited for certain jobs and why various psychological and educational interventions work for some people and not for others (Kaufman, Kaufman & Plucker, 2013;Thorndike, 2005).

Although testing movement has scientific concern to understand and describe intellectual functioning but it is largely motivated by very practical reasons to find the individuals who have requisite abilities to work efficiently in certain professions. The use of intelligence test for selection and recruitment purpose was modernized by US army during World War I &II. Army Alpha (for literate) and Army Beta (for illiterate) were developed and used extensively for screening of recruits (Anastasi & Urbina, 1997; Thorndike, 2005).

After the ensuing success of intelligence for screening of recruits the field of intelligence testing got momentum and got worldwide acceptance for proficiency testing (Cohen &Swerdlik, 2005). The intelligence test has great appealin personnel selection and induction.The development of standardized test and predictive validity in successful human resource induction gave boom to testing and evaluating movement(Gregory, 2007; Thorndike, 2005).

Group administration of IQ tests fostered its widespread use in schools, colleges, industry, and the military (Wasserman & Tulsky, 2005). Researches show that intelligence is the most effective predictor of individual performance at cognitively demanding professions (Gustafsson & Undheim, 1996; Gottfredson, 1997a).Munaf & Ghaus-ur-Rehanam(1996) found a positive relationship between intelligence and job performance.

Intelligence is an all-encompassing ability of human personality with myriad implications on the overall success of life outcomes, most importantly; academic achievements, professional

success, quality of life and social well-being. A critical challenge in the assessment of intelligence has been the level of educational background, language proficiency and age specification. Thus, in order to develop a culture fair and culture free test, nonverbal intelligence testing has been acclaimed as a psychometrically viable method in the field of intelligence testing and measurement.

Majority of the non-verbal intelligence tests like Raven Matrices(Raven & Court, 1998) are single dimensional, consisting of Matrices only. The proposed test is comprised of five subscales: *Matrices, Odd one out, Similarities, Series and Analogies*; each subscale comprising of geometric and figurative materials; symbols, shapes, designs, and patterns all relatively free from language proficiency, educational and cultural background. The multidimensionality improves content adequacy and helps to incorporate different and varied type of items to measure multiple cognitive dimensions of construct. Multidimensional also give face validity and reduces the probability of error of measurement associated with single dimensional test. The present study was, designed to develop a group nonverbal intelligence test for Pakistani youth between theage range of 15 to 24 years.

In Pakistan intelligence tests are used in education, clinical and personnel selection. Every year hundreds of thousands of candidates are assessed to ascertain their suitability and trainability for human resource induction. Armed forces have an advanced personnel selection system for the tri services and intelligence testing is essential for selection and recruitment(Gardezi, 2001; Hussain, 2001). Apart from forces intelligence tests are being applied in civil services competitive exams and educational setting. They also have great significance in clinical and job performance (Munaf & Ghaus-ur-Rehanam 1996).The use of intelligence testing is now considered mandatory in cognitively demanding professions and is used widely in personnel selection and human resource induction.

Conceptualization of Intelligence

There have been consistent efforts by researchers and psychologists to build a single definition of intelligence yet there are as many definitions as researchers (Kaya, Denle & Bulut 2012). The first major study on thedefinition of intelligence was conducted by the editors of the Journal of Education Psychology in 1921 and 17 leading scholars of that time contributed and provided several definitions of intelligence. The second effort was made in 1986 by 20 prominent scholars of the field (Snyderman & Rothman 1987).

Third and most recent effort to define and update the definition of intelligence was made in 1994; definition given by Gottfredson was signed by 52 scholars (Sternberg &Detterman, 1986; Gottfredson, 1997) and stated that: "*Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings "catching on," "making sense" of things, or "figuring out" what to do*" (Gottfredson, 1997b, p. 13).

Fluid intelligence is specific nonverbal intelligence which is related to the ability to deal with novelty, to adapt and thinking *fluidly* to a new, novel and unfamiliar problem. It is defined as the ability to reason and solve problems using new information without relying on previously acquired knowledge and skills (Cattell, 1963).

Theoretical foundations and literature review

McGrew and Flanagan (1998) after structural analysis of intelligence pointed three different research traditions in the development of intelligence testing. The psychometric or structural approach, the information processing approach and the cognitive modifiability approach. The psychometric approach is based on

data reduction techniques, the information processing approach focuses on analogy to conceptualize the information processed during problem solving tasks and the cognitive approachesrely on adaptable and changeable nature of intelligence focusing upon on dynamic assessment procedures.

