CHAPTER II

REVIEW OF RELATED LITERATURE

This study has been undertaken to find out if there is any significant relationship between students’ intelligence scores and their age, sex, and school achievement. The study also aims at studying the differences in the intelligence of the Yemeni secondary school students according to their age, sex and school achievement. In addition to that, one of the purposes of this study is to adapt the Otis-Lennon Mental Ability Test (OLMAT) on the Yemeni culture. For that the researcher has carefully studied the literature and review of researches related to the development of intelligence tests, relationship between intelligence and some related variables as school achievement, and researches related to sex differences in intelligence.

2.1. INDIAN STUDIES:

Ramalingaswami (1970) adopted the performance part of the Wechsler Adult Intelligence scale (WAIS) and collected data from 604 literate persons of Delhi belonging to both sexes between 15 to 45 years. The reliability of the total test ranged from 0.89 to 0.91 in the four age groups. Factor analysis of the subtests for the four
age groups and for both sexes was done separately. A single common factor identified as spearman ‘g’ was found in each analysis.

Trivedi (1970) standardized a ‘culture-free’ test of mental ability for Assamese high school students. The sample consisted of 1310 students of tenth grade. The battery consists of figure arrangement, similarity location, progressive matrices, analogy-based matrices and pattern perception subtests. Both Percentile and T-score norms were prepared on the basis of data collected from the study sample.

Chatterjee and Mukherjee (1971) constructed and developed a non-language test of verbal intelligence. No significant difference was found between non-language test and the two parts of a verbal test containing letter reasoning and verbal reasoning items. The correlation between letter reasoning and total scores on the non-language test was found to be quite high.

Ahuja (1971) constructed a battery consisting of six tests, namely, analogies, classification, disarranged sentences, opposition series and best answers. The reported test-retest reliabilities as well as certain validity coefficients, appear to be satisfactory. Age norms and class norms have been worked out separately for boys and girls because of significant sex differences.

Deshpande (1971) studied the sex differences on Raven’s matrices test, the sample consisted of boys and girls enrolled in class IV of primary schools of Nagpur. He found a significant sex difference in the average scores of boys and girls.
Sharma (1972) attempted to measure children’s intelligence through bicycle drawings. From a preliminary study of 417 drawings, a 75 point scoring scheme, with items of hierarchical difficulty levels was designed, which showed a self scoring consistency coefficient of 0.96. The test has been standardized on 2,863 children from 30 schools and age-norms have been derived. The test-retest reliability was found to be 0.82, and split-half reliabilities for different age groups ranged from 0.84 to 0.92.

Bhatt (1972) translated the different tests of WISC battery into Gujarati after eliminating or changing some items and administered the battery to 440 students from 12 schools in Ahmedabad. Both split-half and test-retest reliabilities were obtained. The coefficient was found to be high. Validity was established against criteria like scores on the Stanford-Binet Intelligence Test, Desai and Bhatt’s Group Test of Intelligence, Shah’s Non-verbal Test of Intelligence, school-marks and ratings. Factor analysis was also carried out.

Mohan (1972) studied Raven’s Standard Progressive Matrices (PMT) and a verbal test of general mental ability. The two tests were administered to 310 students (165 females and 145 males) of Punjab University. The correlation coefficient between PMT and a verbal test of general mental ability was 0.654, and the correlations between the two tests for males and females did not show any difference.

Rao and Reddy (1972) conducted a study entitled, “Can Raven’s progressive matrices test be shortened”. They administered the RPM under nine different conditions of testing to 1,260 higher secondary boys from age group of 13 to 17
years. A shortened version of RPM was also scored along with the standard one and it was found that the correlations were very high when the RPM was administered as a power test without any time limitation. On the other hand, administering the test with time limitations as a speed test yielded significantly lower correlations when compared to the above.

Maniam and Feroze (1973) developed a verbal reasoning test in Tamil for high school students which has parts. The first part consists of eclectic reasoning, mathematical reasoning, syllogistic reasoning and abstract reasoning tests. The second part comprised of scientific reasoning test, indirect test, and synthetic test. The split-half reliability of the total scores was found to be 0.74 and validity coefficient was 0.60 against school marks. Boys were found to be better in reasoning ability than girls.

Faroqi (1974) conducted a series of five studies using the progressive matrices test PMT and found that group testing yielded high scores on the average than the scores under individual condition. The shortened version of PMT calling for about half the time required for the standard version, yielded almost as useful an estimate of child’s intelligence level as the standard one. Sex differences on the PMT were found to increase with increasing age.

Dolke (1975) studied various psychometric properties of progressive matrices test PMT on the basis of data collected from 512 employees of four textile mills. The test-retest reliability after one month was found to be 0.80. The internal consistency
(KR-20) and split-half reliabilities were 0.67 and 0.73 respectively. The concurrent validity coefficient of PMT against scores on the GATB was found to be 0.55. The correlation between PMT and the criterion of occupational success based on merit-rating turned out to be 0.62. Some age-wise norms and occupational-norms were also derived.

Dolke (1976) conducted an investigation into certain psychometric properties of Raven’s Standard Progressive Matrices Test (PMT). An attempt was made to report the work on item analysis, factor analysis, reliability and validity of PMT. The sample included employees from technical and clerical occupations and university students. Their mean score on PMT was 34.15 with a standard deviation of 12.60. For the purpose of item analysis 370 subjects were randomly selected from the total sample of 521. The results revealed that most of the items in PMT were of medium and low difficulty. Many items had poor discrimination. On scaling the test on linear scale it was found that items were not properly spaced in terms of their difficulty. Factor analysis of the test showed that it does not have a relatively simple structure: it measures some other factors along with ‘g’. Reliability and validity coefficients of the test are statistically significant but not very high.

