

*Full Length Research Paper*

# Evaluation of English achievement test: A comparison between high and low achievers amongst selected elementary school students of Pakistan

Zubair Haider\*, Farah Latif, Samina Akhtar and Maria Mushtaq

Department of Educational Training, The Islamic University of Bahawalpur, RYK Campus, Pakistan.

Accepted 5 September, 2012

Validity, reliability and item analysis are critical to the process of evaluating the quality of an educational measurement. The present study evaluates the quality of an assessment constructed to measure elementary school student's achievement in English. In this study, the survey model of descriptive research was used as a research method. Students of both genders (male/female) constitute the population of the study. To measure the achievement level of students, a test of English language as a subject was administered to the sample of 150 students (75 males and 75 females) of 8th grade. They were selected by convenient sampling method from 10 (5 private sectors and 5 public sectors) schools in Pakistan. English Achievement Test for elementary students comprises of 50 multiple-choice items (developed by the author), was used. The collected data were analyzed through descriptive statistics, Z-test and item analysis. The results of the study indicate that male student's achievement level in English is better than the female students at elementary level in selected schools.

**Key words:** English Language, analysis, achievement test, standardization, non-standardization, elementary level, secondary school.

## INTRODUCTION

"Education is the most important asset we have, because our knowledge is the type of wealth that we will never lose no matter what, and the more we share it the more it increases" (Mangal, 2002). The process of acquiring basic knowledge starts with elementary education. Without elementary education, none of one's dreams will be possible. Aggarwal (1991) truly described "Elementary education is like the first step we take in life; we will never be able to run if we do not learn how to walk". According to Suter (2006) "Elementary school level is an important level of formal education in which common basic knowledge and skills are taught, which are required for all citizens in the society".

The elementary education process brings individuals to a level of fundamental competence for solving problems, adapting to social values and applying established social rules. Salma (2000) stated "Children learn much spontaneously, and it is only by careful observation to this spontaneous process that we can build up a sound theory of education and of the school curriculum". Bandura (1986) stated "Knowledge and skills gained in the elementary education, is the basic knowledge and skill to be gained at other educational levels, for this reason other educational levels is also based on the elementary education". Therefore, this fundamental education level affects negatively or positively not only the educational system of the society, but also the other systems of the society. Shahid (2000) observed "Elementary level examinations in Pakistan are the most comprehensive form of testing, typically given at the end

\*Corresponding author. E-mail: [zubairiub@hotmail.com](mailto:zubairiub@hotmail.com).

of the term and one or two times during the semester, a test is more limited in scope, focusing on more particular aspects of the course material”.

“Measurement is central to the construction of a quality student assessment, even in the case of classroom-designed or non-standardized assessments. Measuring variables is one of the necessary steps in the research process” (Morales, 2009). Ebel and Frisbi (1991) define the test in very comprehensive manner “test is a means of measuring the knowledge, skill, feeling, intelligence and aptitude of an individual or group”. Airasian (2005) mentioned “Standardized test is a test, administered and scored in a consistent manner. These tests are designed in such a way that the questions, conditions for administering, scoring procedures, and interpretations are consistent”. The considerations of validity and reliability typically are viewed as essential elements for determining the quality of any standardized test. Johnson and Larry (2006) found “professional and practitioner associations frequently have placed these concerns within broader contexts when developing standards and making overall judgments about the quality of any standardized test as a whole within a given context”.

Wright (2008) pointed “The achievement test whether standardized or non standardized, measures that how much of the material has been mastered and assess the student current status”. These tests are used to determine what a student has learned such as vocabulary, reading, math skill, etc. “Achievement tests are used to evaluate a students or worker's understanding, comprehension, knowledge and capability in a particular area. They are used in academics, professions and many other areas” (Carey, 1994). Wiersma and Jurs (1990) identified “An achievement test is intended to measure what the student has learned or what skills the student has mastered”. “They are typically norm referenced tests that measure the pupil's level of achievement in various content and skill areas” (Gronlund and Linn, 1990). “The achievement test focuses upon examinees attainments at a given point in time” (Jesa, 2005). Anastasia (1982) mentioned “Achievement tests are examinations that are designed to determine the degree of knowledge and proficiency exhibited by an individual in a special area or set of areas”. Moreover, they could be extremely crucial for the students, for they are intended either to make the students pass or fail the test.

Leung (1998) observed “The achievement tests are served as a tool to measure current knowledge levels for the purpose of placing students in an individual environment where they have the chance to advance at a pace that is suitable for their abilities”. These type of tests willfully involve teachers for they will be responsible for the preparation of such tests and giving them to the learners. Egen and Kauchak (1992) defined “Achievement tests measure a person's accomplishment in a subject or task. One instrument may serve both

purposes, acting as an aptitude test to forecast future performances and an achievement test to monitor past and present learning”. It measures how successful students are in achieving objectives of a lesson/course/curriculum.

Hills (1981) found “Achievement test are well suited and it provide educators with objective feedback as to, how much students are learning and understanding”. Thorndike and Hagen (1997) defined “Achievement tests as the type of ability test that describes what a person has learned, thus is called an achievement test”. “Achievement test scores are often used in an educational system to determine what level of instruction for which a student is prepared” (Hudson, 1987). High achievement scores usually indicate a mastery of grade-level material, and the readiness for advanced instruction. Low achievement scores can indicate the need for remediation or repeating a course grade. Results of achievement test also provide a track to the counselor to guide and provide remedy to the students in best possible way.