The psychometric approach begins with the assumption that the nature of intelligence can best be studied by the ways people differ in their abilities. Spearman (1923) in his studies used the factor analysis approach and proposed the unitary view of intelligence. He has firm believe that a general factor *g*, plus one or more specific factors*s* can explain performance on intelligence tests. He explained that people who perform well on a cognitive test tend to perform well on other tests mainly due to the *g* factor. He further concluded that intelligence is a general cognitive ability that can be measured and numerically expressed.

Raymond Cattell first proposed the concepts of fluid and crystallized intelligence which was further developed by John Horn. The theory is primarily concerned with providing a developmental framework for the structure of intelligence for the assessment of adult intelligence. This theory suggests that intelligence is composed of different abilities that interact and work together to produce overall individual intelligence. He distinguished between the abstract, adaptive, biologically influenced cognitive abilities that he called fluid intelligence and the applied, experience based and learning enhanced ability that he called crystallized intelligence (Cattell, 1963).

In the late 1990's, McGrew and Flanagan (1998) proposed the integration of the Cattell-Horn-Carroll models, creating the CHC (Cattell-Horn-Carroll) model. This model consists of a multidimensional view of intelligence, with three cognitive levels. Before 1998, majority of intelligence battery tests measured only a few of the broad cognitive abilities. However, measures grounded in this approach provide a way for practitioners and researchers to assess a much wider range of abilities than prior intelligence

batteries. In a review of the influence of theory on intelligence testing, Floyd, Evans, & McGrew (2003) noted the breadth and flexibility of this model in the cognitive ability research especially in the development of learning disability interventions and mathematical achievement.

Fluid intelligence is the capacity to think and reason abstractly and solve novel problems, independent of any knowledge from the past. It is the ability to analyze problems, identify patterns and relationships that underpin these problems and the extrapolation of using underlying logic. It is necessary for all logical problem solving, e.g., in scientific, mathematical, and technical problem solving. Fluid reasoning includes inductive reasoning and deductive reasoning by building problem-solving strategies. Itcan be measured by tests such as analogies, classifications, and series completions. Crystallized intelligence is the ability to use skills, knowledge, and experience. It is fact based and involves knowledge that comes from prior learning and past experiences. It does not equate to memory, but it does rely on accessing information from long-term memory.

Fluid intelligence is grounded in physiological efficiency, and is thus relatively independent of education and acculturation (Horn & Cattell, 1966). Fluid intelligence is closely related to ones' inherent potentials to be creative and solve new and novel problems. According to Goswami (1992), it is an essential component of cognitive development. He further claimed that fluid ability is in essence Spearman's (1904, 1923) *g*. Fluid intelligence is nonverbal and culture free and is related to person's inherent capacity to learn and solve problems. Fluid intelligence is mandatory in tasks where adaptation to new situation is required. Decades of research on intelligence demonstrated that the best measures of *g* are associated with fluid intelligence (Carroll, 1993; Horn & Cattell, 1966, 1978;Jensen, 1998).

Performance on tests of fluid intelligence are known to predict many aspects of life, including educational and work achievement, and social well-being (Gottfredson & Saklofske,

2009; Neisser et al., 1996; Petrill & Wilkerson, 2000).The recent developments in cognitive sciences, biological science and neuroscience have also advanced our understanding of the neurobiological foundations of fluid intelligence as the measure of g based intellect (Barbey Colom, Paul & Grafman, 2014).

In the light of the above mentioned literature review *fluid abilities* are related to individual's ability to think and act quickly, solve novel problems, and encode short-term memories while *crystallized abilities* are related to use of knowledge and experiences stored in long term memory. Fluid intelligence involves the use of abstract reasoning and is relatively free from the education and language while crystallized intelligence involves reading, vocabulary and general knowledge.

The proposed test is multidimensional, has content adequacy and is free from the confounding variables of language, education and cultural backgrounds, therefore more suitable for Pakistani youth. Furthermore, the test can be administered in group settings for large scale and speedy human resource induction.

The main aim of the present study is the development of a new Group Nonverbal Intelligence Test for Pakistani youth between the age ranges of 15 to 24 years. The study intends to develop a psychometrically sound and standardized intelligence measure; consisting of five subtests: Matrices, Odd one out, Similarities, Series and Analogies, each comprising of geometric and figurative materials; shapes, patterns, symbols and designs. The test can be administered in group settings and has the potential for large scale testing, quick screening, and speedy human resource induction of youth from diversified lingual, cultural and educational backgrounds.

