

# DESIGNING ENGLISH ACHIEVEMENT TEST FOR THE SEVENTH GRADE STUDENTS AT MTS ASH-SHALIHIN

Nur Aliyah Nur<sup>1</sup>, Asma Nurul Inayah\*<sup>2</sup>, Nurfadiah Ramadhani<sup>3</sup>

<sup>1,2,3</sup> Universitas Islam Negeri Alauddin Makassar

Email: [20400120005@uin-alauddin.ac.id](mailto:20400120005@uin-alauddin.ac.id)

## ABSTRACT:

Tests and assessments are always integrated in language teaching and learning as they can be a means of measuring students' knowledge of the language and how they use the language. A large proportion of language tests, especially achievement tests or classroom tests, are paper-based tests. This research is classified as a research and development (R&D) type. The research has been carried out to design an achievement test for the seventh grade students at MTs Ash-Shalihin and tried the test on participants in a class consisting of 29 students. The researchers analyzed and examined the items based on their data presentation, mean and standard deviation scores, reliability (split half), validity, level of difficulty, level of differentiation and fulltime analysis. The result of this research the tests created are basically reliable, valid and not too difficult, in fact some tests are easy. In terms of discriminating power, this test cannot be used to discriminate between clever students and less able students. The researchers need to revise and develop another test with high discriminatory power but the same reliability and validity.

**Keywords: Designing, Achievement Test, Students**

## 1. INTRODUCTION

Language teachers are quite often faced with the responsibility of considering how they are going to measure outcomes and considering what role assessment will play in teaching (Saragih, 2017). A test is part of assessment and it is defined as a method to examine one's ability and knowledge of particular subjects (Brown, 2004). According to Nurdiana (2020), tests and assessments are always integrated in language teaching and learning as they can be a means of measuring students' knowledge of the language and how they use the language. A large proportion of language tests, especially achievement tests or classroom tests, are paper-based tests.

In connection with formal language learning, such as in schools, colleges or universities, achievement tests are mandatory. According to Asrul et al. (2015) daily tests conducted by the teacher in class or even school final exams, have not been able to describe the essence of learning evaluation, especially when it is related to the implementation of the 2013 curriculum. This is because learning evaluation basically does not only assess learning outcomes, but also the processes that educators and students go through in the whole learning process. Besides that, common types of achievement tests are midterms, and final exams. These tests can take various forms, such as multiple-choice questions, fill-in-the-blank and short-form tests, true-false questions, or matching questions.

Achievement tests are universally used in the classroom mainly to measure whether students possess the pre-requisite skills needed to succeed in any unit or whether the students have achieved the objective of the planned instruction, to monitor students' learning and to provide ongoing feedback to both students and teachers during the teaching-learning process, to identify the students' learning difficulties whether persistent or recurring, and to assign grades (Raj, et al.,

2018). According to Brown (2004), achievement tests are designed based on particular units, lessons, or chapters in a course book.

According to Brown (2019), there are several steps involved in the construction of achievement tests. Determine purpose/usefulness, state construct/abilities, draw up specifications, select tasks and arrange them systematically, construct a system of scoring/grading and providing students feedback. In addition, the purpose of feedback for students is to communicate to students to modify thoughts or behavior in order to improve the quality of learning outcomes. Feedback is also a process in which factors that create results can be modified, corrected, and strengthened (Sumarno, 2020).

## **2. METHODS**

A research method is the methodical and scientific process a researcher uses to answer a research question (McClure, 2020). This research is classified as a research and development (R&D) type. According to Gay & Mills (2019), research and development (R&D) is the process of researching consumer needs and then developing products to fulfill those needs. The purpose of R&D efforts in education is not to formulate or test theory but to develop effective products for use in schools. In conducting the designing of the achievement test the researchers used steps to designing an effective test by Brown (2019) namely, determine purpose/usefulness, state constructs/abilities, draw up specification, select task and item types and arrange them systematically, and the last construct a system of scoring/grading and providing students feedback. This research took place at MTs Ash-shalihin, Romangpolong. The researchers carried out the test in the seventh grade students which consisted of 29 students in academic year 2022-2023.

