EFL Teachers' Problems in Constructing English Mid-Semester Test at State Junior High School 12 Padang

Martin Kustati[1]; Vivi Zurniati[2]

[1]Universitas Islam Negeri Imam Bonjol Padang, Sumatera Barat, Indonesia.
[2]Universitas Islam Negeri Imam Bonjol Padang, Sumatera Barat, Indonesia.

<martinkustati@yahoo.com>
<vivizurniati@yahoo.co.id>

Abstrak: Tujuan penelitian ini adalah untuk mengetahui masalah guru EFL pada pembuatan soal Ujian Tengah Semester Bahasa Inggris di SMPN 12 Padang. Penelitian ini menggunakan metode deskriptif kualitatif, dimana data dikumpulkan melalui Wawancara dan analisis dokumen. Wawancara dilakukan untuk menentukan masalah guru dalam pembuatan soal, sedangkan analisis dokumen digunakan untuk triangulasi hasil wawancara. Peneliti memilih dua guru bahasa Inggris yang mengajar di kelas VII. Temuan penelitian ini menunjukkan bahwa pembuatan soal Ujian Tengah Semester Bahasa Inggris di sekolah ini didasarkan pada asosiasi prinsip-prinsip, di mana setiap guru mata pelajaran Bahasa Inggris menghadiri pertemuan untuk membuat soal ujian bersama. Desain ini digunakan oleh sekolah yang mengirimkan guru-gurunya dan sekolah lainyang tidak menirimkan guru-gurunya pada pertemuan pembuatan soal. Disarankan dalam pembuatan soal uijan tengah semester; harus dibebankan kepada guru dan komite sekolah karena semuanya tahu persis kemampuan dan pemahaman siswa. Para guru juga dapat merancang tes dengan memberikan item khusus untuk menggali lebih banyak tentang pemahaman siswa mereka pada setiap mata pelajaran..

Kata Kunci: *Ujian Tengah Semester Bahasa Inggris, Guru EFL, Bahasa Inggris sebagai Bahasa Asing.*

Abstract: The objectives of this study are to find out the EFL teachers' problems in constructing English Mid Semester test at SMPN 12 Padang. This research used descriptive qualitative method, where the data were collected through Interview and document analysis. The interview was conducted to determine the teachers' problems in constructing the test, while, the document analysis to triangulate the result of interview. The researcher chooses two English teachers who teach at class VII. The finding of the study shows that the construction of mid-semester test in this school is based on the principles' association, where each of them attends the meeting to make the test together. This design was used by this school that sends their teachers and also others that the teachers do not attend the meeting. It is recommended that in making mid semester test; it should be giving to the schools' teachers and committee since all of them know precisely the students' ability and comprehension. The teachers may also design the test by giving special items to dig more about their students' comprehension on each subject matter.

Keywords: *English Mid-Semester Test; EFL teacher; English as a Foreign Language.*
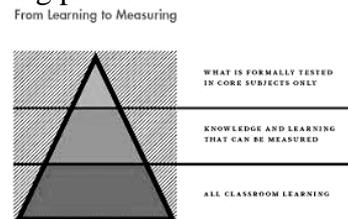
1. INTRODUCTION

The long history of testing has dyed the venture of investigating of human ability. The test methodologies have been studied and inflated by numbers of experts and test practitioners as so many efforts they have done in order to make it meliorate[1]. Standing over these circumstances, the adoption within the educational word is led into a condition of the inseparable factor between test and teaching and learning. In the globe of teaching and learning, the testing has been considered into a particular aspect. It plays a powerful role since that is an arc to enable the arrow dashed.

We take a look into an example in the place where the test has taken a long history. The test initiation and development in America has been started since the nineteenth century. Ravitch[2] stated that the ninetieth century-school tested the students to see if they had mastered what they had taught, and for those who did not pass would leave back.

Nowadays, testing as a hemisphere within the proceedings of teaching and learning in Indonesia where it plays an important role as well. This condition leads to an understand-concept that a test should be known by those who are in education system started from the stakeholders, teachers, and everyone who relates with the test's proceedings. Testing is one of the essentials elements within the process of teaching and learning[3]. All of them cannot be separated one another. The running of teaching and learning, in particular, is directed toward the achievement of those teaching and learning. One another have coupled where the test is used as a binocular tool of whether the tested or in this case is the student has reached the target or whether they are not able to the target yet.