Method

The objectives of the present study were achieved in three phases. In the *first phase* item pool was generated, structure and format of the test was finalized with the help of subject matter

experts and psychologist working in the field of test development and assessment. *Phase two* of the study was dedicated to items analysis, response valence and finalization of test draft. In the *third phase*, test taking time was calculated for each subscale and full test. The details are as follows:

Phase I-Item generation and format of test

Item generation is the first practical-step towards new test construction. Test items were generated in multistage process. The theoretical foundation of test construction was derived from the concept of fluid intelligence. After relevant literature review and referencing of prominent tests, item generation of the test was carried out in line with APA guidelines. Major steps involved in process were as following:

Step 1-Identification of ability Areas. After thorough and extensive study of literature and reputed intelligence tests, major ability areas were identified with the help of subject matter experts and psychologists working in the field of intelligence. The ability areas selected for the proposed test include; reasoning, abstract thinking, figurative and spatial relationship, information processing and conceptual clarity, progressive sequential ability, perceptual speed, recognition and comprehension, classification, cognitive correlates, induction and deduction reasoning. On the basis of test content format of the test was finalized.

Step 2-Format of the test. Format of the test was decided to have multiple choice items. In this format test taker has to choose the correct answer from the given options. Multiple choice formats provide objectivity of marking and ensures uniformity and standardization of marking and scoring. Depending upon the structure and layout of items the proposed test items were divided into five subscales including: Matrices, Odd one out, Series, Similarity and Analogy items covering specified ability areas. This type of arrangement helps to incorporate different and varied types of items to measure multiple cognitive dimensions of construct. Multidimensionality also gives face validity and reduces the probability of error of measurement associated with single

dimensional tests. Each subtest consisting of; geometric symbols, shapes, designs and patterns relatively free from culture, gender and education. Test descriptions of the subtest are as follows:

1. ***Matrices.*** The matrices are best used to assess reasoning power, cognitive functioning and abstract reasoning ability. They require integration of multiple abilities such as visual coding, pattern detection, rule integration, similarity assessment and analogical transfer and problem solving. Matrices of ability measure usually contain a '*g'* factor. In the proposed format figures are arranged in three rows and three columns (3x3). The first two rows and first two columns have three pictures each but the third row and third column has only two pictures and the space for third picture at the lower end is left blank. There is set pattern or sequence to either move left to right row wise or vertically column wise. The test taker has to reason out the matrics pattern and figure out options that complete the pattern.

2. ***Odd one out.*** The odd one out uses popular items to assess the analytical ability, reasoning power, discrimination, association and relatedness ability. These type of items are used both for verbal and non-verbal items. In these format of questions a set of items are given which are related to each other except one which does not go with the others. For the proposed test each question has five figures which are related to each other in one way or the other except one which does not go with the others. The difference is generally due to similarity, position, reflection, rotation, shading, shapes, association or differences etc. The examinee task is to closely observe the figures and reason out the one which is different from the others.

3. ***Similarities***. Similarity question is used to analyze the reasoning power, comparative analysis and relatedness ability and multitasking capacity of the test taker. In the proposed test each similarity question has two set of figures; question figures and the answer figures. Question figures have two figures which have one or more features which make them similar to each other. Similarity

may be based on shading, outlines, and shapes, number of figures or the combination of some or all. The answer figures have four options with only one correct answer. The task is to select the answer figure which is most similar to question figures and write it's number in the answer sheet against the number of the question. These type of questions are not very common but are used for assessment of higher order and multitasking ability.

4. ***Series*****.** Series questions are used to assess the reasoning ability, sequential logic and continuity of relationship and perceptual speed. In series questions a series of shapes, symbols designs or patterns are shown in logical continuation and test takers have to find what will come next in the sequence. In the proposed test each question has four figures that have logical progression/ continuation. Every question is given with five answer options to select the answer option that follow the series.

5. ***Analogies***. Analogy questions are used to assess the inductive and deductive reasoning, encoding and mapping. Questions are generally written in A is to B as C is to '?' format. Analogy is set of items in which comparison is drawn between two things on the base of some definite relationship, resemblances and distinctiveness they share. The questions are presented in the prescribed format and the test taker task is to infer the definite relationship between the first two figures and apply the same analogy to select the answer.

