PRINCIPLE OF LANGUAGE ASSESSMENT OF REVIEW SEMESTER TEST IN INTERLANGUAGE: ENGLISH FOR SENIOR HIGH SCHOOL STUDENTS

LILIA INDRIANI

lilia_indriani@yahoo.com.au

Abstrak

Introduction

Test has powerful effect in the language teaching and learning process. It is important for teachers to use a test in order to measure whether the objectives of language teaching and learning have been achieved or not. Hughes (2003:5) defines testing as the way in which information about people’s language ability can be gathered. The information about people’s ability is very useful and necessary; therefore it needs a test to measure it. Furthermore, Brown (2004:3) defines test as a method of measuring a person’s ability, knowledge or performance in a given domain. Meanwhile, Airasia (2008:9) defines test is a formal, systematic procedure for gathering information. Based on these definitions, the writer can say that a test is an instrument that used by teacher to measure the students’ learning ability. A test is used as the way to know whether the students have achieved the learning objectives or not. It is proof that, at the end of the lesson there must be a test to measure the learning objectives.

Review tests are conducted for one or more of purposes such as to establish a basis for assigning grades, to determine how well each student has achieved the course objectives, to diagnose student problems for remediation and to determine where instruction needs improvement. The function of test is to help the teachers to know whether the learning objectives have been reached and to what extent students master the materials taught. Based on this reason, every teacher must be able to construct good items test to
his/her students. How can we say that it is a good test? Brown, (2004:4) states that a well constructed test is an instrument that provides an accurate measure of the test taker’s ability within a particular domain. Constructing a good test is a complex task involving science and arts. Developing a good test also requires planning: teacher must determine the purpose for the test, and carefully write appropriate test items to achieve that purpose. The most common complaint from students is that the test content did not reflect the material discussed in class. The biggest pitfalls is that the target students hit may not be the target for which teacher aimed; that is, the test scores may not measure what teacher wants them to do. Teacher must make sure the instructional objectives and the test items are congruent. For example, if one of the instructional objectives is for students to apply principles to new situations, then the items test should match the objectives by providing opportunities for students to exhibit that behavior. One simple method for planning a test is creating a table with the task across the top (the learning outcomes or behaviors a student should exhibit as a result of teaching, such as knowledge, understanding, application, analysis, evaluation and synthesis).

Interlanguage is one of the textbook for English subject that is published by Pusat Perbukuan Departemen Pendidikan Nasional (2008). It is used for tenth to twelfth grades of Senior High School. There are ten units and a review semester test after five units. Interlanguage: English for Senior High School Students X are consisted of ten units
with different topics and two review semester tests. They are Let Me Introduce Myself, I’m Sorry to Hear That, That sounds a Nice Idea, Can I See You at 11 a.m.?, I’m Delighted to Hear That, Unit Review Semester 1, It was the Least I could Do, What a Nice Hair Cut, I Find It Very Interesting, I Don’t Believe It!, That’s Very Kind of You, and Unit Review Semester 2. In this book there are also colorful and black-and-white pictures related to the topic, tables, diagrams, and icons to make the students more interesting to read the book. This book is also completed with bibliography, listening script, answer key, index subject, glossary, key to pronunciation, list of abbreviations.

In this research the writer is going to know:

1. Do the review semester tests in Interlanguage: English for Senior High School Students X by Joko Priyana, Virga Renitasari, Arnys Rahayu Irjayanti Published by Pusat Perbukuan Departemen Pendidikan Nasional (2008) fulfill the principle of language assessment?

2. How is the principle of language assessment reflected in the review semester tests in Interlanguage: English for Senior High School Students X by Joko Priyana, Virga Renitasari, Arnys Rahayu Irjayanti Published by Pusat Perbukuan Departemen Pendidikan Nasional (2008)?

**Principles of Language Assessment**

How good of test should be? How can we say that the test is effective? These simple questions need serious attention to be answered in order to construct good test.
Effective test items match the desired instructional outcome as directly as possible. Harmer (2007:167) says that good tests are those that do the job they are designed to do with and which convenience the people taking and marking them that they work. The tests also have positive rather than negative effect on both students and teachers. A good test is valid and has marking reliability. The tests must be practicality. Furthermore, Brown (2004:19) states that tests must require criteria for testing which covers practicality, reliability, validity, authenticity and wash back. These five criteria are discussed in the following sections.