## **3. RESULTS AND DISCUSSION**

### **RESULTS**

Based on the research objectives with reference to the methodology, this research has been carried out to design an achievement test for the seventh grade students at MTs Ash-Shalihin. This research refers to the steps of Brown's design namely, determine purpose/usefulness, state constructs/abilities, draw up specification, select task and item types and arrange them systematically, and the last construct a system of scoring/grading and providing students feedback. The following describes the results of the research and the results of the analysis obtained.

#### **A. Determine Purpose/Usefulness**

An achievement test is related directly to classroom lessons, units, or even a total curriculum. Achievement tests are (or should be) limited to particular material addressed in a curriculum within a particular time frame and are offered after a course has focused on the objectives in question (Brown, 2004). The aim of an achievement test is to determine a student's knowledge in a specific subject area. Achievement tests measure how well students have mastered the subject matter in a course of instruction.

#### **B. State Constructs/Abilities**

Achievement Test is constructed on the basis of the objectives of teaching i.e. knowledge, understanding, and application in the particular subject for which it is to be made. Before constructing the Achievement Test, the teacher should first review the various standardized achievement tests already available for testing in the field of that particular subject. After reviewing them, the latest syllabus is gone through by the teacher (Bawa & Saggu).

#### **C. Draw Up Specification**

In drawing up specifications will simply comprise (a) a broad outline of the test, (b) what skills you will test, and (c) what the items will look like. The results of this study resulted in the development of a blueprint to describe the specifications of the tests carried out. The following is the result of the development of a blue print.

**Table 1.1. Specification Test**

No	Kompetensi Dasar	Learning Outcomes	Materials
1.	<p>3.1 Identify social functions, text structures, and linguistic elements of oral and written interpersonal interaction texts that involve greeting, saying goodbye, saying thank you, and apologizing, as well as responding to them, according to the context of their use.</p> <p>4.1 Compose very short and simple oral and written interpersonal interaction texts that involve greeting, saying goodbye, thanking, and apologizing, and responding to them by paying attention to social functions, text structure, and linguistic elements that are correct and in context.</p>	<ul style="list-style-type: none"> <li>- Listening, imitating, and demonstrating several examples of conversations, with the correct pronunciation and word stress.</li> <li>- Identify the phrase being studied.</li> <li>- Learn to ask unknown or different things.</li> <li>- Get used to applying what is being learned. in interactions with teachers and friends naturally inside and outside the classroom.</li> <li>- Reflect on the process and learning outcomes.</li> </ul>	<p>Greet</p> <hr/> <p>Take Leave</p> <hr/> <p>Say Thank You</p> <hr/> <p>Say Sorry</p>
2.	<p>3.2 Identify social functions, text structures, and linguistic elements of spoken and written transactional interaction texts that involve giving and asking for information related to identity, short and simple, according to the context of</p>	<ul style="list-style-type: none"> <li>- Listening and imitating several examples of self-explanation, with the correct pronunciation and word stress</li> <li>- Identify important phrases</li> </ul>	<p>Personal Identity</p> <hr/> <p>Hobbies</p> <hr/> <p>Members of Family</p>