Step 3-Item generation. Items of the test were generated by theory driven rational approach having empirical support under the guidelines given in manual of APA and criteria adopted by DuBois (1970). As the test was aimed to be a test of nonverbal (fluid intelligence) therefore items for the test were generated from geometric figures, shapes and designs and patterns. Due care was exercised to keep the items relatively free from language and education proficiency and cofounding cultural variables. At the initial stage, a manuscript of 200 items roughly double the items expected to be retained at the final stage were prepared. Items were divided into five subscales i.e., Matrices, odd one out, series,

similarity and analogy. The drawings for the items were created with the help of computer assisted program Auto Cad.

Items were developed as per the guidelines and standards of educational and psychological testing given by joint committee of the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME).The format and guiding principle of item generation and test preparation were as followings:

1. The test should be a nonverbal group test.
2. The format of the proposed test should be dichotomous.
3. The items of the test would be multiple choice items with random occurrence of correct responds in the items.
4. The items of test should be of figurative materials/ types.
5. The test has 5 subtest; including the items of Matrices, Odd one Out, Similarity, Series and Analogy categories.
6. Items included in the non-verbal test should be cultural-fair.
7. The test should be power cum speed test.

Step 4-Subject Matter Experts' opinion. In the next step subject matter experts (SMEs) were consulted. A panel of 6 psychologists having minimum ten years of experience in the field of intelligence testing, administration and application were consulted. 3 scholars were from the field of research and testing and 3 PhD scholars having expertise in the development of intelligence tests were consulted for their input, guidance and technical advice on the items. The items were put up to SMEs for ratings on the five point scale with one being the least desirable. Items with average rating of three or more were retained .The shortlisted items were further refined and in the light SMEs opinion.

Step 5-Preparation of first test draft. In the light of expert opinion and ratings 50 items were dropped and the first test draft consisted of 150 items (30 items per category). Booklets for each subtest with the general instructions for conducting the test

and specific instructions for each subtest were prepared both in Urdu and English languages to minimize the language effect. To make the instruction more comprehensive two examples for each subtest were also incorporated in the test instruction. Answer sheets and scoring keys for each subset was devised.

Phase 2- Item Analysis, response valence and finalization of test draft

Item analysis is an important step of test development. The main aim of item analysis was to check the usefulness of test items and discard the test items that do not come up to the requisite standard. In the second phase two studies were conducted.

Study –I. The details of the Study I of Phase II are given below:

Objectives. Objectives of the first study were as follows:

- To test the feasibility of test format.
- To discard items with ambiguous and unclear drawings.
- To find the difficulty level (too easy or too difficult) of newly constructed test items.
- To determine the response variance.
- To evaluate the clarity and comprehension of test instructions.
- To undertake the revision of items in the light of data and observation of the test takers.
- To check the appropriateness of answer sheets.

Participants. The participants of the study comprised of 100 intermediate level students of Army public school and college Malir, Karachi, Pakistan. The age limits were between 16-18 years (M = 17.6). The students were from different family and cultural backgrounds. The students of this institution were preferred as they provide a good sample due to reason that they come from different socioeconomic classes and cultural backgrounds. Convenient purposive sampling technique was used to collect the data.

Procedure. After getting the permission from school administration a pool of 150 items was tried out on N=100 college level students of army public school and college Malir. The

participants were briefed about the purpose of study and after their verbal consent the test was administered. The instructions were given both in Urdu and English language (written &verbal) with examples. Each subtest was administered separately on a small group of 20-25 students. They were encouraged to share any questions, queries or observations about questions lacking clarity and comprehension. They were also encouraged to feel free to give any suggestions regarding the test administration..

Each test sheet was marked with especially designed marking keys. The difficulty level was computed by dividing the number of test takers who pick the correct option (pass the item) divided by the total test takers who answered or responded to items. The discrimination index (D) of each test items was calculated by contrasting group method. The test takers scores were tabulated in highest to lowest descending order. The 27 % subjects from the highest and 27 % subjects from the lowest formed the upper and lower groups. Subtracting difficulty values of lower group for higher group yielded discrimination index for each item.