Based on what has been asserted above, the link between testing and teaching and learning based on several experts including Bachman[4] in the following picture:

The picture shows that the test is placed in the top of the triangle. The

[1] Lyle F. Bachman, "Modern Language Testing at the Turn of the Century: Assuring That What We Count Counts," *Language Testing* 17, no. 1 (2000): 1–42; Lyle F. Bachman and Adrian S. Palmer, *Language Testing in Practice: Designing and Developing Useful Language Tests*, vol. 1 (Oxford University Press, 1996); Liying Cheng, "The Key to Success: English Language Testing in China," *Language Testing* 25, no. 1 (2008): 15–37.

[2] Diane Ravitch, "A Brief History of Testing and Accountability," *Hoover Digest* 4 (2002): 35–39.

[3] Martin Bygate, Peter Skehan, and Merrill Swain, *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing* (Routledge, 2013); Barry J. Fishman et al., "Linking Teacher and Student Learning to Improve Professional Development in Systemic Reform," *Teaching and Teacher Education* 19, no. 6 (2003): 643–658.

[4] Bachman, "Modern Language Testing at the Turn of the Century"; Bygate, Skehan, and Swain, *Researching Pedagogic Tasks*; Liying Cheng and Andy Curtis, "Washback or Backwash: A Review of the Impact of Testing on Teaching and Learning," *Washback in Language Testing: Research Contexts and Methods*, 2004, 3–17; Dianne Wall, "The Impact of High-Stakes Testing on Teaching and Learning: Can This Be Predicted or Controlled?," *System* 28, no. 4 (2000): 499–509.

second level is the knowledge that can be measured. It is a kind of knowledge may in form of theory or practice. In the lower level of education, the knowledge tested is according to the system of that educational system itself. As the instance is the test within the science major, students may have practical test in adjoining the theory one. The first level in which also becomes the basic level of all is the learning process in the classroom. This process is the fundamental of the knowledge and the test. The rest of the two levels are based on the first level that is the learning process. Therefore, the learning process will determine the knowledge in the next level and test at the top.

In line with the previous concept, the teaching and learning of language also has also formulated by all practitioners. English language testing as one of subject matter within teaching and learning is also tested in order to measure the students' comprehension and so their achievement on it. It is seen that whether the students are having good understanding on the material or they may still need to learn more the material about. So, the test within English language process is essential either as the tool to get the students' score, assess them, ant at last evaluation as the decision maker.

To sum up, the test and teaching and learning process are inseparable inside the educational system. Every people involved should have a good understanding on this conception. The use of the test should have a clear application in order to make the learning process itself can run well and so the students achievement.

2. REVIEW

The definition of test has been stated by many experts and it has also been defined by the word accordingly. Bachman[5] define a test as a procedure intended to establish the quality, performance, or reliability of something, especially before it is taken into widespread use. This definition shows that the test is used to determine the quality of something. This word-definition is guiding into an understanding that a test is usually used to see quality. In the world education, based on this definition, a test is used to see the students' quality. The test is occasionally used to see whether the student's quality is good or not.

Another definition of test is also noted by Salvador [6] that the test construction was originally driven by an interest in the measurement of mental abilities. Their conceptualization drove the technology

---

[5] Bachman, "Modern Language Testing at the Turn of the Century"; Arthur Hughes, *Testing for Language Teachers* (Ernst Klett Sprachen, 2007); Cyril J. Weir, "Language Testing and Validation," *Hampshire: Palgrave McMillan*, 2005.

[6] Salvador Algarabel and Carmen Dasí, "The Definition of Achievement and the Construction of Tests for Its Measurement: A Review of the Main Trends," *Psicologica* 22, no. 1 (2001); Zoltán Dörnyei, *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition* (Routledge, 2014); Zoltán Dörnyei and Tatsuya Taguchi, *Questionnaires in Second Language Research: Construction, Administration, and Processing* (Routledge, 2009); Guillermo Solano-Flores and Elise Trumbull, "Examining Language in Context: The Need for New Research and Practice Paradigms in the Testing of English-Language Learners," *Educational Researcher* 32, no. 2 (2003): 3–13.

that for a long time has been applied to test construction. When achievement began to be measured, the principles of test construction applied were identical to those used in the measurement of abilities.