Pillai (1978) constructed and standardized a test in Tamil to measure the intelligence of children in the age group of 10+ to 15+. The study was also designed to relate intelligence to age, sex, grade, residence and the socio-economic status of pupils. For the final administration, 5,000 pupils were selected from thirty-four schools of one of the fourteen districts of Tamil Nadu, using stratified proportionate
sampling. The test included seven subtests: synonyms, antonyms, analogy, classification, mixed words, reasoning (verbal), and reasoning (numerical). The test-retest reliability was found to be 0.84 and the split-half reliability was 0.88. The content validity was considered on the basis of the various types of behaviors assessed by the subtests. Norms were determined for the total sample, grades and age groups.

**Mishra (1978)** conducted the study to assess factor-structure invariance of the Wechsler Adult Intelligence Scales (WAIS) across two developmental age groups, 18-19 and 25-34, in an ideal factor analytic condition. The WAIS, Differential Aptitude Tests, Culture Fair Intelligence Test, and Advanced Progressive Matrices were administered to 200 subjects (100 subjects in each group). Intercorrelation matrices of the twenty-four variables of both the age groups were subjected to principal component analysis and varimax rotations. Eight factors in each age group were retained following the criterion of positive generalizability. The retained factors in the 18-19 age group were identified as perceptual organization, perceptual inductive reasoning, language usage, associative thinking, conceptual verbalization, verbal reasoning, memory and speed of perceptual thinking. The factors in 25-34 age group were interpreted as language usage, associative thinking, perceptual inductive reasoning, memory, perceptual organization, speed and verbal comprehension. Identical or similar factors were matched across the age groups. Horn’s corrected version of pattern correlation was computed between the factor loadings of each matched factor pair for an invariance analysis. Seven of the eight factors could be
matched and the one unmatched factors pair considered to be the most dissimilar was excluded from the invariance analysis. Marker variables helped for the emergence of certain new factors and raised the magnitudes of WAIS communalities. The invariance analysis revealed that structural invariance could be established for a verbal factor pair (conceptual verbalization of 18-19 age group and verbal comprehension of 25-34 age groups). A complete non-invariance was evident for the memory factor; lack of structural invariance was found for the remaining six factors.

**Puhan** (1979) conducted a study intended to examine WAIS factor structures across 18-19 and 25-35 age groups. The WAIS subscores of one hundred 18-19 and one hundred 25-34 year olds were taken (Puhan, 1974). The raw scores of all the WAIS scales were correlated, and principal component factors were found and rotated to Varimax solutions. In the 18-19 age groups the rotated factors were identified as Verbal Comprehension, Freedom-from-distractibility, Perceptual Assembly, and Symbolic Reasoning. In the 25-34 age group they were interpreted as Perceptual Organization, Verbal Reasoning, and Concentration Speed. The invariance analyses were then performed to examine the structural stabilities of the similar factors. The structure of the Verbal Comprehension factor showed a striking similarity across the age groups. The structure of the Freedom-from-distractibility factor was found lacking invariance requirements. The rest were found as dissimilar factors. The results were discussed in the light of past observations.

**Chakraborty** (1979) also did the study with the major objective of providing an instrument to measure the general mental ability of children of age groups six to
ten (studying in Classes I to V) for solving the problems related to school admission, classification of pupils into homogeneous groups, class promotion, diagnosis, prognosis, educational guidance and research. A battery consists of six tests, stringing Bead Pattern, Picture Sequence, Picture Assembly, Object Profile, Block Design and Symbol Substitution. The standardization sample consisting of 1,000 children (200 from each grade level comprising 100 boys and 100 girls) was selected from 84 per cent of the total subdivisions of Manipur. The reliability coefficients of the battery and its tests were estimated by K-R formula and test-retest method. The reliability coefficients obtained were found to range from 0.66 to 0.91. From factorial analysis it was found that the battery contained only one factor in the six tests studied. The ‘g’ factor loadings and centroid factor loadings of the test varied from 0.75 to 0.87 and 0.76 to 0.87, respectively. The coefficients of correlation between the scores on the battery and scholastic achievement in English, Mathematics and Science were found to range between 0.40 and 0.56. The coefficients of correlation between the battery scores and the scores obtained by the students on WISC and Performance Test of Intelligence were 0.78 and 0.63, respectively. The content, concurrent, factorial, cross and predictive validities of the battery were satisfactory. Grade and age norms were established separately for boys and girls. Grade percentiles for different classes (from I to V), and age percentiles for different age levels (from six to ten) were computed for boys and girls separately.

Sheth (1979), adapted the Wecheler Adult Intelligence scale for Gujarati population by making changes in the items of WAIS, wherever necessary, and
standardized on the population of Ahmedabad city. Six subtests were verbal and five subtests were performance tests. The various subtests were about general information, general comprehension, arithmetical reasoning, similarities, digit span, vocabulary, digit symbol, picture completion, block design, picture arrangement and object assembly. The sample comprised of 400 adults including women. The raw scores of each test were converted into scaled scores using $M = 10$ and $SD = 3$. Reliability of the test was checked by test-retest and split-half methods, and were found to be very high. Validity of the test was determined by comparing its IQs with IQs on Desai-Bhatt, Cattell Culture-Fair tests and Raven’s Standard Progressive Matrices, and the correlations ranged from 0.37 to 0.90. The correlations of the subtests were factor analyzed by Hotteling’s principal axis method and centroid method and eight factors were extracted, the first of them being “g”.