Results

The items with difficulty level more than 0.80 were labelled as too easy and less than 0.20 were labelled as too difficult and therefore, were discarded. Similarly the items with low discriminatory power (less than 0.20) were also discarded being not able to discriminate between the low and high achievers. Distracter analysis of remaining items was also carried out by tabulating the frequency for each answer option. The items with low valence were refined and modified accordingly. On the bases of data analysis 30 items were found too easy (difficulty level more than 0.80) and 7items too difficult (difficulty level less than 0.20). Data revealed that 23 items failed to discriminate between high and low achievers (discriminatory power less than 0.20). On the bases of data 54 items were found lacking requisite standards and thus discarded (Table 1).

Table 1
Details of items discarded and the total discarded and retained items in each subscale (N=100)

Subscales	Items discarded	Total items discarded	Item Retained
Matrices	3, 9, 13, 20, 23, 27	6	24
Odd one out	4, 5, 11, 12, 14, 15, 17, 18, 20, 22, 23, 25,	12	18
Similarity	1, 2, 4, 7, 8, 9, 18, 19, 20, 27	10	20
Series	5, 7, 10, 11, 12, 13, 15, 17, 18, 20, 22, 24, 29	13	17
Analogies	2, 4, 5, 11, 12, 14, 16, 17, 18, 19, 22, 26, 27	13	17

Table 1shows the serial no. of items for each subtest discarded during the analysis. 54 items were discarded and 96 items were retained for the further testing.

After that the shortlisted items were arranged in relevant order of difficulty and a second test draft consisting of 96 items was prepared for further testing. New sets of booklets, answer sheets and keys were prepared.

Study – II. The details of the Study II of Phase II are given below:

Objectives. Objectives of the second study were as follows:

- To ascertain the psychometric properties of the short listed test items.
- To find the difficulty level of test items.
- To find the discriminate index of test items.
- To estimate the internal consistency of test items.
- To find the response variance/ valance of test items.
- Revision, improvement and arrangement of items and response options.
- Get necessary data required for final selection of items.

Participants. The participants of the study comprised of *N*=300 government and private college level students of district Faisalabad. They belonged from different family backgrounds and attended mixed medium of instruction institutions. Data was

collected from Faisalabad, as it has a good blend of rural and urban population. Convenient purposive sampling technique was used to collect the data. The age range of the participants was between 17-20 years (M = 18.6).

Procedure

After getting the permission from authorities each subscale was separately administered on small groups of students from different colleges of Faisalabad. After seeking the consent of students the test was administered. The participants were briefed about the purpose of test and ensured that their test results will be kept confidential. The test instructions were given and examples were briefed were discussed. The students were asked to attempt all questions as quickly as possible and write their starting and closing time at answer sheets. A follow up session was also conducted to get the input of participants about the test. Scoring of the test sheets were carried out with the help of specifically designed scoring keys and data was put through systematic calculations and analysis to find difficulty level, discrimination power and internal consistency of all items. Distracter analysis of short listed items was also carried out for further refinement of answer options.

Results

Items with difficulty value more than 0.70 were discarded being too easy and items with difficulty value less than 0.30were discarded being difficult as per the criterion. The items with discriminatory power less than 0.30 were also discarded as they were discriminating between high and low achievers. On the basis of data analysis 13 items either too easy or too difficult (Table 2),11 items with low discriminatory power (Table 3)and 10 items with low item to total correlation (Table 4) were discarded. Over all 26 items found weaker in psychometric proposition were discarded. After the detailed analysis 80 items were found suitable to be retained for final test draft. The serial no. of items not coming up to the standards on the three indices of difficulty, discrimination and internal consistency were finally discarded (Table 5).

Table 2
Indices showing difficulty level range for the items of each subscale (N=300)

Range	Matrices	Odd one out	Similarity	Series	Analogy
.91-1.0	**0**	**0**	**0**	**0**	**0**
.81-.90	**0**	**0**	**1**	**0**	**0**
.71-.80	**1**	**2**	**1**	**2**	**2**
.61-.70	6	3	6	4	4
.51-.60	8	6	6	5	5
.41-.50	4	4	3	4	3
.30-.40	4	2	3	1	2
.21-.29	**1**	**1**	**0**	**1**	**1**
.11-.20	**0**	**0**	**0**	**0**	**0**
.01-.10	**0**	**0**	**0**	**0**	**0**

Note. **Bold**= Number of discarded items in each subscales

Table 2 shows that 9 items are too easy and 4 items are too difficult. The majority of the items fall in the 0.40 to 0.70 difficulty level range. Overall the results show that 13 items are either very easy or too difficult for set criterion therefore, discarded.