In all kind of achievement tests, item difficulty always secures a very important position. Morales quoted by Linacre (2002), “item difficulty is the characteristics influencing person responses and person ability is the characteristics influencing item difficulty estimates”. A careful consideration is required to be given to the construction of assessments and all items should be written clearly and concisely so that they are not vulnerable to the guess work by the students. For the quality of an evaluation tool, a discussion of validity and reliability is very necessary.

Students' achievement in English language has always been a very concerning matter in Pakistan, because English has become a part and parcel of Pakistan education system. Today, English is the second major language of instruction in Pakistan after Urdu. The main objective of this study was to analyze the psychometric properties of the instrument administered to evaluate the achievement level of male and female students in subject of English at elementary level and also identify low and high achiever students. The test that measured achievement in school English is criterion-referenced, so that test scores directly convey level of competency in defined English domain.

## METHODOLOGY

### Participants

A total of 150 students (75 males and 75 females) from the following schools completed the English achievement test during the ending period of the school year 2008 and 2009. Public schools (Government Girls Model School, Government Canal Colony Girl's School, Government Senior Muslim Girl's School, Government English Public School and Government Girls Tameer-e-Millat School) and private schools (Beacon House School, Educators School, City, School, National Garrison School and Oxbridge

School). All were urban schools and taken from Rahim Yar Khan district situated in the province of Punjab, Pakistan.

### Measure

The English achievement test, a multiple choice assessment designed to measure 8th class students' English ability was administered. The material from the 17 chapters of the English Textbook, issued by Punjab Textbook Board, Lahore for 8th Class was included in the test. The author constructed an English achievement test that comprised of 50 multiple choice items each with four answer choices. The achievement test was piloted with two groups of male and female students. The researcher tried to cover the maximum subject matter in the test. In order to correlate test items with the subject, a table of specification was prepared. The test was content-validated by a professor in the Department of English, the Islamia University of Bahawalpur, suggestions were accepted and the test was revised accordingly.

### Administration and scoring of test

After the preparation of test, co-operating teachers administered the test for female students, while for male students it was administered personally by the researcher. The researcher sought permission from the administrators/principals of the relevant schools well before time. Before administering the tests, specific instructions were also given to the students. The test was administered simultaneously for male and female groups of the students. The students completed the test for two hours under the lookout of their teachers and researcher. The purpose of the teacher-proctor monitoring of the test was to minimize measurement errors that could arise during the actual test. As the answer sheets were also provided with the original test format, the scoring key was used to assess the performance of students on the test. This scientific approach proved to be useful. In this way, the research could score students tests in time.

### Data collection

The sample selection for the method used in the research had been based on convenient calculation. It is assumed that the selected sample represents the whole amount in a high degree. Total of 150 students comprising of 75 males and 75 females were selected as sample from the population. This sample was further divided into 8 groups (4 male groups and 4 female groups). The number of female sample was 40, 30, 35 and 45, respectively, while the number of four male samples was 36, 44, 40 and 30, respectively. Microsoft Office Excel 2007, was used for the analysis and computations involved in analysis. Statistical Package for Social Sciences (SPSS) software version 16 was also used to determine reliability analysis of the test.

## RESULTS

After the administration of the test from students personally by the researcher, the collected data were analyzed and interpreted in the form of percentage, mean performance, standard deviation (SD), Z-test and item analysis. In the research, the methods with co-forms, Spearman Brown correlation coefficient, Kuder Richardson formula and Cronbach's alpha coefficient have been used. The reliability of the survey has been

tested by Cronbach's alpha coefficient. "The researchers agreed on this statement that it is sufficient as a measurement tool to be stated as reliable, if the alpha value is between 0.50 and 0.70" (Johnson and Larry, 2006). The internal consistency of test was also found to be high with a Cronbach's alpha value of 0.66. This value indicates a good reliability for the achievement test. Z-test was applied and the value of critical ratio (CR) was calculated.

### Validity and reliability

The accuracy, or usefulness, of a test is known as its validity. Carey (1994) stated "In additional measurement, validity is often defined as the degree to which a measuring instrument actually serves the purpose for which it intends". In order to assess the reliability of the test, split-half method was also performed computing reliability coefficient of 0.59 for half test and 0.74 for whole test, a value that indicates internal consistencies of the responses in the test. Finally, Kuder Richardson (KR)

formula,  $KR20 = \frac{n}{n-1} \left[ \frac{SD^2 - \sum Pq}{SD^2} \right]$  was used to

determine internal consistency with a value of 0.78 and less accurate, but simple Kuder Richardson formula,

$KR21 = \frac{n}{n-1} \left[ \frac{\sum x(x-1)}{nSD^2} \right]$  was used with a value of

0.70. It was developed for removing the hurdles of KR20, because procedure of KR20 is very difficult for large data. This formula will yield approximately the same results as KR20, but in most cases, the reliability estimate will be smaller.

### Item analysis and discrimination

The test consisted of 50 items. In order to compute the item difficulty and discrimination index, three procedures were applied. In order to identify between high achievers and low achievers, the discriminatory power (D) was calculated, while in order to assess the level of difficulty of each item, the Id was calculated and item discrimination was calculated with the help of the phi-coefficient ( $\Phi$ ). The 150 test papers were in order, highest to lowest scores and 50 papers with the highest total scores and 50 papers with the lowest total scores were selected. So, the three groups formed high achievers (HA), low achievers (LA) and middle achievers (MA). For each question Id, D, and  $\Phi$  were computed separately.