Thakur (1979) designed the study, where the major objective was to construct and standardize a test to measure the general mental ability of students studying in Classes V to VIII of Assamese Medium high and higher secondary schools of Upper Assam. Seven subtests – logical selections, analogies, number series, synonyms-antonyms, proverbs, classifications and best answers – were included in the test. The final version of the test was administered to 3,039 boys and 2,243 girls, selected from classes V to VIII, by adopting the stratified random sampling technique. The reliability coefficients obtained by test-retest, split-half and rational equivalence methods for the entire sample and for different classes of boys and girls separately were found to range from 0.89 to 0.97. Content, construct and
concurrent avidity were established and the obtained coefficients were found to range from 0.41 to 0.88. Class norms and percentile norms were fixed for different grades of boys and girls together as well as separately; norms were also fixed for the entire sample irrespective of grades. Coefficients of correlation between intelligence and academic achievement were found to range from 0.41 to 0.50 for Classes V to VIII.

Desai (1980) conducted the study with the main objective of comparing the performance of different sub-cultures of Gujarat on Raven’s Standard Progressive Matrices, Cattell’s Culture Fair Scale 3, Desai-Bhatt Verbal Group Test of Intelligence and Bavasar Non-verbal Group Test of Intelligence. The sub-cultures were ‘big urban, small urban, semi urban, rural advanced, rural backward, rural muslim, tribal, and tribal institutional’. Eight schools, one from each sub-culture, were selected and one division of Standard VIII was selected from each school randomly. The sample consisted of 338 pupils, and the four tests mentioned earlier were the tools used for this purpose. Product moment coefficient of correlation and factor analysis by the principal factor method along with varimax rotation were the statistical techniques used for data analysis. Girls’ average scores in all sub-cultures were lower than those of boys in the respective sub-cultures but on Cattell’s Culture Fair Scale 3, they were not much different. The mean scores of boys and girls showed progressive decrease from more urban to semi urban, rural and tribal sub-cultures with only one exception that the rural advanced sub-culture showed better averages than the semi-urban group. Because of low norms on the Cattell’s Culture Fair Scale 3, the scope of comparability among various samples reduced to a great extent. Thus,
the Cattell’s culture fair scale 3 did not prove more useful than the verbal and non-
verbal tests used in the investigation. The differences in the average scores of the
verbal test of intelligence were found to be more pronounced than those on Raven’s
and Cattell’s tests. This supports Cattell’s finding that crystallized intelligence
assessed by verbal tests shows greater cultural difference than fluid intelligence
assessed by culture fair tests. Five factors were identified, “g” factor, verbal factor,
deeper reasoning, perception of relationship, and manipulation of correlates.

**Jehan and Ahmad** (1980) evaluated the effect of advantaged and
disadvantaged class on intelligence. Fifty advantaged and fifty disadvantaged children
were identified on the basis of Kuppuswamy’s socio-economic status scale. They
belonged to the age range of 5-9 years and grade level upto 3rd. Seguin Form Board
test of intelligence was administered on each of the Ss individually. The result
revealed that the mean I.Q. of the advantaged group of children was much higher
(107.1) than the disadvantaged group of children (82.7). Parental education,
occupation, income and the living condition of the child plays a significant and
important role in the intellectual development of a child.

**Nayar** (1980) carried out the study to construct and standardize a test for
assessing the social intelligence of adults. Test items were constructed in multiple
choice form with four to six responses likely to be made by the central figure in the
situations. The test consisted of two forms, M and N. The final forms, M and N were
administered to a representative sample of 1,200 and 600, respectively, belonging to
different age groups ranging from 20-40 years and above for purposes of validation.
The coefficients of reliability calculated by the split-half method and the Spearman-Brown prophecy formula for Forms M and N were 0.82 and 0.74, 0.90 and 0.85, respectively. The reliability coefficient between the forms M and N was 0.73.

**Patel (1981)** conducted the investigation with the main objective to develop a non-reading test of general mental ability for Gujarati speaking students of the higher secondary schools of Gujarat State. The standardization sample consisted of 5,725 students studying in the higher secondary schools of Gujarat State. The coefficient of reliability ranged between 0.71 and 0.87 by different methods. The test gave coefficients of validity against teachers’ ratings as 0.59, against examination marks as 0.52, against other tests of intelligence as 0.68 and 0.79. Factor loadings revealed that the test was heavily loaded with “g” factor. Age norms and grade norms were established and deviation-IQs and percentiles for the test were computed.

**Shah (1981)** constructed and standardized a spiral omnibus type group non-verbal test of intelligence for Grades VIII to XII, and the test was meant for the age group of 13 to 17 years. It included six types of nonverbal intelligence tests, namely, similarities, classification, analogies, series, conditions and matrices. The sample of the population of Gujarat was selected by including in it one school from an urban area and one school from a semi urban area from each district of Gujarat except the Danges. From the schools 3,612 pupils were selected by the method of random sampling. The reliability of the test estimated by test-retest method for different age groups ranged from 0.80 to 0.95. The test retest reliability for separate tests ranged from 0.60 to 0.80, the split-half reliability for different age groups ranged between
The validity of the test was obtained by correlating the test with Bhavsar Non-verbal Test, Desai-Bhatt Verbal Test, school examination marks and teachers’ opinions, which were 0.88, 0.90, 0.57 and 0.78, respectively. Factor analysis of the test was carried out by Hottelling’s principal axis method and seven factors were obtained, the first factor being common to all the six tests and hence was thought to be the “g” factor.