Table 3
Indices showing discrimination power of the items of each subscales (N=300)

Range	Matrices	Odd one out	Similarity	Series	Analogy
.91-1.0	0	0	0	0	0
.81-.90	0	0	0	0	0
.71-.80	1	2	1	1	3
.61-.70	3	3	2	3	3
.51-.60	4	5	3	5	3
.41-.50	9	3	7	5	5
.30-.40	4	3	4	2	2
.21-.29	**2**	**1**	**3**	**1**	**1**
.11-.20	**1**	**1**	**1**	**0**	**0**
>.10	**0**	**0**	**0**	**0**	**0**

Note. **Bold**= Number of discarded items in each subscales

Table 3 shows that 11 items are not discriminating adequately between the *g*=high achievers and low achievers. The discrimination power of most of the items fall in the range of .30 to .70.

Table 4

Indices showing item to total correlation for the items of each subscales(N=300)

Range	Matrices	Odd one out	Similarity	Series	Analogy
.91-1.0	0	0	0	0	0
.81-.90	0	0	0	0	0
.71-.80	0	0	0	0	1
.61-.70	1	0	3	2	2
.51-.60	2	3	5	3	3
.40-.50	9	6	3	7	5
.31-.39	7	7	4	3	4
.21-.30	**1**	**1**	**2**	**1**	**1**
.11-.20	**2**	**0**	**1**	**0**	**1**
>.10	**0**	**0**	**0**	**0**	**0**

Note. **Bold**=Number of discarded items in each subscales

Table 4 shows that items have moderate to high item to total correlation. Items with less than correlation coefficient value of .3 were discarded. Data reflects that 10 items found lacking consistency amongst them and thus discarded.

Table 5

Indices showing the items discarded and retained for the each subscales(N=300)

Subtests	Items discarded	Total Items Discarded	Items Retained
Matrices	2, 5, 6, 22	4	20
Odd one out	1, 5	2	16
Similarity	1, 3, 7, 9, 19	5	15
Series	3, 8	2	15
Analogies	2, 4, 9	3	14

Table 5 shows that the serial no. of items for each subtest discarded during the analysis. Total 16 items were discarded and 80 items were selected for final test.

The retained 80 test items were again rearranged in relevant order of difficulty from easiest to difficult. Resultantly minor adjustment in position of distracters was also carried out to ensure random occurrence of correct answers throughout the test. The booklets, answer sheets and scoring keys were accordingly prepared for final draft.

Phase 3-Test taking time for the subtest and the final test

Estimation of test taking time is very important for intelligence and achievement test for comparative performance evaluation. To estimate the test time a separate study was also conducted. The main objective of this phase was to determine the test taking time for the subtests and the final test.

Participants

The participants of this phase comprised of *N*=100 intermediate level students of Army public school and college Malir, Karachi. The age limits were between 17-18 years *(M* = 17.9). Convenient purposive sampling technique was used to collect the data. The data was collected from this institution as it hosts people from varied socioeconomic, educational, cultural and regional backgrounds and thus provides a mini representation of national sample.

Procedure

To determine the test taking time, final test draft was administered on the participants of *N*=100 candidates. Each subtest was administered in standardized way and candidates were asked to complete all items of the test as quickly as possible but enough time was given to complete all the items. They were told to write the test start time and finish time on the answer sheets. The time taken by each candidate was recorded. The time taken by each student was arranged from lowest to highest. The time taken by first 80 best finishers was averaged to set the time limit for each subtest and the total test.

Results

Results show that the time limits for different subtests were in the range from 4 minutes to 20 minutes. The maximum time was observed for matrices as the no. of items for this subtest are more than others. For other subtests the average time limit was calculated to be around five which is relatively stable over the 14 to 16 question limits. The average time for test was calculated to be in the range of 30 minutes approximately (Table 6).

Table 6

Time taken to complete the each subtest and the full test (N=100)

Category of Subtest	Minimum Time	Maximum Time	Average Time
Matrices	9 Minutes	20 Minutes	10 Minutes
Odds one out	4 Minutes	12 Minutes	5 Minutes
Similarity	4 Minutes	16 Minutes	5 Minutes
Series	4 Minutes	15 Minutes	5 Minutes
Analogies	4 Minutes	10 Minutes	5 Minutes
Full Test	25 Minutes	62 Minutes	30 Minutes

Table 6 shows maximum and minimum time taken for the completion of each subtest and the full test. The average test taking time for the full test was calculated to be30 minutes.