	<p>its use. {Pay attention to language elements and vocabulary related to family relationships; pronouns (subjective, objective, possessive).</p> <p>4.2 Compose very short and simple oral and written transactional interaction texts that involve the act of giving and asking for information related to identity, short and simple, taking into account social functions, text structure, and linguistic elements that are correct and in context.</p>	<ul style="list-style-type: none"> <li>- Asking things that are not known or different.</li> <li>- Studying examples of texts describing identity by well-known figures</li> <li>- Exposes his true identity.</li> <li>- Listen to each other and ask questions about each other's identities with their friends</li> <li>- Reflect on the process and learning outcomes.</li> </ul>	
3.	<p>3.3 Identifying social functions, text structures, and linguistic elements of spoken and written transactional interaction texts that involve the act of giving and asking for information regarding the name of the day, month, name of the time of day, time in the form of numbers, date and year, according to the context of its use. (Pay attention to the vocabulary related to cardinal and ordinal numbers).</p> <p>4.3 Compose very short and simple spoken and written transactional interaction texts involving the act of giving and asking for information related to the name of the day, month, name of the time of day, time in the form of numbers, dates, and years, with social functions, text structures, and linguistic</p>	<ul style="list-style-type: none"> <li>- Listening and imitating the explanation of the time when circumstances/events/events occurred, including the name of the day, month, name of the time in days, time in the form of numbers, date and year</li> <li>- Says all the names of the days, months, 1st-31st, times, parts of the day, year with the correct pronunciation and word stress, one by one.</li> <li>- Stating verbally the time of occurrence of various circumstances/events/activities</li> <li>- Asking for the day, date, month, and time of occurrence of circumstances/events/activities with the correct language elements</li> <li>- Writing about the times of important events that are known to the public. The results are published in</li> </ul>	Tell the Time
			Tell the Date
			Tell the Day
			Tell the Month

	elements correct and in context.	class or on the school wall magazine - Reflect on the process and learning outcomes	
4.	3.4 Identify social functions, text structures, and linguistic elements of oral and written transactional interaction texts that involve giving and asking for information regarding the names and numbers of animals, objects, and public buildings that are close to students' daily lives, according to the context of their use. (Pay attention to linguistic elements and vocabulary related to articles a and the, plural and singular)  4.4 Compose very short and simple spoken and written transactional interaction texts that involve the act of giving and asking for information regarding the names and numbers of animals, objects, and public buildings that are close to students' daily lives, taking into account social functions, text structures, and linguistic elements correct and in context.	- Examine several short texts containing mentions of objects in class and school to then read with the correct pronunciation and word stress  - Listening and imitating the teacher asking and mentioning objects in the classroom and at school, with the correct grammar, pronunciation and word stress  - Ask and answer about some objects in and around the house  - Read some short texts about the house and its surroundings, especially the presence of objects and animals  - Ask and answer questions about some of the buildings and objects and animals in and around it  - Reflect on the process and learning outcomes	State Things
			State Animals
			State Public Place

D. Select Task and Item Types and Arrange Them Systematically

In this study the researcher used 4 types of test items, namely multiple choice, fill in the blank, picture-cued task, short answer question. The following is a blue print table that has been developed for compiling assignments and test items.

**Table 2.1 Item Types**

Materials	Test Type	No of Items	Item's Number	Score	Total
Greet	Multiple Choice	2	1, 2	1	2
Take Leave	Multiple Choice	3	3, 4, 5	1	3

Say Thank You	Multiple Choice	2	6, 7		1	2
Say Sorry	Multiple Choice	3	8, 9, 10		1	3
Personal Identity	Fill in the Blank	1	1	<b>SECTION B</b>	5	5
Hobbies	Fill in the Blank	1	2		3	3
Members of Family	Fill in the Blank	1	3		5	5
Tell the Time	Picture-Cued Task	2	1, 2	<b>SECTION C</b>	4	8
Tell the Date	Picture-Cued Task	2	3, 4		4	8
Tell the Day	Picture-Cued Task	2	5, 6		4	8
Tell the Month	Picture-Cued Task	2	7, 8		4	8
State Things	Short Answer Question	3	1 – 3		5	15
State Animals	Short Answer Question	3	4 – 6	<b>SECTION D</b>	5	15
State Public Place	Short Answer Question	3	7 - 9		5	15
4 Test Types		30 Numbers		<b>4 Sections</b>	Total Score: 100	