In Table 1, the value of Id was computed for each item, where the value of Id was greater than 80%, those items were rejected, because they were very easy item. Similarly, the item having the Id value less than 20% were also rejected on account of being difficult items. Therefore, the items number 1, 2, 5, 6, 7, 9, 11, 12, 13,

Table 1. Item wise analysis.

Item no.	Group	Correct responses	Id	D	$\phi$	Item no.	Group	Correct responses	Id	D	$\phi$																																																																																																																																																																																																																																																								
1	HA	49	0.97	0.02	0.01	26	HA	44	0.75	0.26	0.25																																																																																																																																																																																																																																																								
	LA	48					LA	31				2	HA	47	0.92	0.04	0.03	27	HA	50	0.99	0.02	0.01	LA	45	LA	49	3	HA	41	0.77	0.1	0.09	28	HA	46	0.7	0.44	0.54	LA	36	LA	24	4	HA	50	0.9	0.2	0.17	29	HA	43	0.62	0.48	0.6	LA	40	LA	19	5	HA	50	0.87	0.26	0.23	30	HA	46	0.79	0.26	0.24	LA	37	LA	33	6	HA	47	0.85	0.18	0.15	31	HA	44	0.69	0.38	0.41	LA	38	LA	25	7	HA	48	0.85	0.22	0.19	32	HA	27	0.49	0.1	0.25	LA	37	LA	22	8	HA	48	0.8	0.32	0.31	33	HA	34	0.64	0.08	0.07	LA	32	LA	30	9	HA	49	0.89	0.18	0.15	34	HA	32	0.55	0.18	0.19	LA	40	LA	23	10	HA	50	0.82	0.36	0.35	35	HA	43	0.74	0.24	0.23	LA	32	LA	31	11	HA	48	0.84	0.24	0.21	36	HA	40	0.61	0.38	0.44	LA	36	LA	21	12	HA	50	0.94	0.12	0.08	37	HA	36	0.63	0.18	0.18	LA	44	LA	27	13	HA	48	0.87	0.18	0.15	38	HA	41	0.68	0.28	0.28	LA	39	LA	27	14	HA	42	0.77	0.14	0.12	39	HA	35	0.51	0.38	0.48	LA	35	LA	16	15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0	LA	36	LA	8	16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA
2	HA	47	0.92	0.04	0.03	27	HA	50	0.99	0.02	0.01																																																																																																																																																																																																																																																								
	LA	45					LA	49				3	HA	41	0.77	0.1	0.09	28	HA	46	0.7	0.44	0.54	LA	36	LA	24	4	HA	50	0.9	0.2	0.17	29	HA	43	0.62	0.48	0.6	LA	40	LA	19	5	HA	50	0.87	0.26	0.23	30	HA	46	0.79	0.26	0.24	LA	37	LA	33	6	HA	47	0.85	0.18	0.15	31	HA	44	0.69	0.38	0.41	LA	38	LA	25	7	HA	48	0.85	0.22	0.19	32	HA	27	0.49	0.1	0.25	LA	37	LA	22	8	HA	48	0.8	0.32	0.31	33	HA	34	0.64	0.08	0.07	LA	32	LA	30	9	HA	49	0.89	0.18	0.15	34	HA	32	0.55	0.18	0.19	LA	40	LA	23	10	HA	50	0.82	0.36	0.35	35	HA	43	0.74	0.24	0.23	LA	32	LA	31	11	HA	48	0.84	0.24	0.21	36	HA	40	0.61	0.38	0.44	LA	36	LA	21	12	HA	50	0.94	0.12	0.08	37	HA	36	0.63	0.18	0.18	LA	44	LA	27	13	HA	48	0.87	0.18	0.15	38	HA	41	0.68	0.28	0.28	LA	39	LA	27	14	HA	42	0.77	0.14	0.12	39	HA	35	0.51	0.38	0.48	LA	35	LA	16	15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0	LA	36	LA	8	16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20								
3	HA	41	0.77	0.1	0.09	28	HA	46	0.7	0.44	0.54																																																																																																																																																																																																																																																								
	LA	36					LA	24				4	HA	50	0.9	0.2	0.17	29	HA	43	0.62	0.48	0.6	LA	40	LA	19	5	HA	50	0.87	0.26	0.23	30	HA	46	0.79	0.26	0.24	LA	37	LA	33	6	HA	47	0.85	0.18	0.15	31	HA	44	0.69	0.38	0.41	LA	38	LA	25	7	HA	48	0.85	0.22	0.19	32	HA	27	0.49	0.1	0.25	LA	37	LA	22	8	HA	48	0.8	0.32	0.31	33	HA	34	0.64	0.08	0.07	LA	32	LA	30	9	HA	49	0.89	0.18	0.15	34	HA	32	0.55	0.18	0.19	LA	40	LA	23	10	HA	50	0.82	0.36	0.35	35	HA	43	0.74	0.24	0.23	LA	32	LA	31	11	HA	48	0.84	0.24	0.21	36	HA	40	0.61	0.38	0.44	LA	36	LA	21	12	HA	50	0.94	0.12	0.08	37	HA	36	0.63	0.18	0.18	LA	44	LA	27	13	HA	48	0.87	0.18	0.15	38	HA	41	0.68	0.28	0.28	LA	39	LA	27	14	HA	42	0.77	0.14	0.12	39	HA	35	0.51	0.38	0.48	LA	35	LA	16	15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0	LA	36	LA	8	16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20																								
4	HA	50	0.9	0.2	0.17	29	HA	43	0.62	0.48	0.6																																																																																																																																																																																																																																																								
	LA	40					LA	19				5	HA	50	0.87	0.26	0.23	30	HA	46	0.79	0.26	0.24	LA	37	LA	33	6	HA	47	0.85	0.18	0.15	31	HA	44	0.69	0.38	0.41	LA	38	LA	25	7	HA	48	0.85	0.22	0.19	32	HA	27	0.49	0.1	0.25	LA	37	LA	22	8	HA	48	0.8	0.32	0.31	33	HA	34	0.64	0.08	0.07	LA	32	LA	30	9	HA	49	0.89	0.18	0.15	34	HA	32	0.55	0.18	0.19	LA	40	LA	23	10	HA	50	0.82	0.36	0.35	35	HA	43	0.74	0.24	0.23	LA	32	LA	31	11	HA	48	0.84	0.24	0.21	36	HA	40	0.61	0.38	0.44	LA	36	LA	21	12	HA	50	0.94	0.12	0.08	37	HA	36	0.63	0.18	0.18	LA	44	LA	27	13	HA	48	0.87	0.18	0.15	38	HA	41	0.68	0.28	0.28	LA	39	LA	27	14	HA	42	0.77	0.14	0.12	39	HA	35	0.51	0.38	0.48	LA	35	LA	16	15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0	LA	36	LA	8	16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20																																								
5	HA	50	0.87	0.26	0.23	30	HA	46	0.79	0.26	0.24																																																																																																																																																																																																																																																								
	LA	37					LA	33				6	HA	47	0.85	0.18	0.15	31	HA	44	0.69	0.38	0.41	LA	38	LA	25	7	HA	48	0.85	0.22	0.19	32	HA	27	0.49	0.1	0.25	LA	37	LA	22	8	HA	48	0.8	0.32	0.31	33	HA	34	0.64	0.08	0.07	LA	32	LA	30	9	HA	49	0.89	0.18	0.15	34	HA	32	0.55	0.18	0.19	LA	40	LA	23	10	HA	50	0.82	0.36	0.35	35	HA	43	0.74	0.24	0.23	LA	32	LA	31	11	HA	48	0.84	0.24	0.21	36	HA	40	0.61	0.38	0.44	LA	36	LA	21	12	HA	50	0.94	0.12	0.08	37	HA	36	0.63	0.18	0.18	LA	44	LA	27	13	HA	48	0.87	0.18	0.15	38	HA	41	0.68	0.28	0.28	LA	39	LA	27	14	HA	42	0.77	0.14	0.12	39	HA	35	0.51	0.38	0.48	LA	35	LA	16	15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0	LA	36	LA	8	16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20																																																								
6	HA	47	0.85	0.18	0.15	31	HA	44	0.69	0.38	0.41																																																																																																																																																																																																																																																								
	LA	38					LA	25				7	HA	48	0.85	0.22	0.19	32	HA	27	0.49	0.1	0.25	LA	37	LA	22	8	HA	48	0.8	0.32	0.31	33	HA	34	0.64	0.08	0.07	LA	32	LA	30	9	HA	49	0.89	0.18	0.15	34	HA	32	0.55	0.18	0.19	LA	40	LA	23	10	HA	50	0.82	0.36	0.35	35	HA	43	0.74	0.24	0.23	LA	32	LA	31	11	HA	48	0.84	0.24	0.21	36	HA	40	0.61	0.38	0.44	LA	36	LA	21	12	HA	50	0.94	0.12	0.08	37	HA	36	0.63	0.18	0.18	LA	44	LA	27	13	HA	48	0.87	0.18	0.15	38	HA	41	0.68	0.28	0.28	LA	39	LA	27	14	HA	42	0.77	0.14	0.12	39	HA	35	0.51	0.38	0.48	LA	35	LA	16	15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0	LA	36	LA	8	16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20																																																																								
7	HA	48	0.85	0.22	0.19	32	HA	27	0.49	0.1	0.25																																																																																																																																																																																																																																																								