Discussion

Intelligence testing is an important and central step in personnel selection. Majority of the intelligence tests used in Pakistan are in English language which put students of other languages at disadvantage. The English medium tests create unnecessary anxiety in students of Non-English medium institutions and are also anxiety provoking for students of English medium institutions as they feel certain level of difficulty in interpreting the relationship in English. This situation warrants the

development of new test which is relatively free from the confounding variables of language and culture. To address this issue a new test of nonverbal intelligence was planned with the aim to be used in for the assessment of cognitive functioning of youth. The test was also developed to address the demands of large scale group testing and quick screening.

Present study was designed to address the critics associated with foreign language and adapted versions of foreign language tests. Efforts were made to keep the test free from the language proficiency and education loading. The test was especially designed for speedy human resource induction and personnel selection for military and other professions. The proposed test consists of 80 items divided into five subtests; matrices (20), odd one out (16), similarity (15), series (15) and analogy (14 items (Table 5). The format of the test was kept multiple choice and the test items were consisted of non-culture specific abstract geometric and figurative material.

The study was conducted in three phases i.e., item generation, item analysis and test taking time. In first phase initially 200 items were generated approximately double the numbers intended to be used for final test. The proposed test items were discussed with expert of the field for content and qualitative analysis. In the light of experts opinion 150 items were shortlisted and first test draft of five subtests with 30 items each was prepared.

In the next phase two separate studies were conducted for finalization of test items. The items were put through difficult analysis as too easy or too difficult items do not contribute in the overall effectiveness therefore, items with difficulty level over 0.30 and below 0.70 were discarded. Similarly an item which did not discriminate between high and low achievers was also discarded.

After data analysis 80 items were found suitable for inclusion in the final test. Items with difficulty level ranges between 0.30 and 0.70were selected as they were thought to be neither very difficult nor too easy for the test takers. The detailed analysis of the data shows that average difficulty level of all items was approximately 0.50. The difficulty level value of 0.50indicated that 50% of the group would pass the items and 50 % would fail to answer them. A value of 0.50 indicates that the

item has highest level of differentiation between test takers in a group (Anastasi, 1997).Data shows that majority of the items fall in the mid-range on the difficulty distribution index which indicated that the test items had varied difficulty level with majority of the items in the middle range of difficulty level between 41 to 60 (Table 2).Items of difficulty level value closer to 0.50 are considered more useful items for differentiating amongst test takers and thus preferred for test of ability and achievement (Anastasi,1997;Gregory, 2007; Thorndike, 2005).

The discrimination power of each test item was also computed with the differentiation method. Discrimination values indicate how well the items discriminate between the high and low scorers on the test. The discrimination index suggests that test differentiate between high and low achievers (Cohen & Swerdlik, 2005). For the test a discrimination value of 0.30 was set as cutoff point for retention of an item as it provides a good index to differentiate between high and low achievers (Anastasi & Urbina, 2011). The findings shows that discriminate values of the subtest and full test fall in between 0.41 and 0.60 in the middle range of distribution(Table 3).

Item to total correlation was also computed to find the internal consistency of test items and only those items were retained which were found to have a correlation coefficient index of 0.30 or above. Items to total correlation indicate that the test items are internally consistent and test is homogeneous and measure the some universe construct (Gregory, 2007).Distracter analysis was also conducted to find the strength of answer options. The items with low valence was refined, modified or replaced with the new one. The final test draft consisting of 80 items was prepared for further study of time limits.

In the third phase of test development, third study was conducted to decide the test time limit. The average time taken by 80% first finishers of each subtest was estimated and then summed up to decide the test time limit for full test. The time limit for Matrices was averaged at 10 minutes and 5 minutes each for other subtests: Odd one out, Similarities, Series and Analogies subtests. The total time limit calculated for the test was 30 minutes.

Conclusion

Testing and evaluation is a very important aspect of efficient human resource induction. Intelligence tests are constructed to assess learning potentials and predict future success. Looking at the development of tests in advanced countries, the work done in Pakistan is very limited. Considering the limitations of foreign origin and language based tests the present study is an attempt to develop a nonverbal test of intelligence for the youth. Efforts were made to keep the test relatively free from the confounding variables of education, culture and language. The test has great potential for application in education, military and vocational training and has an additional advantage to be used for large scale group testing and screening of individuals for speedy human resource induction.

