
Sahrani
Nida Mufidah
Rusnadi

Antasari State Islamic University of Banjarmasin, Indonesia

This research described the implementation of language assessment principles to evaluate summative test at Madrasah Aliyah Ni’matul Aziz Jelapat Barito Kuala Academic Year 2017/2018. The students in this research are 27 students for class X. The objectives of this research are (1) To know how teacher implements the practicality question for summative test at Madrasah Aliyah Ni’matul Aziz 2017-2018, (2) To know how teacher implements the reliability question for summative test at Madrasah Aliyah Ni’matul Aziz 2017-2018, (3) To know how teacher implements the validity question for summative test Madrasah Aliyah Ni’matul Aziz 2017-2018, (4) To know how teacher implements the authenticity question for summative test Madrasah Aliyah Ni’matul Aziz 2017-2018, (5) To know how teacher implements the washback from summative test at Madrasah Aliyah Ni’matul Aziz 2017-2018. For data collection the researcher did the observation, interview and documentary. The researcher analyzed the data descriptive qualitatively. The result of this research is found that the implementation of language assessment principles on summative test is good. The questions for the test are practice, reliable, valid, authentic, and have a positive washback. Those factors influence the test to be a good test. The teacher has success in making a question for a test. Although the test itself still has a weakness but it is not make the test to be poor in quality.

Corresponding Author: mufidah@gmail.com
Keywords: Assessment; Principles; Implementation; Summative test; Language
INTRODUCTION

Usually, in the middle and the end of semester, teacher has a test or final examination. This test is a summative test. It uses to measure, or summarize, what a student has grasped and typically occurs at the end of a course. A summation of what a student has learned implies looking back and taking stock of how well that student has accomplished objectives, but it does not necessarily point the way to future progress. One of the way to know student’s knowledge by doing an assessment at the middle and the end of year.

Mousavi (2009) states assessment is the way to measure, value level or development of knowledge for a person or student. The result of assessment can be used to measure students’ knowledge including quality of teaching and learning process. This result can be used as references in making syllabus for teaching and learning for the next semester or years.

Ideally, professional teacher should know how to assess the students. This condition is a standard to decide method, technique in teaching and create an efficient process of education. It will give an impact to get the aim of education.

Brown (2009 : 5) stated the quality in teaching and learning process can develop if the system of evaluation has a good quality. Developing quality of teaching and learning process should know student’s characteristic. By doing so, teacher can decide which is the best method and technique for teaching. It will give students more motivation in learning.

Evaluation in teaching and learning has three aspects. There are cognitive, affective, psychomotor. Based on the three aspects, the teacher usually measures student in aspect cognitive. The question to measure student’s cognitive has a relation with student has learned and syllabus. Student cognitive is not the only one to evaluate student’s knowledge.

Brown (2009 : 3) defined test as a subset of assessment techniques. It is prepared administrative procedures that occur at identifiable times in a curriculum. Test can be called a method of measuring a person’s ability, knowledge, or performance in a given domain.
Giving a test to measure something is important. The test to measure student knowledge has some factors which influence student on answering the question. Those factors are student knowledge in understanding material, student preparation on test, and the question of the test which has no relation with material or syllabus. Teacher need to understand technique in making a good question for a test which appropriate with material and syllabus. A good test can be identified by knowing the practicality, validity, reliability, authenticity and washback of the test.

Brown (2009: 7) reminded the function of an assessment depends on the procedure that used. There are commonly divide into two kinds of test, formative and summative. Formative test used to evaluate student in the process of “forming” their ability and knowledge which the aim for making the student to continue that growth process. Summative test is a test which used to evaluate student achievement, knowledge. The test is expected can give a conclusive final report at the end of the project. The report describes or presents the evaluation approaches, instruments, and procedures used, comprehensive assessment, and suggestions about the whole program. Summative reports usually written and detailed also contain data and graphs to complement and explain the message delivered.