This definition shows different point of view from the previous one. This definition is more concerning on the mental abilities. Unless the first definition is stated about the quality, the second one is telling about the mental ability. This though tells us more about the test that it is about the barometer of measurement the mental ability of human.

If it is seen more deeply about this definition, we will find that this definition is related to the human psychology. The expert that has noted this definition pressed more the point of view about the test on the psychology that may involve the process of test itself. This one is under the reason that psychology is an important factor within the text that extremely influences the human ability in doing a test. As other experts of psychology have mentioned that human psychological condition is able to influence the ability to think in this case is to learn.

As one of an extreme example of this statement is environment test. Although this example is not closely related to the psychology, this example deals with factors that also influence the psychological aspect. Even the environment could possibly involve the test, how can we imagine that the psychological one.

The environmental test is dealing with the place where a test is done. This factor directly involves the human ability to do a test. As the instance is a test taker who do the test in silent place,

let say in the classroom, will have different score with the student who do a test in the front desk because many people passing through. This situation and condition will result different outcomes. Even the environment can influence the test taker score and how is the psychology one. It is surely that it involves it extremely.

One test taker that in a good psychological condition will do the test best rather than the test taker that is not in a well situation. Therefore, this definition has brought into an understanding that the test is also about knowing the quality of psychological circumstance.

1. Type of Tests

Bachman[7] stated about test types that a test taker may have. These kinds of test are based on the purpose of the test. They are as the following:

Proficiency Test

The purpose of proficiency test is to test global competence in a language. It tests overall ability regardless of any training they previously had in the language. Proficiency tests have traditionally consisted of standardized multiple-choices item on grammar, vocabulary, reading comprehension, and listening comprehension. One of a standardized proficiency test is TOEFL.

---

[7] Bachman, "Modern Language Testing at the Turn of the Century"; James E. Purpura, "Investigating the Effects of Strategy Use and Second Language Test Performance with High-and Low-Ability Test Takers: A Structural Equation Modelling Approach," *Language Testing* 15, no. 3 (1998): 333–379.

Diagnostic Test

The purpose is to diagnose specific aspects of a language. These tests offer a checklist of features for the teacher to use in discovering difficulties. Proficiency tests should elicit information on what students need to work in the future; therefore the test will typically offer more detailed subcategorized information on the learner. For example, a writing diagnostic test would first elicit a writing sample of the students. Then, the teacher would identify the organization, content, spelling, grammar, or vocabulary of their writing. Based on that identifying, teacher would know the needs of students that should have special focus.

Placement Test

The purpose of placement test is to place a student into a particular level or section of a language curriculum or school. It usually includes a sampling of the material to be covered in the various courses in a curriculum. A student's performance on the test should indicate the point at which the student will find material neither too easy nor too difficult. Placement tests come in many varieties: assessing comprehension and production, responding through written and oral performance, multiple choice, and gap filling formats. One of the examples of Placement tests is the English as a Second Language Placement Test (ESLPT) at San Francisco State University.

Achievement Test

The purpose of achievement tests is to determine whether course objectives have been met with skills acquired by the end of a period of instruction. Achievement tests should be limited to particular material addressed in a curriculum within a particular time frame. Achievement tests belong to summative because they are administered at the end on a unit/term of study. It analyzes the extent to which students have acquired language that have already been taught.Moreover, Bachman in Rahman and Gautam state about the characteristics of test that are included in several points. First, test can be distinguished according to their intended use, such as selection, entrance, readiness, placement, diagnosis, progress, attainment and mastery. Then, tests can differ in content; Achievement tests are based on syllabus, while a proficiency test derives a theory of language ability. Then, different frames of reference can provide the basis for test development and score interpretation norm referenced tests are developed to maximize differences among individual test takers and a test score is interpreted in relation to the score of the test takers. Third, tests can be classified according to the scoring procedure (the act or process of evaluating responses to test situations or evaluating characteristics of whomever or whatever is being rated. It consists of checking the student's response to each item to see if it is correct. Scoring objective tests is purely mechanical process which requires no special skills); objective tests require no judgment on the part of the scorer but in subjective tests, the scorer must judge the correctness of the test taker's respons. Beside, Black and Dylan [8]