E. Construct a System of Scoring/Grading and Providing Students Feedback

1. System of Score Grading

**Table 3.1 Score Grading Section A**

Item's	Correct answer score	Wrong answer score
--------	----------------------	--------------------

Number		
1	1	0
2	1	0
3	1	0
4	1	0
5	1	0
6	1	0
7	1	0
8	1	0
9	1	0
10	1	0
<b>Total Score</b>	10	

Questions number 1&3

**Table 4.1 Score Grading Section B**

Score	Criteria
5	If students can fill in 5 questions correctly
4	If students can fill in 4 questions correctly
3	If students can fill in 3 questions correctly
2	If students can fill in 2 questions correctly
1	If students can fill in 1 questions correctly
0	If students can't fill the questions correctly

Question no. 2

**Table 5.1 Score Grading Section B**

Score	Criteria
3	If students can fill in 3 questions correctly
2	If students can fill in 2 questions correctly
1	If students can fill in 1 questions correctly
0	If students can't fill in questions correctly

Question No. 1&2

**Table 6.1. Score Grading Section C**

Score for each number	Criteria
4	If the student can answer the question correctly based on the picture
0	If the student cannot answer the question correctly based on the picture
Total max score for each question	4

Question no. 1-6

**Table 7.1. Score Grading Section D**

Score for	Criteria
-----------	----------

<b>each number</b>	
5	If students can mention 3 answer correctly according to the question
3	If students can mention 2 answer correctly according to the question
1	If students can mention 1 answer correctly according to the question
0	If students can't mention 3 answer correctly according to the question
Total max score for each question	5

### Question no. 7-9

**Table 8.1. Score Grading Section D**

<b>Score for each number</b>	<b>Criteria</b>
5	If students can answer the question correctly
0	If students can't answer the question correctly
Total max score for each question	5

### 2. Students Feedback

To find out the feedback test for students, it is necessary to describe it through the frequency distribution, standard deviation, reliability, validity, difficulty level of tests, and discrimination power.

#### a. Frequency Distribution

The following table contains the imaginary scores of a group of 29 students on a particular test consisting of 30 items. The table contains a frequency distribution showing the number of students who obtained each mark awarded: tallies that are the strokes representing the number of students obtaining the same scores: the frequency and the percentage of each score.

**Table 1.2. The Frequency Distribution of Scores**

No	X	F	FX
1	98	1	98
2	97	2	194
3	96	3	288
4	95	3	285
5	94	1	94
6	93	3	279
7	92	2	184
8	87	1	87
9	86	1	86
10	85	2	170
11	84	2	168

12	79	1	79
13	75	1	75
14	64	1	64
15	61	1	61
16	45	1	45
17	34	1	34
18	32	1	32
19	27	1	27
TOTAL		29	$\sum FX = 2350$

b. Standard Deviation

After knowing the frequency distribution scores of female students, then the results of the standard deviation of the test results. The standard deviation measures the degree to which the group of scores deviates from the mean. Standard derivation is useful for providing information concerning characteristics of differ groups (Heaton J, 1988). The following is the result of calculating standard deviation:

**Table 2.2 Standard Deviation**

No	X	D	d <sup>2</sup>
1	98	17.97	322.9209
2	97	19.97	398.8009
3	97	16.97	398.8009
4	96	15.97	255.0409
5	96	15.97	255.0409
6	96	15.97	255.0409
7	95	14.97	224.1009
8	95	14.97	224.1009
9	95	14.97	224.1009
10	94	13.97	195.1609
11	93	12.97	168.2209
12	93	12.97	168.2209
13	93	12.97	168.2209
14	92	11.97	143.2809
15	92	11.97	143.2809
16	87	6.97	69.4909
17	86	5.97	35.6409
18	85	4.97	24.7009
19	85	4.97	24.7009
20	84	3.97	15.7609
21	84	3.97	15.7609
22	79	-1.03	1.0609
23	75	-5.03	25.3009
24	64	-16.03	256.9609
25	61	-19.03	362.1409
26	45	-35.03	1227.1009
27	34	-46.03	2118.7609
28	32	-48.03	2306.8809
29	27	-53.03	2812.1809
TOTAL			$\sum d^2 = 15,847.4752$