Here are some previous studies related with this research. The first research entitled: “Analisis Butir Soal Ulangan Semester Gasal Pada Mata Pelajaran Matematika Kelas XI IPA di MAN Negara Kabupaten Hulu Sungai Selatan Tahun Pelajaran 2013-2014” written by Yusti Fauzi. Yusti’s thesis focus on giving feedback for the teacher to make a good item for a test. The similarity with that thesis is to analyze each item questions on mathematic course. While this research focuses on analyze each questions validity, reliability, and distractor in summative test. The differences with this research are researcher will give some technique to evaluate summative test by giving a questionnaire for the teacher who teach English course.

The second research with entitled “An Analysis of English Items of The Eleventh Grade at Man Negara Hulu Sungai Selatan Regency Academic Year 2009/2010”. written by Dewi Amiaty, Dewi’s thesis talked about the effectives of
distractor. The similarity is on the research in summative test and difference to research the effectiveness of distractor but the researcher focus on principles of language assessment in summative test.

The researcher chooses this school because the teacher at that school has a teacher who graduated from English Department. The teacher has a good knowledge in teaching English based on her background study. The other reason, that teacher does not evaluate his question on summative test such as the practicality, validity, reliability, authenticity and washback of the test.

This research conducted on class IX A. The researcher decides to choice class IX because class IX is a beginner in learning English Course. This condition will decide student future about English subject. When student gets lower score they feel bad with English course. In others cases their friends will judge the students who get lower score as a fool student because they get lower school. The score is not a standard to decide student knowledge. The teacher should know the question for the test whether valid or reliable to the test. Test is expected to measure student knowledge and the material that has been learned.

The other purpose of this test is what happen when student passed summative test, give a washback for student, and motivate student to learn English course. Based on the problem the researcher wants to evaluate the question in final semester to know the practicality, validity, reliability, authenticity, washback. Therefore, this research is aimed to know the assessment principle of summative test conducted in Madrasah Aliyah Ni’Matul Aziz Jelap Barito Kuala Academic Year 2017/2018.

By doing this, hopefully can contribute to Enriching the researcher’s knowledge in analyzing question test and applying researcher’s knowledge on this department, For teacher this research will make teacher to be more professionalism in making an item question for a test, and adding the literacy source of UIN Antasari Library

METHOD

This research used qualitative research method to know how the teacher implements the principle of language assessment in a test for tenth grade senior
high school at Madrasah Aliyah Ni’matul Aziz Jelapat Barito Kuala Regency. The researcher chose this school because the final exam for English subject uses multiple choices and the teacher does not evaluate the question of the test.

Subject of the research is question of the test for tenth grade Senior High School at Madrasah Aliyah Ni’matul Aziz Jelapat Barito Kuala Academic Year 2017-2018, while object of this research is students’ answer sheet final semester of English subject. The researcher analyzed student’s answer to know the practicality, validity, reliability and authenticity.

The source of data that related to this research are respondents such as English teacher, headmaster and students of tenth grade of senior high school at Madrasah Aliyah Ni’matul Aziz Jelapat Barito Kuala. The researcher also need some informants such as the headmaster, student, and administration staff and all written report that related to the data.

The observation conduct on tenth grade senior high school at Madrasah Aliyah Ni’matul Aziz Jelapat Barito Kuala Regency. This technique is used to get the data which is related with statement of problems. The researcher should observe the real situation in the field and gain a better picture of summative test. While observing, the researcher used field note to record the situation. Beside used observation field, the researcher also used observation checklist.

Moreover, the researcher did the interview for English teacher, headmaster and all of the persons who related with the research. Researcher interviewed the headmaster to collect the data about school structure and school infrastructure. The other interview conducted with the teacher who teach English to collect the data witch related with the teacher theory in implementing principles of language assessment in the test. The data such as the paper question for the test, key answer, procedure of scoring, students’ score, etc.