---

[8]   Paul Black and Dylan Wiliam, Inside the Black Box: Raising Standards through Classroom Assessment (Granada Learning, 2005); Bygate, Skehan, and Swain,

stated about the importance of testing for teacher in educational system. He said that the teacher is responsible for making many decision's that effect child's education – the sequence of material to be taught, activities and books, to be used and ways to organized students to learn. In addition, teachers must evaluate each child's progress over the time.

Designing classroom language tests deal with some guidelines and process on how to make and revise the existing tests. This subject matter aim at the ability of the test's maker in designing a good language test as well as to revise a test that has been made before. Moreover, in designing the test, there are six questions to be answered by the test's maker. The questions are dealing with some aspects of test such as the following:

The purpose of the test
- The objective of the test
- Objective and purpose specification
- Selecting the test item types
- The administration of the test
- The kind of scoring, grading, and/or feedback

These five questions should form the basis of designing, administering, and making maximum the test

2. Four Assessment Scenarios

Furthermore, the first question above is related to the purpose of the test. There are four assessment scenarios to arise the within real classroom situation. The

Researching Pedagogic Tasks; John R. Frederiksen and Allan Collins, "A Systems Approach to Educational Testing," Educational Researcher 18, no. 9 (1989): 27–32; Diane Ravitch, The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education (Basic Books, 2016); Dianne Wall, "The Impact of High-Stakes Testing on Teaching and Learning: Can This Be Predicted or Controlled?," System 28, no. 4 (2000): 499–509.

following kind of activities may pop up particular result which enables the test's makers to see the real condition of their testee. Those kinds of works are such as reading quiz, grammar unit test, midterm essay, listening and or speaking final exam.

The above scenarios will make the test's maker see how far the testee's ability is. It is hoped that through seeing the result of those activities, some concepts in making a good test will be reached.

3. Determining the Purpose of A Test

To determine the purpose of a test is about to determine the usefulness of the test itself. To make sure that the test has the right purpose and usefulness, the test's maker need to take a look the following point. They are to administer a test in the course; determine the relative significance to the course; mark the end of the lesson, unit, or period of time; the comparison to the other students' performance; the use of the result to determine whether the students meet certain determine curricular standard; the students to be recipients of beneficial washback; the use of the result as a mean to allocate pedagogical efforts in the days or weeks to follow; and the impact for the teacher, students, before and after the test. Some of these points are becoming some consideration in determining the purpose and usefulness of a test. A right-purpose test will be able to answer the above points as well as to follow them.

4. Designing Clear, Unambiguous Objectives

In determining the right purpose of the test, it deals with what test maker need

the students do. One way which he/she can do is by looking back up to the course of the class about what has been studied. This point means that what materials have given to the students become the navigator to design the test. As the instance is the tester wants the testee to do an oral performance, in line with the materials, then after that he/she designs the test which tests the oral ability. To do so, the test's maker can list the purpose of the test based on the classroom teaching and learning has gone through. If one gets lucky, he/she may be able to determine it when the material has been studied. On the other side, if he/she less lucky he/she can make the list of test purpose in line with the given materials.

5. Drawing Up Test Specifications

Test specifications mean a blue print of the test. This blue print will become the guideline plan for fulfilling the instrument to use in the test. There are some items inside of a blue print,-a description of its content; item types; tasks; skill to be included; how the test will be scored; and how it will be reported to the students. After determining these guidelines, it is hoped that they may able the test's maker to determine what want to be achieved so that the right purpose can be got.

6. Devising Test Items

This point contains a warning that a test is not always a clear and linear process. Sometimes, a shift may appear to destroy the rightfulness of a test. Therefore, the test maker needs to note the guideline of the test making just some points that have been asserted above. Making a clear purpose of the test by looking back up to the materials

or even blue print of the test are important action to do by a test's designer.