Banker (1981) constructed and standardized an abstract reasoning test. The final test was administered on 5,277 students of ninety-one different schools of fifty-nine different places of Saurashtra. Reliability of the test was established by test-retest method (0.81), split-half method (0.94), Rulon formula (0.94), and Kuder-Richardson formula (0.95). The three types of validity established were congruent validity ($r = 0.84$), concurrent validity ($r = 0.63$) and predictive validity ranging from 0.62 to 0.72. Separate norms were established for boys and girls of Grades VIII and IX, in the form of percentile ranks, standard scores, T scores, Stanines and letter grades.

Joshi (1982) adapted the Wechsler Preschool and Primary Scale of Intelligence for the children of Ahmedabad city in Gujarat. The original Wechsler scale consists of six verbal subtests, namely, information, vocabulary, arithmetic, similarities, comprehension and sentences, and five performance subtests, namely, animal house, picture completion, mazes, geometric design and block design. In the adaptation, some changes were made to suit the tests to Gujarati children. The tests were administered to thirty-seven children of age group four to six and a half years.
for item analysis, fixation of time limit and development of the scoring scheme. One hundred and eighty boys and 180 girls were tested with the Gujarati adaptations of WPPSI and their verbal performance and total scale IQs were calculated. Conversion tables for all these were prepared. The reliability of the scale was determined by split-half technique for all different age groups and also for individual tests which ranged from 0.28 to 0.94, by the test-retest method, the reliability ranged from 0.63 to 0.93. The validity of the scale was determined by correlating the WPPSI IQs with the Stanford-Binet IQs, Draw-a-Man Test IQs, school marks and teachers’ ratings which ranged from 0.26 to 0.96. The scoring key and the tables of conversion of raw scores into scaled scores and then to IQs were presented in the manual.

Chatterji (1983) made a study which aimed at finding the academic group differences in intelligence among intermediate college students. Jalota’s (1976) Hindi version of the Group Test of General Mental Ability (1972) was administered to a sample of 760 students of class XII belonging to four academic groups viz., Arts, Science, Commerce and Agriculture. The study revealed that students of Science group stood highest on the intelligence scale and students from Arts faculty were at the lowest end. There was no significant difference between the intelligence level of commerce and Agriculture groups, but each of them was significantly more intelligent than Arts Group.

Mohan and Bhatia (1985) conducted a study, where the purpose was to study psychomotor performance in children as a function of intelligence and sex. Two simple psychomotor tasks- Backward Figure Writing (BFW) and Tapping were used
on 60 children belonging to the age group of 10 to 14 Years. Three levels of intelligence; gifted, normal and mentally retarded were taken, each group consisting of 20 children (10 boys and 10 girls). Each subject worked for 5 minutes in the pre-rest period, rested for 1 minute and again worked for 2 minutes in the post-rest period on both the psychomotor tasks. Results indicated that in both the pre-rest and the post-rest periods, the subjects with higher intelligence consistently showed a better performance on BFW and tapping. Sex emerged to be a significant determiner in pre-rest BFW, pre-rest and post-rest tapping.

Chatterji (1986) developed a non-verbal intelligence test for the children from the age level 13 to 15 years of class IX and class X. The test consisted of items presented through geometrical figures and were of the multiple choice type. Both the difficulty and discrimination indices of the items were calculated. Most of the items included in the test were within the satisfactory range of difficulty values and had sufficiently high discrimination. Split-half reliability, after correction was 0.95. Concurrent validity with external criterion was 0.62.

Rangari (1987) studied the intelligence of the tribal and the non-tribal students of eleventh and twelfth standards. The sample consisted of 217 tribal and 255 non-tribal students selected from Arts, Science and Commerce colleges in Aurangabad city in Maharashtra State. Nafde’s Non-Verbal Test of intelligence was used to measure the subjects’ intelligence. The results indicated that the non-tribal students as a total group were significantly higher in intelligence than the tribal students. Moreover, the non-tribal males and the non-tribal urban students were
higher in intelligence than their tribal counterparts. But, the non-tribal females and the non-tribal rural students did not differ significantly from their tribal counterparts.

Shamshada (1988) attempted to compare boys and girls with regard to intelligence, neuroticism, scholastic achievement and need achievement. The sample comprised of 1,008 students covering equal number of boys and girls, who were drawn from Srinagar, Sopore, Baramullah and Anantang. The relevant data was collected using Tandon’s Group Test of General Mental Ability, Dadherjee’s Incomplete Sentences Blank and Annual Examination Marks. The collected data was treated using mean, SD and ‘t’ test. The major finding was that girls were superior to boys in intelligence and scholastic achievement.