1. Practicality

Practicality in language testing refers to administers, economy and fair. (Brown 2004:19) states that an effective test is practical. This means that test is not excessively expensive, stays within appropriate time constraints, is relatively easy to administer, and has a scoring/evaluation procedure that is specific and time-efficient.

A test that is prohibitively expensive is impractical. A test of language proficiency that takes a student five hours to complete is impractical it consumes more time and money than necessary to accomplish its objective. A test that requires individual one-on-one proctoring is impractical for a group of several hundred test-takers and only a handful of examiners. A test that takes a few minutes for a student to take and several hours for an
examiner to evaluate is impractical for most classroom situations. A test that can be scored only by computer is impractical if the test takes place a thousand miles away from the nearest computer. The value and quality of a test sometimes hinge on such nitty-gritty, practical considerations.

2. Reliability

A reliable test is consistent and dependable. If you give the same test to the same student or matched students on two different occasions, the test should yield similar results. The issue of reliability of a test may best be addressed by considering a number. Schumacher and McMillan (2001:181) says that reliability refers to the consistency of measurement, the extent to which the scores are similar over different forms of the same instrument or occasions of data collection. This means that reliability in language test refers to accuracy, consistency and stability of measurement by a test. Airasia and Russell (2008:243) state that reliability concerned with the stability and consistency of assessment. They said that the logical way to obtain information about reliability of student’s performance is to observe and score two or more performances or products of the same kind.

3. Validity

The most complex criterion of an effective test and arguably the most important principle is validity, "the extent to which inferences made from assessment results are appropriate, meaningful, and useful in terms of the
purpose of assessment. Schumacher and McMillan (2001:281) says that test validity is the extent to which inferences and uses made on the basis of an instrument are reasonable and appropriate. Validity is a judgment of appropriateness of a measure for specific inferences, decisions, consequences or uses that result from the scores that are generated. This means that validly in language testing refers to what to test for students. Airasia and Russell (2008:242) state that concerned with whether the information obtain from an assessment permits the teacher to make an appropriate decision about a student’s learning. Furthermore, they said that test instrument that irrelevant to the discussions (materials) it was intended to help make, it is invalid.

4. Authenticity

Authenticity is define as the degree of correspondence of the characteristics of a given language test task to the features of a target language task. Brown (2004:28) points that in a test, authenticity are presented in the following ways: the language in the test is as natural as possible, items are contextualized rather than isolated, topics are meaningful (relevant, Interesting) for the learner, sonic thematic organization to items is provided, such as through a story line or episode and tasks represent, or closely approximate, real-world tasks.

5. Washback

Every type of tests has washback effect. It occurs when teachers see the forms of their students are going to
have to take and then as a result, start teaching for the test. Washback generally refers to the effects the tests have on instruction in terms of how students prepare for the test (Brown, 2004:28).

Achievement Test

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 address in a curriculum within a particular time frame and are offered after a course has focus on the objectives in question. Achievement tests can also serve the diagnostic role of indicating what a student needs to continue to work on in the future, but the primary role of an achievement test is to determine whether course objectives have been met and appropriate knowledge and skills acquired by the end of a period of instruction. Achievement tests are often summative because they are administered at the end of a unit or term of study. They also play an important formative role. An effective achievement test will offer wash back about the quality of a learner's performance in subsets of the unit or course. This wash back contributes to the formative nature of such tests, (Brown 2004:47-48). The specifications for an achievement test should be determined by the objectives of the lesson, unit, or course being assessed, the relative importance (or weight) assigned to each objective, the tasks employed in classroom lessons during the unit of time, practicality issues, such as the time frame for the test and
turnaround time and the extent to which the test structure lends itself to formative wash back.

Achievement test is applied at the end of the course of study. Hughes (2003:12) explains model of achievement test. He states that achievement tests are directly related to language courses, their purposes being to establish how successful individual students, group of students, or the courses themselves have been in achieving objectives. Brown (2005:11) says that achievement test is design with very specific reference to a particular course which is directly based on course objective. This means that the purpose of the test is to achieve the learning objectives in the syllabus from all the materials taught to students and it is done at the end of the course of study. Meanwhile, Hughes (2003:13-15) mentions that there are two kinds of achievement tests namely final achievement tests and progress achievement tests. These two kinds of tests are explained in the following section:

1. Final achievement tests
   They are those administrated at the end of the study. Another term of final achievement test is summative test. Summative test is, used at the end of semester or in order to measure what has been achieved both groups and individuals (Hughes 2003:5). These tests are written and administrated by minister of education, official examining boards, or members of teaching institutions (schools). The content of these tests must be related to the courses syllabus or on the books and other materials used. The
Methodology

This research is a descriptive qualitative research since it tries to describe the language phenomenon descriptively without using statistics. The data are taken from the review semester tests in Interlanguage: English for Senior High School Students X by Joko Priyana, Virga Renitasari, Arnys Rahayu Irjayanti Published by Pusat Perbukuan Departemen Pendidikan Nasional (2008). The techniques of data collection are observing and note taking. Observing is done by reading the textbook and observing the principles of language assessment in the review semester tests. After observing and note taking, the principle of language assessment found are noted and collected. Data are then analyzed using the theory of principle of language assessment from Brown (2004: 19).

Subject of the Research

The subjects of the research are two review semester tests in Interlanguage: English for Senior High School Students X by Joko Priyana, Virga Renitasari, Arnys Rahayu Irjayanti Published by Pusat Perbukuan Departemen Pendidikan Nasional (2008). It is consisted of 30 multiple choice questions, that are divided into three parts, they are listening section (1-8), speaking section (9-11 in review test 1 and 9-13 in review test 2), and reading section that consist of three kinds of genre of text (12-16 text 1, 17-20 text 2, 21-30 text 3 in review test 1 and 14-20 text 1, 21-26 text 2, 27-30 text 3).
Analysis

In the review semester 1 test, it is consisted of 30 items. All of them are multiple choice items. It is divided into 8 items for testing listening, 3 items for testing speaking and 19 items for testing reading comprehension. In the review semester 2 test, it is consisted of 30 items. All of them are multiple choice items. It is divided into 8 items for testing listening, 5 items for testing speaking and 17 items for testing reading comprehension. The principle of language assessment found are:

1. Practicality

This review semester 1 and 2 test is practical because:

a. It is not excessively expensive because 30 items only in 8 pages.
b. It stays within appropriate time constraints. Although there isn’t time limitation in the test but usually it is done in English hour that is 2 x 35 minutes for one meeting.
c. It is relatively to administer because it is multiple choice items. There is also the answer key for the administrator/ the teacher.
d. It has a scoring/ evaluation procedure that is specific and time efficient. The administrator/ the teacher only count the correct items then divide it with 3. It is so easy and time efficient to give scoring for this test.

2. Reliability

It is a reliable test because it is consistent and dependable. The issue of reliability of a test may best be addressed by
considering a number of factors that may contribute to the unreliability of a test, they are:

a. Student-related reliability: temporary illness, fatigue, a 'bad day', anxiety and other physical and psychological factors.

b. Rater Reliability: human error, subjectivity, and bias may enter into the scoring process.

c. Test administration reliability: the conditions of the test administration such as in photocopying variations, the amount of light in different parts of the room, variations in temperature, and even in the condition of desks and chairs.

d. Test reliability: the nature of the test itself.

3. Validity

Validity is the extent to which inferences made from assessment results are appropriate, meaningful and useful in terms of the purpose of the assessment. Validity is the extent to which a test measures what it claims to measure. It is vital for a test to be valid in order for the results to be accurately applied and interpreted. The test in review semester 2 is about listening, speaking and reading (narrative, descriptive and news item). I can say the test is valid because it covers three skills and more than one genre based on the syllabi for semester 2. The text for descriptive entitle 'My Very Best Friend' is too long. It makes the students feel depressed. The question is also not balance with the length of the text, just 5 questions.
4. Authenticity
In the test the authenticity may be presented in the following ways:
a. The language in the test is not as natural as possible. It is too formal.
b. Items are isolated.
c. Topics are meaningful, relevant and interesting for the learners.
d. There are some thematic organizations provided.
e. The items represent real-world tasks.

Discussion

1. Instruction
In the review semester 1 and 2 tests, the instructions that are used are:

   I. Listen to short conversations between two people. Choose the best answer to each question. The listening script is in the Appendix.

The instruction is clear for the students. It doesn’t necessary mention “the listening script is in the appendix”. It will make the students open the appendix then listen to the cassette.

   II. Choose the best expressions to complete each of the following dialogues.

The instruction is clear for the students.
There is no instruction for testing speaking in review semester test 2. It is better to give instructions to for the speaking test. Just saying:

   II. Choose the best expressions to complete each of
the following dialogues.
There is no instruction for testing reading in review semester 1 test. Just saying:

Questions 12-16 are based on the following text.