In addition, the researcher also gain the data by documentary which consists of video, document, picture and other document that related with the research. In this research the researcher used documentary to collect data dealing with the principles of language assessment in the test. This technique was aimed looking for all written reports or documents whish related with the research. The data were condition of teacher, administration staff and faculties of the school.
The researcher answers the first question by doing observation to look for information about the teacher’s test at Madrasah Aliyah Ni’matul Aziz Jelapat Barito Kuala Regency. The second, the researcher analyze the result of the interview to get more information how is the implementation principle of language assessment in summative test. The last, the researcher analyze the result all of data from observation, interview and documentary. The results for this research is used to several teachers to learn how to teach English carefully and know how is a good test for evaluating student.

FINDINGS AND DISCUSSIONS

Principles of language assessment can be applied to be formal test or summative test but with the ultimate recognition that these principles also apply to assessments of all kind. This principle uses to measure that a test effective, appropriate, useful, or, in down to earth terms, a “good” test. A good can be identify by testing a test: Practicality, reliability, validity, authenticity, and washback. Before having the result of the research, it is important to note the big picture of this principles.

Practicality

Practicality refers to the logistical, administrative issues involved in making, giving, and scoring an assessment instrument. Following Mousavi, (2009 : 518) practicality include ‘cost, the amount of time it takes to construct and to administer, ease of scoring, and ease of interpreting/reporting the result. Some attributes of practicality : (a) Stays within budgetary limits, (b) can be completed by the test taker within appropriate time constrains, (c) has clear direction for administration, (d) appropriately utilizes available material resources, and (e) consider the time and effort involved for both design and scoring

Furthermore, practicality is the first step before make a test design. Teacher should consider the practicality of the test. Is it appropriate with the attribute of practicality principle.

Reliability

A test that result a consistent and dependable can mention as a reliable test. The test that try to the same student or matched student on two different
occasions, the test should yield similar result. The pattern principles of reliability is include: (a) Consistent in its conditions across two or more administration, (b) has clear direction for scoring or assess, (c) has uniform rubrics for scoring/evaluation, (d) lends itself to consistent application of those rubrics by scorer, and (e) contains items/tasks that are unambiguous to the test taker.

The reliability of the test has the pattern principles. When the question has a clear direction, it will be easier to do and mark student’s answer. The issue of the reliability of test can be better understood by considering a number of factors that make the test to be unreliability. There are four possible factors regarding fluctuation in (a) The student, (b) the scoring, (c) the test administration, and (d) the test itself (Bachman 1990).

**Validity**

Test validity is the most critical factor to be judged and considered in the total of a foreign language testing. A test is valid when it measures effectively what it is intended to measure.

Validity really is not a simple concept; however, the concept of validity reveals a number or aspect, each of which deserves our attention. Hughes (2003) classifies validity into four: content validity, face validity, construct validity, and criterion-related validity.

Content validity is concerned with the extent to which the test is representative of a defined body of content consisting of topics and Moreover, the test should reflect instructional objectives or subject matters. But it is not expected that every knowledge or skill will always appear in the test; there may simply be too many things for all of them to appear in a single test.

Wiersma (1990) divided content validity into two parts, content validity of teacher constructed tests and content validity of published tests. Content validity of teacher-constructed test essentially depends on the sampling of items. If the test items adequately represent the domain of possible items, the test has adequate content validity. When a test is not content valid, there are two consequences. First, the students cannot demonstrate skills that they possess if they are not tested. Second, irrelevant items are presented that the students will likely answer.
incorrectly only because the content was not taught. Both of these consequences tend to lower the test scores; as a result, the test score is not an adequate measure of student performance relative to the content covered by instruction.

The second part is content validity of published tests. Teachers may, at least on occasion, use published tests, some of which accompany curriculum materials. The tests constructed for a specified textbook or set of materials usually have high content validity if the materials are used as intended for instruction. Sometimes materials are used as supplementary and are only partially covered, in which case any accompanying tests would at least need to be reviewed for content validity.