	LA	37					LA	22				8	HA	48	0.8	0.32	0.31	33	HA	34	0.64	0.08	0.07	LA	32	LA	30	9	HA	49	0.89	0.18	0.15	34	HA	32	0.55	0.18	0.19	LA	40	LA	23	10	HA	50	0.82	0.36	0.35	35	HA	43	0.74	0.24	0.23	LA	32	LA	31	11	HA	48	0.84	0.24	0.21	36	HA	40	0.61	0.38	0.44	LA	36	LA	21	12	HA	50	0.94	0.12	0.08	37	HA	36	0.63	0.18	0.18	LA	44	LA	27	13	HA	48	0.87	0.18	0.15	38	HA	41	0.68	0.28	0.28	LA	39	LA	27	14	HA	42	0.77	0.14	0.12	39	HA	35	0.51	0.38	0.48	LA	35	LA	16	15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0	LA	36	LA	8	16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20																																																																																								
8	HA	48	0.8	0.32	0.31	33	HA	34	0.64	0.08	0.07																																																																																																																																																																																																																																																								
	LA	32					LA	30				9	HA	49	0.89	0.18	0.15	34	HA	32	0.55	0.18	0.19	LA	40	LA	23	10	HA	50	0.82	0.36	0.35	35	HA	43	0.74	0.24	0.23	LA	32	LA	31	11	HA	48	0.84	0.24	0.21	36	HA	40	0.61	0.38	0.44	LA	36	LA	21	12	HA	50	0.94	0.12	0.08	37	HA	36	0.63	0.18	0.18	LA	44	LA	27	13	HA	48	0.87	0.18	0.15	38	HA	41	0.68	0.28	0.28	LA	39	LA	27	14	HA	42	0.77	0.14	0.12	39	HA	35	0.51	0.38	0.48	LA	35	LA	16	15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0	LA	36	LA	8	16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20																																																																																																								
9	HA	49	0.89	0.18	0.15	34	HA	32	0.55	0.18	0.19																																																																																																																																																																																																																																																								
	LA	40					LA	23				10	HA	50	0.82	0.36	0.35	35	HA	43	0.74	0.24	0.23	LA	32	LA	31	11	HA	48	0.84	0.24	0.21	36	HA	40	0.61	0.38	0.44	LA	36	LA	21	12	HA	50	0.94	0.12	0.08	37	HA	36	0.63	0.18	0.18	LA	44	LA	27	13	HA	48	0.87	0.18	0.15	38	HA	41	0.68	0.28	0.28	LA	39	LA	27	14	HA	42	0.77	0.14	0.12	39	HA	35	0.51	0.38	0.48	LA	35	LA	16	15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0	LA	36	LA	8	16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20																																																																																																																								
10	HA	50	0.82	0.36	0.35	35	HA	43	0.74	0.24	0.23																																																																																																																																																																																																																																																								
	LA	32					LA	31				11	HA	48	0.84	0.24	0.21	36	HA	40	0.61	0.38	0.44	LA	36	LA	21	12	HA	50	0.94	0.12	0.08	37	HA	36	0.63	0.18	0.18	LA	44	LA	27	13	HA	48	0.87	0.18	0.15	38	HA	41	0.68	0.28	0.28	LA	39	LA	27	14	HA	42	0.77	0.14	0.12	39	HA	35	0.51	0.38	0.48	LA	35	LA	16	15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0	LA	36	LA	8	16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20																																																																																																																																								
11	HA	48	0.84	0.24	0.21	36	HA	40	0.61	0.38	0.44																																																																																																																																																																																																																																																								
	LA	36					LA	21				12	HA	50	0.94	0.12	0.08	37	HA	36	0.63	0.18	0.18	LA	44	LA	27	13	HA	48	0.87	0.18	0.15	38	HA	41	0.68	0.28	0.28	LA	39	LA	27	14	HA	42	0.77	0.14	0.12	39	HA	35	0.51	0.38	0.48	LA	35	LA	16	15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0	LA	36	LA	8	16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20																																																																																																																																																								
12	HA	50	0.94	0.12	0.08	37	HA	36	0.63	0.18	0.18																																																																																																																																																																																																																																																								
	LA	44					LA	27				13	HA	48	0.87	0.18	0.15	38	HA	41	0.68	0.28	0.28	LA	39	LA	27	14	HA	42	0.77	0.14	0.12	39	HA	35	0.51	0.38	0.48	LA	35	LA	16	15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0	LA	36	LA	8	16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20																																																																																																																																																																								
13	HA	48	0.87	0.18	0.15	38	HA	41	0.68	0.28	0.28																																																																																																																																																																																																																																																								
	LA	39					LA	27				14	HA	42	0.77	0.14	0.12	39	HA	35	0.51	0.38	0.48	LA	35	LA	16	15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0	LA	36	LA	8	16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20																																																																																																																																																																																								
14	HA	42	0.77	0.14	0.12	39	HA	35	0.51	0.38	0.48																																																																																																																																																																																																																																																								
	LA	35					LA	16				15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0	LA	36	LA	8	16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20																																																																																																																																																																																																								
15	HA	47	0.83	0.22	0.19	40	HA	8	0.16	0	0																																																																																																																																																																																																																																																								
	LA	36					LA	8				16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33	LA	15	LA	15	17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20																																																																																																																																																																																																																								
16	HA	44	0.59	0.58	0.82	41	HA	28	0.43	0.26	0.33																																																																																																																																																																																																																																																								
	LA	15					LA	15				17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04	LA	33	LA	20																																																																																																																																																																																																																																								
17	HA	49	0.82	0.32	0.3	42	HA	22	0.42	0.04	0.04																																																																																																																																																																																																																																																								
	LA	33					LA	20																																																																																																																																																																																																																																																											