Mahnaz (1994) studied intelligence and an academic achievement of tenth Grade students of east and west part of Pune city with particular reference to the socio-economic background. The sample consisted of 1946 students from 39 schools, and the Kulman –Anderson intelligence test was used to measure the intellectual level of the students. The ANOVA test results indicate that there is not much difference in intellectual abilities of students of different type of schools. The results show that there is a negligible difference in IQ mean scores of boys and Girls. The correlation results show a low and insignificant correlation between academic achievement and IQ (r = 0.26).
The following conclusions can be made on the basis of the above review:


The third part studied the differences in intelligence according to some variables as those of Deshpande (1971) studied the sex differences on Raven’s Matrics test. Desai (1980) compared the performance of different sub-cultures of Gujarati. Jehan and Ahmad, (1980) evaluated the effect of advantaged and disadvantaged class on intelligence, Chatterji (1983) studied the academic group differences in intelligence, Mohan and Bhatia (1985) studied psychomotor performance in children as a function of intelligence and sex, Rangari (1987) studied the intelligence of the tribal and the non-tribal students, Shamshada (1988) compared boys and girls with regard to intelligence, neuroticism, scholastic achievement,
Mahanaz (1994) studied intelligence and an academic achievement of tenth Grade students of east and west part of Pune city.

The findings showed that:

1. There were significant differences in intelligence score between students in different ages according to the studies conducted by Ahuja (1971), Sharma (1972), Dolke (1975), Pillai (1987), Chakraborty (1979), and Patel (1981).

2. There were differences between boys and girls in the average of intelligence scores according to the studies conducted by Ahuja (1971), Deshpande (1971), Faroqi (1974), Desai (1980), and Shamshada (1988). But the study of Feroze (1973) shows significant differences in reasoning ability only and Mohan (1972) shows no significant differences between boys and girls.

3. Thakur (1979) and Chatterji (1983) commented on the relationship between intelligence test and academic achievement. But Mahanaz (1994) found an insignificant correlation between academic achievement and IQ.

2.2. ARABIAN STUDIES:

Malkaw (1979) developed group intelligence test for ages 4-9 years that assumed three basic abilities: verbal ability, numerical ability, and logical thinking ability. Factor analysis showed loading on two factors, viz, verbal factor and logical factor. Reliability was established by test-retest, and split-half methods. The
reliability coefficients were 0.93, 0.95, respectively, validity of the test was established by correlation with school achievement, factor analysis, and differences in performance between levels of age by ANOVA.

**Yaqoub** (1979) conducted a study to determine the psychometric properties of Arabic version of Otis-Lennon mental ability test Elementary level form (J). The sample consisted of 486 students. Three indices of validity were obtained: concept validity, concurrent validity, and test homogeneity. The reliability of the test was obtained by (KR-20) formula. Its values ranged from 0.80 to 0.84. According to the test-retest method, its values ranged from 0.72 to 0.90. The third reliability index was obtained by split half method, and its values ranged from 0.76 to 0.87.

The objective of **Habahibh’s** study (1986) was to develop group intelligence test for Jordanian children in the age 12 – 15 years, in order to measure mental ability. The intelligence test consisted of seven subtests, word classification, arithmetic, number series, synonyms, antonyms, analogy and opposites. The sample consisted of 400 students, the validity was calculated by: two way ANOVA, percentage of successes, correlation between total score and school achievement criterion, and factor analysis. From factor analysis it was found that the test was loaded with three factors, verbal factor, numerical factor and reasoning factor. The split-half reliability coefficient ranged from 0.84 to 0.90, the test-retest reliability coefficient was 0.92.

**Yaqoub** (1988) conducted a study, where the purpose was to study the psychometric properties of the Otis-Lennon mental ability test elementary I level
form (K). The sample consisted of 576 students. The following validity indices were obtained: Concurrent validity, Discriminant validity and Homogeneity test. The reliability of the test was obtained by test-retest method, its values ranged from 0.77 to 0.87. By the split half method, its values ranged from 0.77 to 0.86. The third reliability index was obtained by Cronbach Alpha formula. Its values ranged from 0.83 to 0.89.

Amorah (1991) conducted a study entitled, “Comparative study of psychometric properties of Otis-Lennon mental ability test elementary I level form J and K”. It aimed at comparing the test validity and the test reliability for the two forms J and K. The sample consisted of 217 students. The validity was obtained by three methods: concurrent validity, discriminate validity, and construct validity. All values were significant at less than 0.01 level. The reliability was obtained by test-retest method and Cronbach Alpha formula. The values by test-retest method were 0.71 for form J and 0.77 for form K, and by Cronbach Alpha, 0.84 for form J and 0.88 for form K.

Korashy (1995) conducted a study, where the Rasch Model was applied to the selection of items for a mental ability test. The major objective was to apply Rasch measurement principles to the selection of items for an Arabic version of the Otis-Lennon Mental Ability Test Advanced level form J for secondary and university students in Kuwait. The sample consisted of 599 male and female students from various grades in secondary schools and the university. Fifty items were selected for the final form of the test. The reliability and validity for each item and for the total
test were calculated. The item separation of the test reached 6.19 with a reliability of 0.97, which means that these items were very well discriminated by the study sample. The item validity was counted by three ways. The first, the mean square residual, MNSQ, is a Rasch statistic that corresponds in meaning to the item point-biserial correlation in classical models. The second, the item-fit statistic, and the third, the item outfit. These results support the evidence of the validity of the test.

Al-Kofahi (1997) conducted a study entitled “The psychometric properties of an Adapted Version of Otis–Lennon Mental Ability Test Intermediate level form J”. The aim of this study was to determine the psychometric properties, difficulty and discrimination for the test items. In addition, it aimed at finding the test validity and reliability indices and driving norms for adapted Otis-Lennon mental ability test intermediate level form J on the Jordanian environment. The OLMAT was administered to 390 students (216 males, 174 females) in grade 7th, 8th, and 9th. The concurrent validity for three grades was 0.79, 0.81, and 0.86. The reliability obtained by (KR-20) formula was 0.96 and by test-retest was 0.91. As for norms, the percentile ranks were obtained for the raw scores and the deviation IQs for the different levels of ages and grades.