According to Tinambunan (1998) the content validity may be defined as the extent to which a test measures a representative sample of the subject matter content and the behavioral changes under consideration. It can be understood, that in this case the content of the test which interpreted through the test is the important thing and must be able measure what it is intended to measure.

There are two importance of content validity. First, the greater test’s content validity, the more likely it is to be an accurate measure of what it is supposed to measure. Secondly, such a test is likely to have a harmful backwash effect. Areas which are not tested are likely become areas ignored in teaching and learning. The best a safeguard against this is to construct full test specification and to ensure that the test content is a fair reflection of these.

The next is Face validity. Face validity is a surface or appearance of test. As Alderson (2001) stated, face validity refers to the test’s surface credibility or public acceptability. It is mean the test’s surface that will be examined to the students must describe or show the good construct test and it can be acceptability by the examinee. While Hughes (2003) stated, Face validity refers to the appropriateness of test items. It is mean the form of test that given to the examinee should be appropriate and complete instructions.

Substantially, there is no different view among definition above. They would like to elaborate that a test is regarded as having face validity, if its appearance is acceptable, it is clearly readable, and it has a clear instruction in answering the test.
Besides, there is also construct validity. William said that construct validity is concerned with the psychological constructs that are reflected in the scores of a measure or test. It means, the result of testing which has done will be described in the form of scores.

Construct validity deals with construct and underlying theory of the language learning and testing. J.B. Heaton (1990) states that if the test has construct validity it is capable of measuring certain specific characteristics in accordance with a theory of language and behavior and learning. The statement mentioned gives a describing that the test made by teacher where it has construct validity otomatically it can measure certain specific characteristic accordance theory language.

While, Kenneth (1997) said that construct validity is the systematic analysis of tests score designed to assess whether there is a basis for validity. This statement explain that it should be basis, in this case the theory of language and behavior that sistematic in designing the construct validity Criterion-Related Validity

Criterion validity is based on the correlation between scores on the test and scores on a criterion. The corelation coeficient is the criterion validity coefficient. It can be understood, in the criterion validity there is a relation between scores on the test which resulted by the students and the scores that standarized and both can be influence one to another.

According to Hughes (2003) there are two kinds of criterion related validity: concurent validity and last is predictive validity. Concurrent validity is constanted when the test and standarisation are arranged at about same time, Kenneth D. Hopkins (1997), Educational and Psychological Measurement and Evaluation, while the predictive validity is focuss the level when a test can guess examinee’s future action.

**Authenticity**

The fourth major principle of language testing is authenticity, a concept that is difficult to define, especially within the art and science of evaluating and designing test. Bachman and palmer (1996) in Douglas Brown stated authenticity
as “the degree of correspondence of the characteristics of a given language test task to the features of a target language task” and then suggested an agenda for identifying those target language tasks and for transforming them into valid test items.

Many test items types fail to simulate real world tasks. They may be contrived or artificial in their attempt to target a grammatical form or a lexical item. The sequencing of items that bear no relationship to one another lacks authenticity. One does not look very long to find reading comprehension passages in proficiency test that do not reflect a real-world passage. In a test, authenticity may be present in the following ways : (a) Contains language that is as natural as possible (b) has items that are contextualized rather than isolated (c) includes meaningful, relevant, interesting topics (d) provides some thematic organization to items, such as through a story line or episode, and (e) Offers tasks that replicate real-world tasks.

The authenticity of test tasks in recent years has increased noticeably. Two or three decades ago, unconnected, boring, contrived items were accepted as a necessary component of testing. Things have changed. It was once assumed that large scale testing could not include performance of the productive skills and stay within budgetary constraints, but now many such tests offer speaking and writing components. Reading passages are selected from real-world sources that test-takers are likely to have encountered or will encounter. Listening comprehension sections feature natural language with hesitations, white noise, and interruptions. More tests offer items that are “episodic” in that they are sequenced to form meaningful units, paragraphs, or stories.