Table 1. Contd.

18	H A L A	49 44	0.93	0.1	0.08	43	H A L A	29 18	0.47	0.22	0.26
19	H A L A	47 35	0.82	0.24	0.21	44	H A L A	36 22	0.58	0.28	0.31
20	H A L A	48 42	0.9	0.12	0.1	45	H A L A	37 12	0.49	0.5	0.71
21	H A L A	48 23	0.71	0.5	0.59	46	H A L A	43 40	0.83	0.06	0.05
22	H A L A	45 36	0.81	0.18	0.16	47	H A L A	46 26	0.72	0.4	0.23
23	H A L A	47 40	0.87	0.14	0.11	48	H A L A	48 18	0.66	0.6	0.83
24	H A L A	48 34	0.82	0.28	0.26	49	H A L A	48 38	0.86	0.2	0.17
25	H A L A	47 38	0.85	0.18	0.15	50	H A L A	48 35	0.83	0.26	0.23

18, 27, 40 and 49, were rejected and item numbers 10, 15, 17, 19, 22, 24, 25, 46 and 50 modified. The value of D was computed on each item. All those items having  $D = 0.20$  were rejected, because those items were unable to discriminate between the HA and LA on the basis of this criterion. Items 1, 2, 3, 18, 20, 23, 27, 32, 33, 40, 42 and 46 were rejected and items 5, 9, 22, 25, 34 and 37 were suggested to be modified. For further discrimination, the value of  $\Phi$  was also computed. All these items having value of less than 0.20 were also rejected. This criterion held to drop the following items 1, 2, 3, 4, 6, 9, 12, 18, 20, 23, 27, 33, 40 and 46 and item numbers 22 and 49 were suggested to be modified. On the basis of all these values of  $I_d$ ,  $D$  and  $\Phi$ , several items were recommended to be dropped: items numbered 1, 2, 9, 18, 22, 23, 27, 40 and 46.