7. Designing Multiple-Choice Items

In designing multiple-choice items, the test makers need to note the technique tests only recognition knowledge. Next, their guess may have a considerable effect on test scorer and the technique severely restricts what can be tested. The other weaknesses, it is very difficult to write successful items and beneficial washback may be minimal. Cheating may be facilitated

Furthermore, there are four points of guidelines for designing multiple-choice classroom-based and large-scale situation. First, they design each item to measure a single objective. Second, they state both stem and options as simply and directly as possible; Next, they make certain that the intended answer is clearly the only correct one. Next, they (optional) use item indices to accept discard, or revise items, and this forth point has three indices. Item facility (IF) – deals with the difficulty or easiness of each item; item discrimination (ID) – deals with item's differentiate between high and low test's taker; ND Distracter efficiency (DE) – deals with destructor "lure" a sufficient number of the test's takers, especially lower-ability ones, and those responses are somewhat evenly distributed across all destructors.

8. Administering the Test

Here are some considerations to ensure the administration of the test:
   a. Provide appropriate pre-test information on the conditions for the test:

1) The materials that students should bring with them.
2) Kinds of the items that will be on the test.
b. Suggestions of strategies for optimal performance Evaluation criteria
   1) Offer a review of components of narrative and description essay
   2) Give students a chance to ask any questions, and provide responses

   For more specific administration details, the students should arrive early; try everything out it advance (technology devices); have extra paper:
   1) Start on time.
   2) Distribute the test itself.
   3) Remain quietly.
c. Remain the students to complete their work when the time is out, Scoring, Grading, and Giving Feedback.

Scoring, grading, and giving feedback are important aspect of test. They will reflect the test itself. Therefore, the test's maker cannot ignore this last process of test. These three processes are started from the scoring, then continue to the grading and finally end with giving feedback. The concept of feedback is the same with the washback.

Tests and assessments, however, are two dimensional things. Both of them are not containing the exactly same aspects as the test-takers gotten. It is stated that every assessments carry the test, and not every test contain the assessment. In other word, it can be said that assessments bring the test every time a test-maker making and designing it. This theoretical principle of test and assessment set a snow-balling effect toward those tests and assessments. The assessments in which constitute of some kinds of types may become the alternatives for tests as the aspects of assessments contains the test itself.

The following discussions are talking about some alternatives from assessment toward the tests so that a test is able to do. Some types within the assessments are enabling the teacher or other people who need to do a test testing the students and other testees.

In designing a test, the standardization is important. Furthermore, in the educational world, such as what has been asserted above, the standardized-test is a kind of test that is acceptable. In the field of education, which is a formal one, the standardization is crucial. The students will not be allowed to get the un-standardized kind of test. The stakeholders of an instructional program will do their best in giving anything to their students.

Some countries have done their best effort to do so. As one of the instances is United States (US) in which they have a program called NCLB (No Child Left Behind) as the standardization of their education, and including the test as well. The standard of government program has standardized the education itself. It means that that NCLB has standardized the education system.

In Indonesia, the education is standardized by the BNSP (Badan Nasional Standar Pendidikan). There are 8 points which are standardized in this institution. They are as the following:

1. Standard of graduate-competence

2. Standard of content
3. Standard of process
4. Standard of educators and staff
5. Standard of infrastructure
6. Standard of management of education
7. Standard of financing
8. Standard of evaluation

This standard of education will also be able to standard the test. Point number eight from eight points above sets the test as one of its elements. Having some points above, it can be inferred that the test within the teaching and learning process is crucial and essential. Thus all the above conception should be comprehend by the test maker and so the teachers who often use the test within the process of teaching and learning.

3. METHODS

This research is a field research. The determination of the sample in qualitative research according to (Miles, Huberman, & Saldana, 2013) does not only consider the person to be interviewed, but also considers the setting, events, and social processes that occur in the community. These four things are the parameters in determining these research informants.

The aim of this study is to provide empirical evidence of the problems faced by Four English teachers in SMP 12 Padang in constructing English Mid-Semester test. Descriptive Qualitative research was used in this study where the data taken from Interview and document analysis.