**Washback**

A facet of consequential validity discussed above is “the effect of testing on teaching and learning” otherwise known in the language assessment field as washback. To distinguish the impact of an assessment, discussed above, from washback, think of the latter as referring almost always to classroom-based issues such as the extent to which assessment affects a student’s future language development.
Following Messick (1996, : 241) emphasized that the washback effect may refer to both the promotion and the inhibition of learning, thus emphasizing what may be referred as beneficial versus harmful (or negative) washback. Alderson (2001) stated washback an important enough concept to define a washback hypothesis that essentially elaborated on how tests influence both teaching and learning. Cheng, Watanabe, and Churtis (2004) told an entire anthology to the issue of washback, and Spratt (2005) challenged teacher become agents of beneficial washback in their language washback. The following factors that provide beneficial washback: (a) Positively influences what and how teachers teach, (b) positively influences what and how learners learn, (c) for learners a chance to adequately prepare, (d) gives learners feedback that enhances their language development, (e) is more formative in nature than summative test, and (f) provides conditions for peak performance by the learner.

In large-scale assessment, washback often refers to the effects that tests have on instruction in term of how student prepare for the test. “Cram” courses and “teaching to the test” are examples of washback that may have both negative and positive effect. The current worldwide use of standardized test for gate-keeping posposes can lead student to focus on simply gaining an acceptable score rather than on language development. On the positive side, many enrollees in test-preparation courses report increased competence in certain language-related task.

The Implementation of The Practicality of Question for Summative Test at Madrasah Aliyah Ni’matul Aziz Academic Year 2017/2018

On the first observation, the researcher came to that school on October 7th 2017. The researcher observed the teaching process. From that observation, the researcher found the teacher taught by using a book “Path Way To English” for senior High School Grade X written by Th. M. Sudarwati and “Look Ahead An English Course” for senior High School Grade X written by Th. M. Sudarwati.

The teacher told that the test would conduct on October 9th 2017. Therefore, the researcher came to the school and joined with the student. The test began at 7:45 WIB until 8:45 WIB. When observed the test, the researcher found that the questions paper consisted of 20 multiple choice questions. This data is supported by the result of interview with the teacher, both stated that the questions...
were made in simple way consisted of 20 questions (Interview, October 14th 2017). The teacher stated to the researcher that he chose multiple choices because it would be easy for giving scoring. The teacher made the question one week before the test conducted. When made the questions, the teacher also made the answer key. When asking about the cost, the teacher told if the cost for the test was very cheap. It just spent for about Rp. 12.000,- and funded by the school.

When observed the test, researcher knew that the teacher gave 60 minutes for answering the questions. It means students just had 3 minutes for answering each question. After submitting their answer, the teacher closed the lesson and did not correct students’ answer on that day. The data is also supported by the interview conducted on October 14th 2017.

There are some criteria of practicality that a test has to fulfill. Based on Mousavi (2009), practicality include cost, the amount of time it takes to construct and to administer, ease of scoring, and ease of interpreting/reporting the result. Based on the interview and observation, the questions and answer key for the test were made by the teacher. The cost for the test was also very cheap and appropriate with school’s estimation.

The teacher chose multiple choices for testing his student. He also made answer key for each question to help him in scoring. The set time frame for the test also appropriate with the number of question.

Based on the observation and interview result, it is clear that the teacher had a good preparation to conduct a test. The teacher had implemented principle of practicality of language assessment.

The Implementation of The Reliability of Question for Summative Test at Madrasah Aliyah Ni’matul Aziz Academic Year 2017/2018

The researcher came to the school on October 9th and observed the test. All of student in the class could read clearly and understood the question. The teacher told that he had printed the paper question to make the student could read the questions.

When the time was running for 40 minutes, some students submitted their answer sheet to the teacher and the rest submitted when the time was over. The
teacher told to his student to submit their answer sheet with the paper question to avoid cheating.