$$SD = \sqrt{\frac{15,847.4752}{29}} = \sqrt{546.465} = 23.38$$

$$SD = 23.38$$

c. Reliability

In term of reliability, the writer uses single-test single trial method with split-half reliability, applying Pearson product moment correlation and Spearman-Brown odd even model correlation. In this case, test score of every student is divided into halves (odd and even number), the odd numbers are classified into variable X while the even numbers are classified into variable Y, after that the scores of each half are correlated with the other half, and the coefficient derived is adjusted by means of the Pearson product moment formula. The following is the result of calculating reliability:

**Table 3.2. Pearson Product Moment Correlation**

No.	X (odd)	Y(even)	X <sup>2</sup>	Y <sup>2</sup>	XY
1.	98	97	9604	9409	9506
2.	97	96	9409	9216	9321
3.	96	96	9216	9216	9216
4.	95	95	9025	9025	9025
5.	95	94	9025	8836	8930
6.	93	93	8649	8649	8649
7.	93	92	8649	8464	8556
8.	92	87	8464	7569	8004
9.	86	85	7396	7225	7310
10.	85	84	7225	7056	7140
11.	84	79	7056	6241	6636
12.	75	64	5625	4094	4800
13.	61	45	3721	2025	2745
14.	34	32	1156	1024	1008
15.	27		729		27
	$\sum X = 1211$	$\sum Y = 1139$	$\sum X^2 = 104949$	$\sum Y^2 = 98049$	$\sum XY = 100873$

$$r_{xy} = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}}$$

$$r_{xy} = \frac{(15)(100873) - (1211)(1139)}{\sqrt{(104949 - (1211)^2)(98049 - (1139)^2)}}$$

$$r_{xy} = \frac{1513095 - 1379329}{\sqrt{(104949 - 1466521)(98049 - 1297321)}}$$

$$r_{xy} = \frac{133766}{\sqrt{(-1361572)(-1199272)}}$$

$$r_{xy} = \frac{133766}{\sqrt{1361572 \times 1199272}}$$

$$r_{xy} = \frac{133766}{\sqrt{1632895175584}}$$

$$r_{xy} = \frac{133766}{1277847.869}$$

$$r_{xy} = 0.105$$

The result of this calculation 0.105 shows that there is a correlation between the two sets of scores. The result of this calculation is then analyzed using Spearman-Brown odd even model correlation to see the reliability of the test.

$$r_{tt} = \frac{2r_{hh}}{1+r_{hh}}$$

$$r_{tt} = \frac{2(0.105)}{1+(0.105)} = \frac{0.216}{1.105} = 0.195$$

d. Validity

The researcher takes three items as the samples of the calculation of validity tests (those three samples indicate good, middle, and low validity):

- 1) Determining the proportion of testes correctly answering the analyzed item:

$$p = \frac{\sum x}{N} = 1, 0.8276 \text{ and } 0.2759$$

- 2) Determining the proportion of testes incorrectly answer the analyzed item:

$$q = 0.7241, 0.1724 \text{ and } 0$$

- 3) Calculating the mean score of the total score:

$$M_t = 71.97$$

- 4) Calculating the mean score of testes correctly answering the analyzed item:

$$M_p = 85, 73.30, \text{ and } 54$$

- 5) Calculating the standard deviation of the total score: SD = 23.38

- 6) Calculating the item validity coefficient: 0.6157, 0.4133, and -0.098

After calculating the validity of the data ( $r_{pbi}$ ), there are three kinds of validity that the writer found from the data, as followings:

- 1) Good validity 0.5 – 1 = Section A (1), Section D (1-9)
- 2) Medium validity 0.3– 0.49 = Section A (5, 6), Section B (3), Section C (6)
- 3) Low validity under 0.3 = Section A (2, 3, 4, 7, 8, 9, 10), Section B ( 1, 2), Section C (1, 2, 3, 4, 5, 7, 8).

e. Difficulty Level and Discrimination Power

The difficulty level or the index of difficulty of an item simply shows how easy or difficult the particular item proved in the test. In this case, the writers divides the scores into two halves, 27% the upper group and 27% the lower group. Those two halves are divided based on the students' score, from the highest to the lowest. There are 29 samples of this try-out test, so there are 8 students in the upper group with higher scores and there are 8 students in the lower group with the lower scores.

The discrimination index of an item is the extent to which an item differentiates between high- and low-ability test-takers. The procedure in obtaining the index of discrimination has few differences from the index of difficulty is done. However, they are same in some steps, especially steps 1 (dividing the students into upper and lower group as attached in the separate sheet). To determine the index of discrimination, the writer uses the following formula (Brown, 2008: 59)

$$ID = \frac{\text{Total of students who answer correctly from Upper Group} - \text{from Lower Group}}{\frac{1}{2} \text{ Students}}$$

**Table 4.2. Discrimination Power**

Item	Group	A	B	C	D	KEY	IF	ID	Remark
1	UG	7	1	0	0	A	0.88	0	Easy
	LG	7	1	0	0				
2	UG	1	1	5	1	C	0.50	0.25	Medium
	LG	1	3	3	1				
3	UG	2	3	3	0	B	0.38	0	Medium
	LG	0	3	2	3				
4	UG	7	1	0	0	A	0.63	0.38	Medium
	LG	4	1	1	2				
5	UG	8	0	0	0	A	0.88	0.25	Easy
	LG	6	0	0	2				
6	UG	0	7	0	1	B	0.63	0.50	Medium
	LG	1	3	0	4				
7	UG	2	4	0	2	B	0.56	-0.13	Medium
	LG	1	5	2	0				
8	UG	0	8	0	0	B	0.75	0.50	Easy
	LG	1	4	3	0				
9	UG	2	0	6	0	C	0.63	0.25	Medium
	LG	3	1	4	0				
10	UG	0	0	8	0	C	0.63	0.75	Medium
	LG	4	0	2	2				

## DISCUSSION

Based on the table data deviation above, in this data, the standard deviation of the scores from this test is 23.38 which indicates that the scores are not largely spread out over the mean score 80.03. On the other hand, the data is quite varying since the standard deviation is quite big. As cited from (Heaton J, 1988) that “A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data are spread out over a large range of values”. In addition, if the score of standard deviation is smaller than the mean score it means that the data have no a big deviance each other and vice versa. Reliability of the tests if the resulted calculation ( $r_{tt}$ ) is the same or greater than 0.70, the evaluated test is highly reliable. Conversely, if the resulted calculation ( $r_{tt}$ ) is smaller than 0.70, the evaluated test is not highly reliable. In this case, the test reliability score ( $r_{tt}$ ) is 0.195 which is smaller than 0.70. So, it indicates that the evaluated test is not highly reliable.

Based on table 4.2 and its calculations, almost all questions are at the medium level. Multiple choice questions about number 1 is the most easy item in this test because the score of index facility (IF) the question number 1 is 0.88 and score of index dicrimination is 0. In addition, the level of difficulty contained in this test is based on the syllabus used by the teacher. Item number 1, 5, 8 are easy, and the rest of it are medium. This shows that all the items cannot differentiate between smart students and less able students. Based on the author's observations of students when conducting tests, this can happen because the participants are students with the same level of achievement, so that the abilities possessed by students are almost the same.