It is better to give instructions to for the reading test. Meanwhile in Review semester 2 test the instruction is clear for the students.

III. Read the texts and choose the best answer to each question.

Questions 14-20 are based on the following text.

2. Listening Section
In the review semester 1 test there are 8 items for testing listening. Number 2, 3, 5, 6 and 8 are not a good test because there is more than one answer so it makes bias. The students will be confused which is the correct answer. It is not good distracters because the distracters can be the right answer too.

Here are the discussions:
2. What is Virga's intention?
   a. to introduce Retno
   b. to introduce herself
   c. to ask how Retno is
   d. to state how she is
   e. to greet Retno

The key answer is B 'to introduce her self'. But the answer E 'to greet Retno' is also correct.

See the audio script:
   Retno: Hi, my name is Retno.
Vigra: Hi, Retno. I'm Vigra

From the audio script, we can see that Vigra also greets Retno with 'Hi, Retno', so the answer E is also correct.

3. What does Arnys mean?
   a. She says that she is pleased.
   b. She wants to borrow the novel.
   c. She will lend Ruben the novel.
   d. She responds to Ruben's thanks.
   e. She would like to thank Ruben for the novel.

The key answer is A ‘She says that she is pleased.’ But it is false. The correct answer is D ‘She responds to Ruben’s thanks.’ See the audio script:

Ruben: Thank you very much for lending me the novel.
Arnys: Don’t mention it.

From the audio script, we know that Arnys says, ‘Don’t mention it’. It means that she responds to Ruben’s thanks.

5. What does Ayu mean?
   a. She wants to make chocolate pudding with Retno.
   b. She does not want to make a chocolate pudding.
   c. She wants Retno to make her a chocolate pudding.
   d. She accepts Retno’s invitation.
   e. She declines Retno’s invitation.

The key answer is D ‘She accepts Retno’s invitation.’ But answer A ‘She wants to make chocolate pudding with Retno.’ is also correct. See the audio script below:

Retno: Tomorrow I’m going to make a chocolate pudding. Would you like to come?
Ayu: Thank you, I’d like to.
From the script, we know that Ayu accepts the invitation and it means that she wants to make the chocolate pudding with Retno too. So the distracter is not good.

6. What does Arnys imply?
   a. She accepts Adib’s invitation.
   b. She declines Adib’s invitation.
   c. She wants to have dinner at another place.
   d. She wants to have dinner with somebody else.
   e. She asks Adib to have dinner at another restaurant.
   
   The key answer is B ‘She declines Adib’s invitation.’ But answer D ‘She wants to have dinner with somebody else.’ Is also correct. See the audio script below:
   
   Adib : Shall we have dinner at the Japanese restaurant tonight?
   Arnys : I’d like to, but I have another plan.
   
   It is not clear what is Arnys’s plan, so the answer D is also acceptable. It is not a good distracters.

8. What does Retno imply?
   a. She really wants to win the poetry reading competition.
   b. She wants to express that she is pleased.
   c. She wants to join the poetry reading competition.
   d. She wants to thank Marcell.
   e. She wants to tell Marcell that she won the competition.
   
   The key answer is A ‘She really wants to win the poetry reading competition.’ But the correct answer is B ‘She wants to express that she is pleased.’. See the audio script below:
   
   Marcell : Hey Retno. They told me that you won thr poetry reading competition.
Retno : Thank goodness!
From the script we know that Retno wants to express that she is pleased.

The audio script is not a good conversation for testing listening because the validity is very low. The reasons are: the conversation does not happen naturally, we don't know is Anita is the first speaker or Retno is. It is very difficult to differentiate which one is Anita or Retno. It is better to use the first or the second speaker in the questions.

In review semester 2 test there are 8 items in testing listening. All the key answer is correct. Although the script sounds not naturally and too formal for speaking, it is better than unit review semester 1. All the distracters are good.

3. Speaking Section

In review semester 2 test there are 3 items for speaking, 9, 10, and 11. The question and the distracters in 9 and 10 are not valid because they lead to ambiguity. The conversation in the questions is too formal. It is rare to happen in the real world.