The researcher asked about scoring. He answered that he prepared an answer key when made the questions. It helped him to correct student answer sheet. The teacher also had a rubric for scoring. He also said the rubric made him easy to score students’ answer sheet after corrected it.

The researcher also looked to the student who faced the test in order to know students’ condition when answering the question of the test. All of student seen clam, nice and easy to answer, beside there were also some student seen nervous.

Beside looked the student, the researcher also pay attention to the environment of the class. Around the class where the test conducted there was one class that still need reparation. It made a little noisy out of the class. The teacher immediately closed the door to decrease the noisy around.

When researcher saw the questions, the researcher found some weaknesses for the test. The weaknesses were the test at Madrasah Aliyah Ni’matul Aziz Jelapat Barito Kuala Regency did not put a test such as listening test or speaking test. The grammar for the question was not complete sentence. That was for number one.

1. What is consonant letter......
   
   a. Unvoiceless  b. Vowel  c. Like B,C and D  d. Like A and E

   e. alphabet

   For option a. Unvoiceless that word still a strange word for student who sit in tenth grade Senior High School. They did not know about voice and voiceless. They also have not learned suffixes and prefixes yet. The last weakness was in the paper questions itself. There was not a direction or guideline to answer the question. The finding of this data got from observation and also supported by the interview when conducted on October 14th 2017.

A reliability makes a test to be a good test. There are some factors that made the test to be reliable, namely student, scoring, test administration and the
test itself (Brown, 2009). Looking to the data on observation and interview with the teacher at the school, test at Madrasah Aliyah Ni’matul Aziz can be called as a reliable test. The test at that school was consistent, the question was not leak, paper question was clean, had a rubric for scoring, the student conditions and the environment of the class was good. Although there still had a weak in reliability such as did not put listening and speaking test and the grammar for the question was not a complete sentence.

**The Implementation of The Validity of Question for Summative Test at Madrasah Aliyah Ni’matul Aziz Academic Year 2017/2018**

The researcher interviewed the teacher on October 14th 2017 about the validity of question in the test. He told that this test aimed to evaluate student knowledge on grammar, vocabulary and reading skill. He divided the question into grammar, vocabulary and reading skill.

The first part was for general knowledge and grammar session. It consists of 12 questions. Here is one of the example,

2. *How many totally of consonant ....*

   a. 5  b. 21  c. 26  d. 2  e. A,E,I,O,U

The teacher could evaluate students’ knowledge on grammar based on the questions number 1-12. He checked which numbers or questions students got wrong answer the most. It gave the information that students still did not understand the material or forget the material.

The second part was for vocabulary. The questions for this part were number 13 to 17. It consists of procedure test about making a *popsicles*. Here is the example,

13. *What is the purpose of the text?*

   a. to explain the use of paper cups  
   b. to explain how to serve popsicles  
   c. to explain how to use juice in food  
   d. to describe popsicles  
   e. to explain how to make popsicles
Procedure text had many vocabularies. He used this part to evaluate students’ knowledge on vocabulary by giving a procedure text. The teacher told that if the students know the meaning of vocabulary in procedure text they could arrange the procedure and understood the process.

The last part was for reading skill. Number 18-20 was reading section. The question was story line. Here is the example,

**18. Who was the naughtiest rabbit?**


The last part or for number 18-20 was reading section. The question for this section student was not only needed a good skill reading but also a good knowledge on vocabulary. The teacher said the question for this part force student to read the text repetitious when answered the questions. The purposed for this part was if student had a good skill on reading and knew the meaning of the story they could be easy to answer the question. Due to on that question the teacher could guess students’ skill on reading.

The result of this observation was also supported by the interview result with the teacher. The questions for the test were taught by the teacher. The test was appropriate, useful and meaningful for the student. The questions related with the material and it measured students’ skill in reading, vocabulary and grammar.