## 4. CONCLUSION

After the construction and trial tests took place at MTS Ash-Shalihin, Rompolong. The participants are the seventh grade students for the 2022-2023 academic year. The researchers tried the test on participants in a class consisting of 29 students. The researchers analyzed and examined

the items based on their data presentation, mean and standard deviation scores, reliability (split half), validity, level of difficulty, level of differentiation and fulltime analysis.

In this data, the standard deviation of the scores from the try out test is 23.38 which indicates that the scores are not largely spread out over the mean score of 80.03. On the other hand, the data is quite varying since the standard deviation is quite big. As cited from (Heaton J, 1988) that "A low standard deviation indicates that the data points tend to be very close to the mean, whereas a high standard deviation indicates that the data are spread out over a large range of values". In addition, if the score of standard deviation is smaller than the mean score it means that the data have no big deviation from each other and vice versa.

To interpret the test reliability ( $r_{tt}$ ), Sudijono in Jabu (2008) provides criteria. If the resulted calculation ( $r_{tt}$ ) is the same or greater than 0.70, the evaluated test is highly reliable. Conversely, if the resulted calculation ( $r_{tt}$ ) is smaller than 0.70, the evaluated test is not highly reliable. In this case, the test reliability score ( $r_{tt}$ ) is 0.195 which is smaller than 0.70. So, it indicates that the evaluated test is not highly reliable. After calculating the validity of the data ( $r_{pbi}$ ), there are three kinds of validity that the writer found from the data, as followings: Good validity 0.5 – 1 = Section A (1), Section D (1-9) 2. Medium validity 0.3– 0.49 = Section A (5, 6), Section B (3), Section C (6) 3. Low validity under 0.3 = Section A (2, 3, 4, 7, 8, 9, 10), Section B (1, 2), Section C (1, 2, 3, 4, 5, 7, 8).

Finally, the tests created are basically reliable, valid and not too difficult, in fact some tests are easy. In terms of discriminating power, this test cannot be used to discriminate between clever students and less able students. The researchers need to revise and develop another test with high discriminatory power but the same reliability and validity.

## REFERENCES

- Asrul, Ananda, Rusydi., & Rosnita. (2015). *Evaluasi Pembelajaran*. Bandung: Citapustaka Media.
- Bawa, & Saggi, S. Assessment for Learning Module. *Central University of Punjab*, 5. Retrieved from Central University of Punjab.
- Brown, H. D. (2004). *Language Assesment: Principles and Classroom Practices*. New York: Pearson Education.
- Brown, H. D. (2019). *Language Assesment: Principles and Classroom Practice Third Edition*. New York: Pearson.

- Gay, L. R., E. Mills, G., & W. Airasian, P. (2019). *Educational Research: Competencies for Analysis and Applications Tenth Edition*. Boston: Pearson Education International.
- Heaton J, B. (1988). *Writing English Language Tests*. London: Long Man.
- Manjunatha. (2019). Descriptive Research. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 864.
- McClure, K. S. (2020). *Selecting and Describing Your Research Instruments*. American: American Psychological Association.
- Nurdiana. (2020). Designing an Online Achievement Test for General English Classes. *Journal of English Teaching as a Foreign Language (JETAFL)*, 18.
- Raj, Agarwal, A. D., Jain, K., Sharma, Shukla, Kumar, S., & Sunaina. (2018). Unit-8 Achievement Tests. In A. Dev, K. Jain, Sharma, Shukla, S. Kumar, & Sunaina, *Unit-8 Achievement Tests* (p. 6). New Delhi: IGNOU.
- Saragih, F. H. (2017). Testing and Assessment in English Language Instruction. *Universitas Negeri Medan*, 74.
- Sumarno. (2020). Hubungan Strategi Umpan Balik (Feedback), Motivasi Berprestasi dan Hasil Belajar Dalam Pembelajaran PPKn di SMK. *Jurnal Penelitian Inovasi Pembelajaran*, 41.