Reference

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educationaland psychological testing*. Washington DC: American Educational Research Association.

Anastasi, A., & Urbina, S. (1997). *Psychological testing*. (7th ed.), Upper Saddle River, NJ: Prentice Hall.

Barbey, A. K., Colom, R., Paul, E. J., & Grafman, J. (2014). Architecture of fluid intelligence and working memory revealed by lesion mapping. *Brain Structure & Function, 219*, 485–494. Retrieved from http://dx.doi.org/10.1007/ s00429-013-0512-z

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

Cattell, R. B. (1963). The theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1-22. Retrieved from http://dx.doi. org/10. 1037/ h0046743

Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement, 15*, 139-164.

Cohen, R.J., & Swerdlik, M. E. (2005). *Psychological testing and assessment* (6thed). Boston: McGeaw-Hill.

Deary, I., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13–21.

DuBios, P. H. (1970). *A history of psychological testing*. Boston: Allyn and Bacon.

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell– Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school age years. *Psychology in the Schools, 40(2),* 155-171.

Gardezi, A. H. (2001). *Development and standardization of an indigenous non-verbal test for adolescents* (Unpublished Doctoral dissertation). National Institute of Psychology, Quaid-i-Azam University, Islamabad, Pakistan.

Goswami U. (1992). *Analogical reasoning in children*. Hillsdale, NJ: Lawrence Erlbaum

Gottfredson, L. S. (1997a). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24*(1), 13-23.

Gottfredson, L. S. (1997b). Why *g* matters: The complexity of everyday life. *Intelligence*, *24*, 79–132.

Gottfredson, L., & Saklofske, D. H. (2009). Intelligence: Foundations and issues in assessment. *Canadian Psychology, 50(3)*, 183-195. doi:10.1037/a0016641

Gregory, R. J. (2007). *Psychological testing*: *History, principles, and applications* (5th ed.). Boston: Pearson Education.

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8,* 179-203.

Gustafsson, J. E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. C. Berliner & R. C. Calfee. *Handbook of educational psychology* (pp. 186-242). New York, NY: Simon & Schuster Macmillan.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligence. *Journal of Educational Psychology, 57*, 253-270. doi: 10. 103 7/h 0023816

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

Jensen, A. R. (1998). *The g factor. The science of mental ability*. Westport, Connecticut: Praeger Publishers.

Kaufman, J., Kaufman, S., &Plucker, J. (2013*). Contemporary theories of intelligence*. *Oxford handbooks online.* Retrieved from http://oxford hand books. com/ view/ 10. 1093 /oxford hb/ 97801 9537 6746. 001.0001/oxfordhb-9780195376746-e-51.

Kaya, F., Delen, E., & Bulut, O. (2012). Test review: Shipley-2. *Journal of Psychology & Educational Assessment, 30*, 593-597.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR):Gf–Gc Cross– Battery Assessment.* Boston, MA: Allyn and Bacon.

Munaf, S., & Ghaus-ur-Rehanam, M. (1996). The relationship between intelligence and job performance. *Pakistan Journal of Psychology, 27*, 3-10.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S., Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51(2),* 77-101.

Petrill, S. A., & Wilkerson, B. (2000). Intelligence and achievement: A behavioral genetic perspective. *Educational Psychology Review, 12*, 185-199.

Raven, J. C., & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales*. Oxford Psychologists' Press.

Reisberg, D. (2013). *The Oxford handbook of cognitive psychology*. New York, NY, US: Oxford University Press**.** Retrieved from: http://dx.doi.org/10.1016/j.intell.2017. 06.001 Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist, 42*, 137-144. doi:10.1037//0003-066X.42.2.137

Spearman, C. (1904). General intelligence: objectively determined and measured. *American Journal of Psychology, 15*, 291-293.

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

Sternberg, R. J., & Detterman D. K. (1986). What is Intelligence? *Contemporary viewpoints on its nature and definitions* (pp. 19-21). Norwood, NJ: Cambridge University Press

Thorndike R. M. (2005). *Measurement and evaluation in psychology and education* (6th ed), New Jersey: Pearson Education Inc.

Wasserman, J. D., & Tulsky, D. S. (2005). The origins of intellectual processing. In D. P. Petrill, S. A., & Wilkerson, B. (2000). Intelligence and achievement: A behavioral genetic perspective. *Educational Psychology Review, 12*, 185-199