Table 1 shows that 36 (72%) of the items are average items and the remaining 14 (28%) belong to difficult and easy items. It could be implied with the results that achievement test was fairly difficult, because more than half of the students got most of the items correctly. But, considering that the examinees were English major, the result could also mean that they really have the ability to answer even difficult items, because English language has rigid qualifying test to proceed with their field of specialization. Thus, to be able to major in English, the students must have attained an above average score in the high and higher secondary school entrance examination.

Of the 41 items considered in the test, only 9 or 18% came up to be poor items. These items were dropped or rejected. Only four marginal items need to be improved. Thirty seven or 74% of the items were either good or very good items. This means that generally, the achievement test items truly represented the learning ability of test takers in English language. Most of the items in test can discriminate well the high and low achieving groups.

After the application period of data collection tools has been completed, the answering papers were controlled. In the first step of evaluation, the scores of 150 students were presented in the form of frequency distribution (Figure 1). The data of frequency distribution was converted and presented in form of polygon. The graph shows that the distribution was negatively skewed. A large majority of the students are seen on the right end. On the left hand, the population is very thin. The polygon indicates that the test was somewhat easy for the sample population included in the research.

As indicated in Table 3 and Figure 2, the mean performance of group 5 is better and the mean performance of group 3 which is low. Comparison shows that the male mean performance was better than the female mean performance. Standard deviation is the most reliable and authentic measure of dispersion. It shows the dispersion of scores on the scale. When the value of standard deviation is higher, it means that the spread of score on two sides of arithmetic mean is greater, while the low magnitude of standard deviation

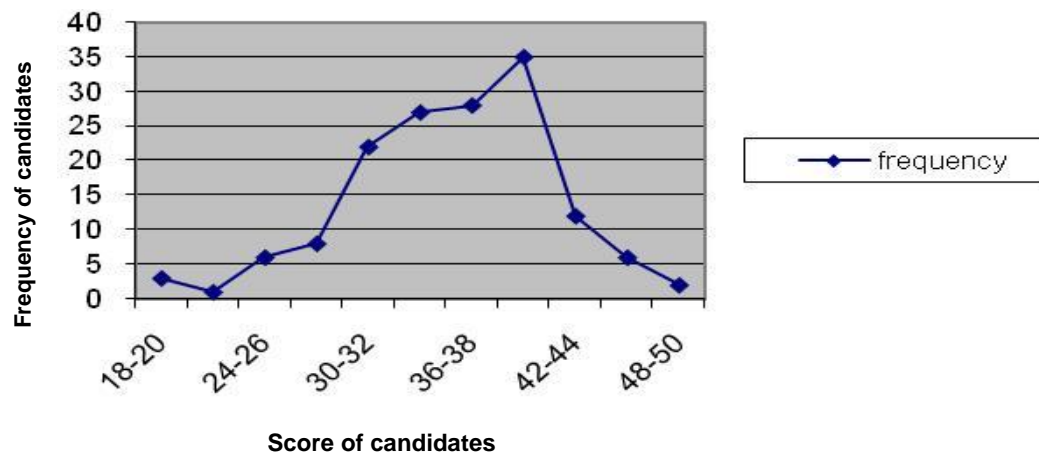


Figure 1. Frequency distribution.

Table 2. Frequency distribution.

C-I	Frequency
18 - 20	03
21 - 23	01
24 - 26	06
27 - 29	08
30 - 32	22
33 - 35	27
36 - 38	28
39 - 41	35
42 - 44	12
45 - 47	06
48 - 50	02
Total	150

indicates the lesser spread of score on the scale. Group 4 has the highest value of standard deviation (6.4) which shows that the spread of scores in this very group was greater. In the same way, in group 1, standard deviation (3.9) was the least, which indicates that the spread of score in this group was very low.

Table 4 shows that the combined mean performance of male students is somehow greater than the combined mean performance of female students in the test. The table further indicates that the scores of combined standard deviation of male group were greater than the female group. Then to check the difference between the mean performance of the male and female samples, Z-test was applied and the value of critical ratio (CR) was calculated and presented in graphical form as shown in Figure 3. Critical value < test value, the calculated value of Z-test is 1.80 and it is smaller than the table value of 1.96; the difference is statistically significant which means that the male students were better performance in the test as compared to the female students.