Data analysis was done qualitatively, in order to describe and explain in detail to see the

implementation of Arabic Intensive Program, especially in regard to the planning, implementation, and evaluation process. Geetz (1983) explained that basically, qualitative analysis seeks to find the meaning that refers to the knowledge in its function as an individual motor in social life. Indeed, according to Spradley (1980), qualitative analysis cannot be separated from the context of the events observed. It can be identified based on the place where events, behaviors, and activities of the community occur. Qualitative data are analyzed by steps: first, the model of writing memos at the time of the research; second, data coding; third, searching for themes and categories; fourth, discussing data; and fifth, drawing conclusions, (Alwasilah, 2002).

4. FINDING AND DISCUSSION

In the field of test implementation, there are many possibilities can occur during the making, the use, and even the result that the test may produce. All of those factors involve so many elements started from the farthest until the nearest one toward the test. All of them are the key factors of the test itself as they are responsible to the test that they have made.

The test, however, is implementation. It should be based on the theory so people can practice it. In making and designing the test such as the mid semester test or even the semester one, every school in different district has dimensional ways. In Padang, the making of the test is based on the MKKS (Musyawarah Kesatuan Kepala Sekolah). This association of the principle makes them to send some

teachers to attend the meeting in order to make the test collaboratively.

In doing so, the writers took the sample of even mid semester test at State Junior High School 12 Padang. The test was held on March 11th 2017. This mid semester test was provided by the principles' association in Padang district or what we say as the MKKS. The way to make this test is by sending one of the English teachers to attend the meeting and collaboratively designing and making this mid semester test. This one is also ruled in semester test as this school is a state school. This has been a law of state school to provide the mid semester and semester school. From the meeting of teacher who come from several schools resulted the test for the next is used to test the students. The making of this test also is based on the syllabus that is provided by the province government in which also used by the teachers. Moreover, after finishing the distribution of test is done with giving the test sheet to the students so that they can do the examination.

Thus, it can be inferred that the developing mid-semester test in this school is based on the principles' association, where each of them will attend the meeting to make the test together. This design will after that use by those schools that send the teacher and also others that the teachers do not attend the meeting. Therefore this program also called as association of principles as teacher is the representation in order to make the mid semester test.

Based on circumstances above, mid-term test is made by teachers who attend the principle's association. Some condition may appear because of this. It is better for schools to make their own id-term test rather to use the association test. This is because every school has got different situation and condition. One teacher, who designs the test, he/she may not know other school's condition. Therefore, the test should be design by the teacher who teaches in the school, or at least the representation of the school. Teachers who know about school condition's and students' general ability.

The idea of principle's association is good. But, rather to have too general overview on test's design, it will be better to make it more specific. One solution can do to make this mid-semester test better is by gathering school's teachers to sit together designing the midterm test. The teachers' themselves who are teaching and designing the test for students. Although many theories such as what has been asserted above have provided by the experts, the implementation of the test can be far away and can be appropriate with. There are six points that are going to be provided in the following as the comparison of theory and practice within the test.

Dealing with the purpose the test is stated that it is used to make the students is able to pass the national examination. The test and also daily examination are intended to make the students become familiar with the test so they will not feel difficult to answer the questions within the national examination. According to [9], the purpose of the test is dealing with the usefulness of the test itself. It is accordingly to the profits of the test. The statement about the test seems to

---

[9] Bachman, "Modern Language Testing at the Turn of the Century"; Hughes, *Testing for Language Teachers*; Weir, "Language Testing and Validation."

be right since the national examination is occasionally becomes the top target of the learning. The students must pass the test because the big consequences of the failure of this national examination. Additionally, the pictures within the test sheet, for some numbers, are not clearly printed. The validity of the test is minus here. For example is question number 1 is asking the students to guess what are the people doing in the picture. However, the picture is blur and unable to see clearly.

The next point that can be seen is the devising the test items. This mid semester test consists of 50 numbers of questions in which all of them are in form of multiple choices. There is no essay, fill in the blanks or other kinds of item within the test. This fact shows that the test is not varieties in form of questions. It is stated that a good test carries a variation within the test. To see students' ability of language, the test needs to include other aspects of language. In this case, the test maker may also provide the test with giving other kind of questions such as the essay, fill in the blank, true false questions, or even the matching.