The English teacher at Madrasah Aliyah Ni’matul Aziz Jelapat Barito Kuala regency conducted a test to evaluate students’ knowledge and skill. Based on the data which got from the interview and observation, the teacher made a question to evaluate students’ grammar, vocabulary and reading skill. The number of question was 20 questions that divided in three parts. Those parts were for grammar, vocabulary and reading skill. Validity as the important point in a test can make a test to be effective and arguably. The extent of validity made assessment result to be appropriate, meaningful, and useful in terms of the purpose of the assessment (Gronlund, 1990). This test can be called as valid. The questions were appropriate with teacher’s target and the test was useful and meaningful.
The Implementation of The Authenticity of Question for Summative Test at Madrasah Aliyah Ni’matul Aziz Academic Year 2017/2018

The test was made by the teacher himself based on the material which had been taught in the class. He constructed the question as simple as he could to make his student understood. Here were the examples of that question.

1. What is consonant letter......
   b. Unvoiceless  b. Vowel  c. Like B,C and D  d. Like A and E
   e. alphabet

2. How many totally of consonant ....
   a. 5  b. 21  c. 26  d. 2  e. A,E,I,O,U

3. Rina .... me ..... were his students last year.
   a. not only and but also  b. And and Both  c. Both and And
   d. Either and or  e. Neither and nor

4. She can’t attend her class ....... she is sick.
   a. because  b. While  c. so  d. In order that  e. When

The other question for number has been taken from a website. He put the link where he took the questions. Here are the example,

*The text is for questions 13 – 17*

*How to make a popsicles*

*Instructions*

*Things you need:*

- Aluminum foil
- Fruit juice or other liquid
- Paper or plastic cups
- Handles

1. Collect some small cups to use as popsicle molds. You can use paper cups, Plastic cups or empty yogurt cups.
2. Pour any sort of fruit juice you like into the small cups, about ¾ full.

3. Make a lid for each cup for a small amount of aluminum foil. Cover each cup with the foil, sealing to prevent leaks.

4. Put something in the cup to use as a handle. This could be recycled popsicle sticks or chop sticks, or plastic spoons or forks. Do this by piercing the center of the aluminum foil with the handle and pressing it straight down until it hits bottom.

5. Put the cups in the freezer and let them freeze until hard.

6. Take the cups out of the freezer, remove the popsicles and eat them. The popsicles can be removed easily by letting them sit for a bit to melt slightly or running them under hot water.

(http://www.ehow.com/how2043741makepopsicles.html#ixzz1MPwr9fg)

The link where he took the question put under the procedure text. The teacher said he put the link to make who read the question could access the question from the original sources. He took that topic because the content of procedure text is general and student knew the procedure. It helped student to understand the topic and questions.

The result of this research is supported by observation and interview result with the teacher. The question for the test made one week before the test conducted. The teacher had a good preparation to make the question. The topic for the test was famous and student knew the material. Structure for the sentences in the question was short and simple student could answer clearly.

Bachman and Palmer (1996) in Brown defined authenticity as “the degree of correspondence of the characteristics of a given language test task to the features of a target language task” and then suggested an agenda for identifying those target language tasks and for transforming them into valid test items. The language should be easy to understand and useful for student to the future.

The questions which used for summative test at Madrasah Aliyah Ni’matul Aziz Jelapat Barito Kuala Regency is made by the teacher. The teacher made the question based on the material he had been taught before the test conducted. The
language for question in the test was clear and short. Student could understand all the questions. The teacher put the website for question which he got from the website. He did it to tell the sources where he got the question.