The above findings reveal that

1- There is no study conducted in Yemen to develop intelligence test or study the relationship of intelligence test with some variables.
2- The Arabian studies can be divided into two parts, the first part deals with development tests as those of Malkaw (1979) and Habahibh (1986) and the second part deals with adaptation of foreign tests to use in Arabic Language as of those of Yaqoub (1979), Yaqoub (1988), Amora (1991), Korashy (1995) and Al-kofahi (1997).

3- The results of Arabic studies indicate that the OLMAT is suitable for Arabic culture and has good reliability and validity.

4- The findings show differences in intelligence test between age levels as those of Malkaw (1979) and Al-Kofahi (1997).

2.3. STUDIES FROM OTHER COUNTRIES AND CULTURES

Proger et al (1971) studied the relative predictive and construct validation of the Otis-Lennon Mental Ability Test (OL-MAT), The Lorge-Thorndike Intelligence Test (L-TIT), and The Metropolitan Readiness Test (MRT) in grades Two and Four. The sample consisted of 322 students from second grade and 316 students from fourth grade. The relative predictive and construct validities of the Oits-Lennon, Lorge-Thorndike, and Metropolitan readiness tests were investigated by several multivariate analyses: canonical correlation, factor analysis, and stepwise regression. Canonical correlation and principal-components factor analysis were employed to study the factorial construct validity of the O-L MAT, L-TIT, and MRT. The results demonstrated some similarities and differences between the two analytical approaches to construct validation. Further, stepwise multiple regression was used to
establish the relative predictive validities of the three tests, in selected verbal and numerical areas. In brief, the O-LMAT appears to be at least as effective a predictor of verbal and numerical achievement, as measured by the Stanford Achievement test (SAT), and the use of teacher rating (TR) as is L-T IT and MRT.

Another research was conducted by Estabrook (1984), under the title “A Canonical correlation Analysis of the Wechsler Intelligence Scale for Children-Revised and the Woodcock-Johnson Tests of Cognitive Ability in a sample referred for Suspected Learning Disabilities”. The sample consisted of 107 boys and 45 girls in grades 1 – 7. A canonical analysis was completed to examine the overlap between the two tests. Three significant canonical correlations were obtained. The redundancy index showed that approximately 28.6% WJSCA subtest variance is predictable from the linear combination of the WISC-R subtests, and approximately 32.7% of the WISC-R subtest variance is predictable from the EJTCA subtests. Analysis of the structure correlations indicates that the first canonical variants share a general intelligence factor, the second a perceptual speed factor, and the third a numerical-memory factor.

The other study carried out by Antonak et al. (1982), is titled Otis-Lennon Mental Ability Test (OLMAT), Stanford Achievement Test (SAT), and Three Demographic Variables (sex, school attended, and years in district) as Predictors of Achievement in Grades 2 and 4. The sample consisted of 91 students from second grade classes in 1977-78, who were also presented as forth grades during the 1979-80 academic year, and 103 students from the second grades in 1979-80. A series of
multivariate statistical analysis was used to determine the relationships that exists among these three variables. The results revealed that the best predictor of achievement at either grade 2 or grade 4 was the I.Q. variable alone. The correlations among the OLMAT-IQ and the SAT variables were strong and positive for all pupil samples. The correlation between the OLMAT scores for the sample of 91 students when tested as second graders in 1977-78 and again as fourth graders in 1979-80 was 0.74. The correlations among the SAT subtests led to questions concerning the test’s construct validity and its use for differential diagnosis of educational problems of children. The nonsignificant sex differences in both the mean SAT and OLMAT IQ scores were found. A detailed study of the correlations between the OLMAT and SAT did not support the continued use of the group IQ test as part of the district’s comprehensive testing program.

The other study presented by Grossman et al (1983) under the title of “Validity of the Slosson and Otis-Lennon in predicting achievement of gifted students” studied the efficacy of the Slosson Intelligence Test and the Otis-Lennon Mental Ability Test to predict academic achievement, as measured by selected subtests of the Stanford Achievement test (SAT). The sample consisted of 46 children of middle-class origin (24 males and 22 females) from a Midwest, urban public elementary school district. The results of a multivariate multiple regression analysis indicated that the Slosson and Otis-Lennon significantly predict SAT Vocabulary, Reading Comprehension, and Mathematical Concepts subtests, with the Otis-Lennon in comparison with the Slosson accounting for a significantly higher proportion of the
variance with SAT scores. Regression equations are provided for determining expected SAT scores, based upon observed Otis-Lennon and Slosson IQs. Implications of the findings with regard to the screening of students for gifted programs are delineated.

Kaeser and Reynolds (1985) examined the II subtests of the Wechsler Preschool and Primary Scale of Intelligences (WPPSI) statistically for evidence of sex differences. The authors used the data gathered for the WPPSI in the early 1960s, which consisted of 600 males and 599 females selected on the basis of age, geographic region, urban-rural residence, race and father’s occupation. Results provided fundamental support for previous work with older children showing better female performance on memory and psychomotor tasks and better male performance on spatial tasks, although the distinctions in performance were not as clear as with older children.