It is expected that by giving these variations, it can show the students comprehension more. By only giving one type of questions, it may not explore the students' ability of language. Even more in language test, it is important to dig about the students' ability by also giving variation within the test so that if it is found that the students are having good ability in one side of the test meanwhile the students is not well in other kind of test.

5. CONCLUSION

Having some points above, it can be seen that the designing and making the test need to improve. Giving more variations in questions may become the alternatives for the test maker to test the test in order to know more about their comprehension of the test. The additional kind of questions such as the essay, fill in the blank, and other kind of questions can be provided so that the variation of the test may arise the students' competence on language. Some points are related to review presents also need to attend by test designers. Test is not a simple thing. But rather it is complicated. A test cannot be regarded through only arranging some points of questions based on some topics learnt. However, more than that, a test is also suitable for students by their proficiency. To sum up, this mid semester test need to be improved and the making and designing of the test should be giving to the schools' teachers and committee since all of them are knowing precisely the students' ability and comprehension. They may also design the test by giving special items to dig more about their students' comprehension about each subject matter.

6. REFFERENCES

Algarabel, Salvador, and Carmen Dasí. "The Definition of Achievement and the Construction of Tests for Its Measurement: A Review of the Main Trends." *Psicologica* 22, no. 1 (2001).

Bachman, Lyle F. "Modern Language Testing at the Turn of the Century: Assuring That What

We Count Counts." *Language Testing* 17, no. 1 (2000): 1–42.

Bachman, Lyle F., and Adrian S. Palmer. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Vol. 1. Oxford University Press, 1996.

Black, Paul, and Dylan Wiliam. *Inside the Black Box: Raising Standards through Classroom Assessment*. Granada Learning, 2005.

Bygate, Martin, Peter Skehan, and Merrill Swain. *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing*. Routledge, 2013.

Cheng, Liying. "The Key to Success: English Language Testing in China." *Language Testing* 25, no. 1 (2008): 15–37.

Cheng, Liying, and Andy Curtis. "Washback or Backwash: A Review of the Impact of Testing on Teaching and Learning." *Washback in Language Testing: Research Contexts and Methods*, 2004, 3–17.

Dörnyei, Zoltán. *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Routledge, 2014.

Dörnyei, Zoltán, and Tatsuya Taguchi. *Questionnaires in Second Language Research: Construction, Administration, and Processing*. Routledge, 2009.

Fishman, Barry J., Ronald W. Marx, Stephen Best, and Revital T. Tal. "Linking Teacher and Student Learning to Improve Professional Development in Systemic Reform." *Teaching and Teacher Education* 19, no. 6 (2003): 643–658.

Frederiksen, John R., and Allan Collins. "A Systems Approach to Educational Testing." *Educational Researcher* 18, no. 9 (1989): 27–32.

Hughes, Arthur. *Testing for Language Teachers*. Ernst Klett Sprachen, 2007.

Purpura, James E. "Investigating the Effects of Strategy Use and Second Language Test Performance with High-and Low-Ability Test Takers: A Structural Equation Modelling Approach." *Language Testing* 15, no. 3 (1998): 333–379.

Ravitch, Diane. "A Brief History of Testing and Accountability." *Hoover Digest* 4 (2002): 35–39.

———. *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. Basic Books, 2016.

Solano-Flores, Guillermo, and Elise Trumbull. "Examining Language in Context: The Need for New Research and Practice Paradigms in the Testing of English-Language Learners." *Educational Researcher* 32, no. 2 (2003): 3–13.

Wall, Dianne. "The Impact of High-Stakes Testing on Teaching and Learning: Can This Be Predicted or Controlled?" *System* 28, no. 4 (2000): 499–509.

———. "The Impact of High-Stakes Testing on Teaching and Learning: Can This Be Predicted or Controlled?" *System* 28, no. 4 (2000): 499–509.

Weir, Cyril J. "Language Testing and Validation." *Hampshire: Palgrave McMillan*, 2005.