**The Implementation of The Washback of Question for Summative test at Madrasah Aliyah Ni’matul Aziz academic year 2017/2018**

When interviewed the teacher on October 14th 2017, researcher knew that teacher gave student answer sheet and their score on the next meeting. The teacher wrote a feedback to the student by writing a note on students’ answer sheet. Here were the example of teacher’s feedbacks: *excellent, good job, study hard, focus on study do not speak when the teacher was explaining.* Another way to give feedback was by telling the students in front of the class: *For student who got high score please defense it, for students who got low score or medium score please study hard and enhance your score. You can get better if you do the best for your effort.*

The teacher told students’ weakness when shared their score. Here were some examples of student weakness (a) a half of student got wrong answer for question number one. The teacher told to do not be nervous when answering the questions because it will make them lost of focus. (b) The question for number 13-17 was a procedure text which had been learned before. He told to the students to study at home for repeating the material. Students’ weakness for this section was lack of vocabularies in the text. Teacher gave advised to students for improving their vocabulary by writing a new vocabulary on their diary or book to remember the vocabulary which had been learned. (c) The question for number 18-20 was reading part, students must improve their skill in reading by learning more. The teacher said they could practice to read a book or story line to improve their skill on reading. From the interview, the researcher found that score of students for the test were various. There were some students who got high score and low score. The teacher told to the researcher that he found a positive washback from the test. It can be seen for the student who got high score. They had more motivation to study hard and maintain their score. Beside, the other students who got low score also showed a positive washback. They seemed focus on study and asked the material if they were not understood with the material.

*Sahrani*

*LET: Linguistics, Literature and Language Teaching Journal Vol.8 No.2 2018*
When shared students’ answer sheet and score, the teacher also gave a homework for student who got low score. The question for the task was the question of middle test itself. The student answered the question alone at home and should submit the answer next meeting. The teacher said that purposes of the homework were to give a chance to students to know their mistake when answered at the middle test and improve their score by doing the homework.

The result of this research is supported by the interview result with the teacher on October 14th 2017. The teacher conducted the test to evaluate students’ knowledge and his method, technique, approach in teaching learning effective or not. The test evaluated all of the process teaching and learning. It influenced teacher and students’ motivation in teaching and learning.

In the end, the test for tenth grade Senior High School at at Madrasah Aliyah Ni’matul Aizz Jelapat Barito Kuala Regency had been implemented the principles of language assessment in a test. The principles of language assessment were practicality, reliability, validity, authenticity and washback. The implementation for principles of language assessment itself made the test to be appropriate and good test.

After conducting a test, the teacher can see the effect of testing on teaching and learning. It had known in the language assessment field as a washback. The washback effect may refer to both the promotion and the inhibition of learning, thus emphasizing what may be referred to as beneficial versus harmful (or Negative) washback (Messick, 1996 : 241).

Based on the data, washback from the test at Madrasah Aliyah Ni’matul Aziz Jelapat Barito Kuala Regency got a positive washback. It can be seen from students’ motivation and discipline when learning English. Student felt enthusiastic in the class and asked the material which they did not understand. It proved the test have given a positive washback.

CONCLUSIONS AND SUGGESTIONS

Conclusions

From all of the data described about implementation principles of language assessment in that test at Madrasah Aliyah Ni’matul Aziz Jelapat Barito
Kuala Regency, it can be concluded that teacher has successfully implemented the principle of practicality. Beside, the principle of reliability also seen clearly in this test. Moreover, the principle of validity and authenticity appear as well on the questions used by the teacher. In terms of washback, this summative test has given a positive washback.

**Suggestions**

There are some suggestions that can be proposed by the researcher. First, the teacher can make a pre-test before conducting a middle test, in order to measure students’ understanding of the material and as reference to make the question for middle test. The pre-test can be a question for homework if there is no time to conduct in the class. Second, when giving a formative or summative test, teacher should give a feedback to the student. The feedback can be applause in the class, say ‘good job’, praise to the student. It can motivate the student to be more diligent and they will feel their effort is valuable. Third, the teacher can make other test which can improve or evaluate all of student skill. The test can accommodate all skill like reading, speaking, writing, listening, grammar, vocabulary, etc. The test was not only focus on some skill like reading, grammar and vocabulary, but the test content can exercise all of students’ skill.

**REFERENCES**


*Sahrani*

*LET: Linguistics, Literature and Language Teaching Journal Vol.8 No.2 2018*