## DISCUSSION

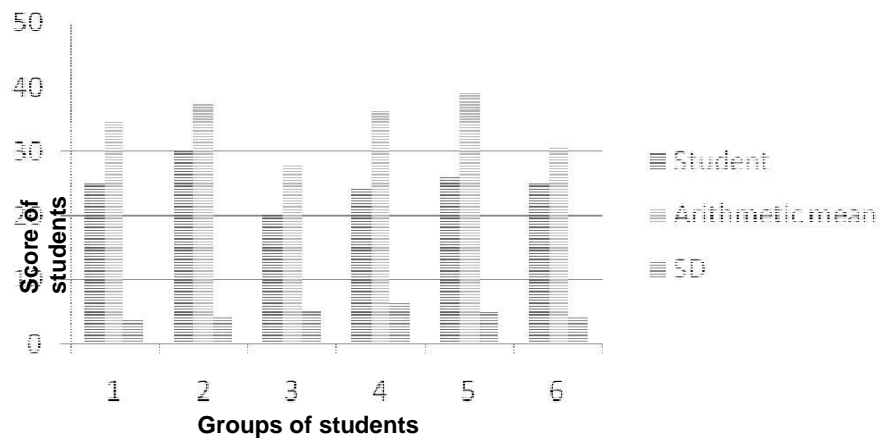
On the basis of test results, the researcher re-examined all items tagged for review. The items rejected were revised thoroughly and were made the first items in an effort to place easier items first on the student assessment. Some items were rewarded after it was felt that students were over analyzing the questions. The item with the negative item correlation (item 18) was considered to be deleted, because the item in general was confusing. These improvements will enhance the reliability and the quality of the test; subsequently, improve students' academic achievement and performance.

Overall, result of the analysis could figure out that the achievement test in general was a good test, although, some items were removed, revised and rephrased, most of the items were found to be good items.

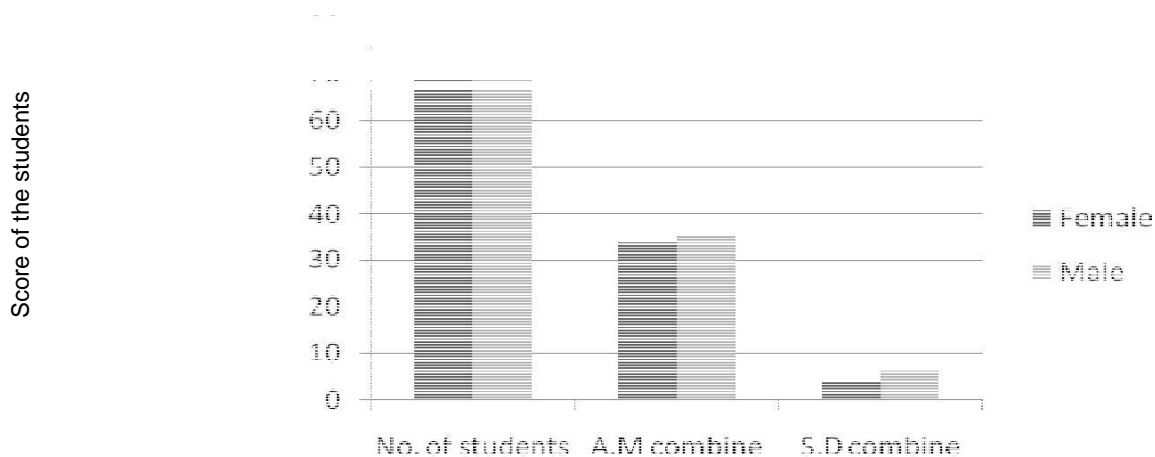
The results obtained from Table 2 indicate frequency distribution and most of the students lay between the score of 39 and 41. The polygon indicates that the test was balance for the sample included in the research. On the basis of percentage performance of the students, the highest value was 90% and lowest value was 40% in female groups, while the highest value was 98% and lowest value 42%, which shows the better performance of male students than the female. The mean performance in female groups was as 34.5, 37.4 and 27.9, the combined mean of female group was 33.9. The mean performance in male groups was as 36.3, 39.1 and 30.6, the combined mean of male group was 35.4. The spread of score in female groups was 3.9, 4.5 and 5.2, the combined SD of female group was 4.36. The spread of score in male groups was as 6.4, 4.8, and 4.3, the combined SD of male group was 6.32. The computed value of Z-test was 1.80 which is less than the table value (1.96), which means the difference is significant showing that the male students were statistically better performer than the female students. Three methods were used to compute

**Table 3.** Mean and standard deviation of students' test score.

Group	Group no.	No. of students	Mean	SD
Female	01	25	34.5	3.9
	02	30	37.4	4.5
	03	20	27.9	5.2
Male	04	24	36.3	6.4
	05	26	39.1	4.8
	06	25	30.6	4.3

**Figure 2.** Mean and standard deviation of students' test score.**Table 4.** Difference between mean performance and standard deviation.

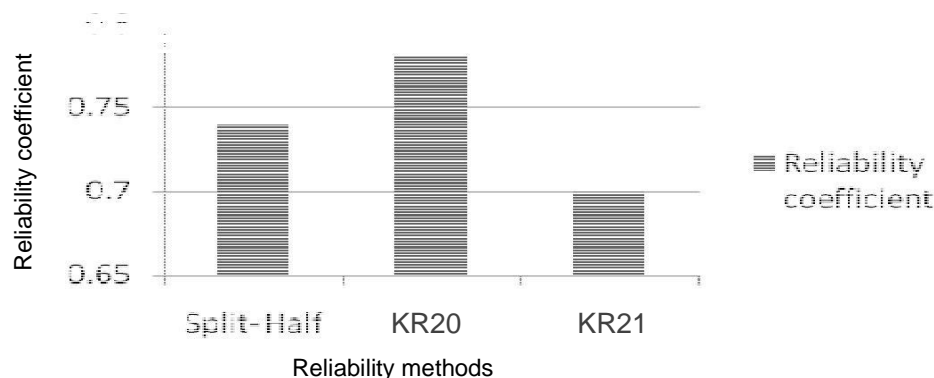
Group	No. of students	Combined mean	Combined SD
Female	75	33.9	4.36
Male	75	35.4	6.32

**Figure 3.** Graphical presentation of the difference between mean performance and standard deviation.

**Table 5.** Reliability coefficient by different methods.

Method	Reliability coefficient
Split- Half	0.74
KR20	0.78
KR21	0.70

\*KR = Kuder Richardson formula.