Helmes (1987) conducted a study under the title “Concurrent Validation of AH2 as a Brief Measure of Intelligence in Canadian University Students”. The purpose of this study was to examine a recently developed British General Ability Test (AH2) and examine the performance of a group of Canadian University students on AH2, and correlating its scores with those from a multiple-choice vocabulary test and a widely used nonverbal test of abstract reasoning. This test was intended for ages 10 and above and covered a wide range of abilities through the use of two different time limits and a steep item difficulty progression. It was composed of three 40-item subsets: verbal (V), numerical (N), and pictorial (P), AH2 thus was designed
to provide information on several abilities in a relatively short period of time (approximately 30-45 minutes for the short time limits) over a wide range of ability. The sample consisted of 130 introductory psychology students (65 males, 65 females) at university of Western Ontario. The results indicated that there was a significant sex difference for the verbal subtest. Correlations with age were modestly significant for all scales except AH2 Numerical. Internal consistency reliabilities (coefficient alpha) were 0.79, 0.87 and 0.83 for AH2 verbal, numerical and perceptual scales respectively. There were no significant differences between the means for the Canadian students and the British students group, although the differences were significant for verbal subtest (p < .06).

Cooper and Fraboni, (1988) conducted a study entitled “Relationship between the Wechsler Adult intelligence Scale-Revised and the Wide Range Achievement test-Revised in a Sample of Normal Adults”. The Wechsler Adult Intelligence Scale-Revised (WAIS-R) and the wide Range Achievement Test-Revised (WRAT-R) were administered to 121 adults, 80 males and 41 females, and correlations were calculated between WRAT-R subtests and WAIS-R verbal, performance, and full scale IQs, and WAIS-R scaled scores. A Multiple Regression analysis was carried out using the WRAT-R subtests as independent variables and full scale IQ as the dependent variable. The results of the multiple regression analysis indicated that all three WRAT-R subtests were moderate predictors of WAIS-R full scale IQ than performance IQ. All correlations between IQ and WRAT-R subtests were significant (p = 0.1)
Hyde and Linn (1988), conducted a study entitled “Gender Differences in Verbal Ability: A Meta- Analysis”. They located 165 studies that reported data on gender differences in verbal ability. The weighted mean effect size (d) was + 0.11, indicating a slight female superiority in performance. The differences in verbal ability no longer existed. Analyses of effect sizes for different measures of verbal ability showed almost all to be small in magnitude: for vocabulary, d = 0.02; for analogies, d = - 0.16 (slight male superiority in performance; for reading comprehension, d = 0.03; for speech production, d =0.33 (the largest effect size); for essay writing, d = 0.09; for anagrams, d = 0.22; and for tests of general verbal ability, d =0.20. For the 1985 administration of the Scholastic Aptitude Test-Verbal, d = -0.11, indicating superior male performance. Analysis of tests requiring different cognitive processes involved in verbal ability yielded no evidence of substantial gender differences in any aspect of processing. Similarly, an analysis by age indicated no striking changes in the magnitude of gender differences at different ages, countering some conclusions made by some studies that gender differences in verbal ability emerge around age 11. For studies published in 1973 or earlier, d = 0.23 and for studies published after 1973, d = 0.10, indicating a slight decline in the magnitude of the gender difference in recent years. The implications of these findings are discussed, including their implications for theories of sex differences in brain lateralization and their relation to changing gender roles.

Lee and Lam, (1988) conducted a study, the purpose of which was to assess the developmental and the cultural invariance of the factor structure of the Wechsler
Intelligence Scale for Children Revised (WISC-R). Both the WISC-R and an adapted version of the WISC-R, adapted for use in Hong Kong, called the HK-WISC, were examined by using confirmatory factor analysis. Age groups of children 7, 10, and 13 years old were studied to assess the developmental invariance of the WISC-R and of the HK-WISC. The data was the matrices of correlations between subtests as reported in the WISC-R (Wechsler, 1974) and the HK-WISC (Yung, 1981) manuals. The correlation matrices, averaged over all age groups were examined in the study. The findings of this study suggest that the three-factor model is invariant across the two different cultural groups of American children and Hong Kong Chinese children as well as across three different age groups of 7, 10, and 13 for the HK-WISC. The cross-cultural and cross-age invariance of the factor structure provide strong support for the use of the Wechsler Scale for different age and cultural groups, provided that the test is properly normed and adapted for use with the particular groups of individuals.

Anderson et al (1989) investigated the temporal stability of WISC-R IQ scores for learning-disabled students (88 males and 25 females) from four school systems in Louisiana. Pearson product moment correlations yielded coefficients that were considerably lower than those previously reported: \( r = 0.55, p < .001 \) for the Verbal IQ; \( r = .63, p < .001 \) for the Performance IQs; and \( r = 0.58, p < .001 \) for the full scale IQs. Results of t-test analyses indicated that only the Verbal IQ scores were significantly different when the initial evaluation (\( M = 89.4 \)) was compared to the re-evaluation (\( M = 85.3 \)) \( p < .001 \). The results suggest that the WISC-R may be less
stable for the learning-disabled population than for other groups and that the average 3-year test-retest time lapse was an influential factor in the reduced reliability of this instrument.

Canivez (2000) investigated the predictive and construct validity of the Developing Cognitive Abilities Test (DCAT) in a heterogeneous sample of 863 sixth grade students. Level H of the DCAT was administered during the student’s sixth grade year and selected subtests of the Iowa Tests of Basic Skills (ITBS) were administered eight months later during their seventh grade year. Results showed that correlations between the DCAT and Iowa Tests of Basic Skills ranged from 0.50 to 0.74 with a median $r = 0.635$. Correlations also supported the construct (convergent) validity of the DCAT when compared to the ITBS with the DCAT verbal subtest correlating significantly higher with the ITBS vocabulary reading, and language usage than either the DCAT quantitative or spatial subtests which are not as verbally oriented. The DCAT quantitative subtest was associated with the ITBS mathematics problem solving to a greater extent than either the DCAT verbal or spatial subtests.