**Figure 4.** Graphical presentation of the reliability coefficient of the different methods.

the reliability coefficient; the calculated values of these methods were as in split-half (0.74), KR20, (0.78), and KR21 (0.70) (Table 5 and Figure 4). As all the calculated values of reliability coefficient are more than 50, which means the test was concerned to be reliable to a great extent. Items were analyzed through  $I_d$ ,  $D$  and  $\Phi$ , where the calculated values of  $D$  fluctuated from 0 to 0.60,  $I_d$  from 0.16 to 0.99 and  $\Phi$  from 0.00 to 0.83. Although, it could be said that English major have the advantage in taking the test, it should not stop there. The test was made to measure the knowledge and achievement that was supposedly acquired by a student regardless of his/her area of specialization.

## Conclusion

This study stands as an improvement to existing studies. Basically, it was to find out the quality of the items in a test, try them out, analyze their behavior practically and discriminate high achievers and low achievers (male and female) students. The data used in this study help us to draw a number of sharp conclusions about evaluation of English achievement and elementary education. (1) For the English achievement test, item difficulty was computed for each item, where the value of item difficulty was greater than 80% and less than 20%, those items were rejected because they are very easy and very difficulty item. (2) Items having discrimination index = 0.20 or less were rejected, because they were unable to discriminate between HA and LA on the basis of the set

criterion. (3) On the basis of mean performance in test score, it is observed that male mean performance is better than female students for 8th grade in selected elementary schools. (4) The combined mean and combined standard deviation of male group was greater than female group in the test, the calculated value of Z-test 1.80 is also smaller than table value 1.96, the difference among male and female mean performance is statistically significant, which means male students are better performer than female students at 8th grade in selected elementary schools in Pakistan. (5) Calculated value of reliability coefficient was more than 50 of the three methods, split-half (0.74), KR20 (0.78), and KR21 (0.70), which means the test was concerned to be reliable to great extent.

Based on all the foregoing process, it is important to mention that these results serve to improve the quality of the test items, strengthen one of the central ideas of English achievement: the assessment of basic capacity and understanding acquired throughout the whole scholastic life. By using them, it is possible to discriminate, and select the best out of all students who want to enter the high schools.

## LIMITATIONS

This research was limited to the subject of English for class VIII, the evaluation on other subjects should also be constructed in this pattern. Emphasis was paid on the prose of the text of textbook only; future researcher



should also include the poetry of the textbook. Students should be aware of the model achievement tests in the classroom. The standardized achievement tests should be used in all schools to rate the effectiveness on the basis of students performance.

## REFERENCES

- Aggarwal JC (1991). Theories and practical of education. New Delhi: Vikas Publishing House. p. 43.
- Airasian P (2005). Classroom Assessment. New York: McGraw Hill, Inc. p. 74.
- Anastasia A (1982). Psychological Testing. New York: McMillan Publishing Co. p. 102.
- Bandura A (1986). Social foundations of thought and action. NJ: Prentice Hall. p. 202.
- Carey LM (1994). Measurement and evaluation School learning. London: Allyn and Bacon. p. 47.
- Ebel RL, Frisbi DA (1991). Essentials of Educational Measurement. New Delhi: Prentice Hall. p. 123.
- Egen P, Kauchak D (1992). Educational psychology: classroom connections. New York: Macmillan Publishing Company. p. 58.
- Gronlund NE, Linn RL (1990). Measurement and Evaluation in Teaching. New York: Macmillan Publishing Company. p. 147.
- Hills J (1981). Measurement and Evaluation in classroom (2<sup>nd</sup> Edition). Columbus: Charles E Merrill Publication Company. p. 36.
- Hudson RA (1987). Sociolinguistics. Cambridge: Cambridge University Press. p. 22.
- Jesa M (2005). Efficient English teaching. New Delhi: Apt Publishing Corporation. p. 66.
- Johnson B, Larry C (2006). Educational Research. London: SAGE Publication. p. 152.
- Leung J (1998). Printing Style and academic Achievement. Columbus: Merrill Palmer Q. p. 45.
- Linacre JM (2002). What do Infit and Outfit, Mean-Square and Standardized Mean? Rasch Meas. Trans. 16(2):878.
- Mangal SK (2002). Statistics in Psychology and Education. New Delhi: Hall of India PVT. p. 85.
- Morales RA (2009). Evaluation of Mathematics Achievement Test: A comparison between CTT and IRT. Int. J. Educ. Psy. Assess. 1(1):19-26.
- Salma W (2000). Elementary Education. Islamabad: Allama Iqbal Open University Publishers. pp. 61-62.
- Shahid SM (2000). Evaluation and Measurement. Lahore: Majeed Book Depot. p. 38.
- Suter WN (2006). Instructions to Educational Research. London: SAGE Publication. p. 242.
- Thorndike RL, Hagen EP (1997). Measurement and Evaluation in Psychology and Education. New York: John Wiley and Sons, Inc. p. 179.
- Wiersma W, Jurs G (1990). Educational Measurement and Testing. London: Allyn and Bacon. p. 169.
- Wright RJ (2008). Educational Assessment. London: SAGE Publication. p. 87.