Fabregat et al (2000) conducted a study about sex differences in general intelligence defined as ‘g’ among young adolescents. They studied two independent samples. The samples were a total of 1565 young adolescents (797) girls and (768) boys. The congruence coefficients between the ‘g’ factors extracted for each sex suggested a near identity, and then the sex difference in ‘g’ was represented on each of the subtests in terms of a point-biserial correlation. These correlations were
included with the full matrix of subtest intercorrelations for factor analysis. The results reveal the factor loading of sex on g, which in this study suggest a null sex difference.

**Diseth** (2002) conducted a study, the purpose of which was to investigate the relationship between intelligence, approaches to learning and academic achievement. The sample consisted of 89 Norwegian undergraduate psychology students. Intelligence was measured by means of three different tests. The WAIS Vocabulary test was used as a measurement of crystallized intelligence, and Monneslands Verbal Analogies Test was a measure of fluid intelligence and the Sandford/Rybakoff Spatial Test was a measure of spatial intelligence. General intelligence did not correlate significantly with any of the approaches to learning variables or with examination grade. The WAIS vocabulary test appears to be the most relevant measurement of intelligence with respect to academic achievement. In general, the present findings do not give much support to the relationship between intelligence and achievement, as would be expected.

**Colom and Lopez** (2002) presented a study on sex differences in fluid intelligence (Gf) among high school graduates. The sample consisted of 4072 high school graduates (1772 females and 2300 males). Three measures of Gf were used in this study: The PMA Inductive Reasoning Test, the Advanced Progressive Matrices (APM), and the Culture Fair Intelligence Test (Scale 3). The mean scores obtained by females and males on the three tests, as well as the standardized sex difference (d), the difference in IQ points, and t values as tests of the statistical significance of the differences were obtained. The results reveal that females outperform males in the
PMA reasoning test, that males outperform females in the Raven, and that there is no sex difference in the Culture-Fair Test. There is no systematic difference favoring any sex in the measure of Gf, and that there is no sex difference in the best available measure of Gf (the culture-Fair Test), it is concluded that the sex difference in fluid intelligence is non-existent.

Codorniu-Raga and Vigil-Colet (2003) conducted a study on sex differences in psychometric and chronometric measures of intelligence among young adolescents. The objective of this study was to analysis the effects of sex on psychometric and chronometric measures of intelligence in a sample of 234 secondary school students (149 boys and 94 girls). The intelligence tests administered were the Spanish adaptation of Thurston’s Primary Mental Abilities (PMA), and the Spanish adaptation of Cattell’s ‘g’ Factor Test (TEA). They computed the descriptive statistics for all the measures in the study according to age and sex. They performed a multivariate analysis of variance and a univariate analysis of variance for each variable. They found that there was a significant increase in those abilities related to writing and reading speed for the age range from 11 to 14 years. The sex factor showed no significant effects either for the psychometric measures most related to ‘g’, such as Cattell’s factor test or the reasoning scale of the PMA test. Sex did show significant effects on many of the psychometric measures related to specific abilities such as the spatial, word fluency and numerical scales of the PMA. Girls outperformed boys in the word fluency and numerical scales, and boys outperformed girls in the spatial scale.
Lynn, Fergusson and Horwood (2005) conducted a study to find the sex differences on the WISC-R in New Zealand. Sex differences on the WISC-R were examined in a sample of 897 children from 8 and 9 years who were gathered during the course of the Christchurch Health and Development Study (CHDS). The CHDS is a longitudinal study of a birth cohort of 1265 children born in the Christchurch Urban region during mid 1977. The findings revealed that boys scored significantly higher than girls on the subtests of information, vocabulary, block design and object assembly, while girls scored significantly higher on coding. Boys obtained slightly but not significantly higher scores on the verbal, performance and full scale IQs. The results were in general similar to the sex differences in the standardization samples of the WISC-R in Scotland, the Netherlands and the United states.

On the basis of the above review, the following conclusions can be made:

The above studies can be divided into three parts, the first deals with the development of intelligence test through studying the reliability and validity of some test as those of Proger et al (1971), Estabrook (1984), Antonak et al (1982), Helmes (1989), Lee and Lam (1988), Anderson et al (1989), and Canivez (2000).


The third part studies the relationship between intelligence and academic achievement as of those Antonak et al (1982), and Cooper and Fraboni (1988).
The above findings reveal that:

1. There are significant differences in intelligence scores between age levels as studies conducted by Lee and Lam (1988), Codorniu-Raga and Vigil_Colet (2003).

2. There are no significant differences between boys and girls according to the studies conducted by Fabregat et al (2000), and Colom and Lopez (2002).

3. There are differences in intelligence scores between boys and girls as those of Kasser and Reynolds (1985), Helmes (1987), Codorniu-Raga and Vigil-Colet (2003), and Lynn, Feryusson and Horwood (2005).

4. Grossman et al (1983) and Cooper and Fraboni (1988) found significant relationship between students intelligence Scores and academic achievement, Antonak et al (1982) found that the best predictor of achievement was I.Q variable alone. The findings of Diseth (2002) stress the idea that there is no significant relationship between intelligence and achievement.