9. Ruben : Can I see you at 9 a.m. tomorrow?
Ayu : _____. I have a class at 9 a.m.
   a. Sorry, I don't think so.
   b. Sure, that will be fine.
   c. I'll be waiting for you.
   d. Yes, I think so.
   e. Great. It's a perfect time.
The key answer is B ‘Sure, that will be fine’. The appropriate answer is A ‘Sorry, I don’t think so’ because from the conversation Ayu says that she has a class at 9 a.m. It means she can’t meet Ruben tomorrow at 9 a.m. But the question is not valid because Ayu can meet Ruben in the class at 9 a.m. because they are class mate. If the second works, it means the correct answer will be B, C, and D, although answer E is also acceptable.

10. Arnys : Marcell promised to take me to the boutique, but he did not come. He didn’t tell me why he couldn’t make it.

Adib : _____
   a. How upsetting.
   b. How marvelous.
   c. That sounds nice.
   d. That’s great.
   e. That’s fantastic.

The key answer is D ‘That’s great’. The correct answer is A ‘How upsetting’ because we can see from Arnys’ statement that she is disappointed because Marcell broke his promised. The question is not valid too.

Meanwhile in the review semester 2 test there are 5 items for testing speaking. It is added 2 items than in review semester 1. All the key answer is correct. The distracters are good ones.

4. Reading Section
In the review semester 1 test the question for reading start
from 12 until 30, the percentage is not balance for each of the skills. The reading test is divided based on the genre. There are recount (12-16), procedure (17-20), and narrative (20-30).

**RECOUNT**

The text for the recount text is a good one because it is authentic text that is adapted from L.A. Hill, 1963. There are 5 items. Below is the discussion:

12. The text tells about ...
   - a. joining a football match
   - b. winning a football match
   - c. attending a football match
   - d. the school football team
   - e. playing football

This question leads to main idea of the text. The answer key is C ‘attending a football match’ but the word ‘attending’ is not appropriate the correct answer is A ‘joining a football match’. The distracters are good, although I found the different pattern in D ‘the school football team’, not using gerund.

The distracters in number 13 are not good because there are two sentences that lead to the answer that are:
   - a. They shouted to support their school team.
   - b. They shouted to the headmaster.

The word ‘shouted’ is repeated in A and B. The correct answer for number 13 is A based on the text.

**PROCEDURE**

The text is authentic that is adapted from www.cest-mamaison.net. The distracters are good. The key answer for
number 17–20 are correct.

**NARRATIVE**
The text is authentic that is adapted from www.st.rim.or.jp. The distracters are good. The key answer for number 21–30 are correct, except number 23.

23. How was Queen Bidara Putih like?
   a. beautiful, sensible, and sensitive.
   b. beautiful, uncaring, and sensitive.
   c. pretty, wise, and indifferent.
   d. pretty, sensible, and indifferent.
   e. beautiful, wise and unconcerned.

None of them are correct based on the text, the correct one is ‘beautiful, wise and sensitive.’

Meanwhile in review semester 2 test there are 13 items in reading. The texts are narrative, descriptive and news item. I will discuss one by one because there are a lot of mistakes in reading text. Question 14-20: ‘The White Butterfly’. The text is authentic because it is adapted from www.pitt.edu.

15. How was Takahama’s characters?
   a. He was a humorous person.
   b. He was an indolent person.
   c. He was a pathetic person.
   d. He was a sociable person.
   e. He was a hilarious person.

The key answer is D ‘He was a sociable person’ but in the text there is no sociable appeared there. The correct one is ‘He was a loyal person’. Loyal is different with sociable.
18. The word 'it' (paragraph 3, sentence 2) refers to ...
   a. the garden
   b. the cemetery
   c. the tomb
   d. the white butterfly
   e. the gate
The key answer is C ‘the tomb’, but the correct answer is D ‘the white butterfly’.

19. Which of the following words is the antonym of 'resolved'?
   a. undecided
   b. resolute
   c. unyielding
   d. determined
   e. unbendable
The distracters are not good because the word resolved in the form of past participle so all the distracters must be in the same form.

Question 21-26 : ‘My Very Best Friends. The last paragraph doesn’t sound English : ‘That’s all about me and friends of mine.’ It must be ‘That’s all about my friends and me’.

22. The following are Firsty’s characters, except ...
   a. cheerful, humorous and sensitive
   b. cheerful, confident and strong-minded
   c. determined, sensitive and funny
   d. determined, humorous, sensitive
   e. sensitive, confident, serious
The key answer is B ‘cheerful, confident and strong-minded’
but it should be A, C, D, and E. We should omit the word 'except' if the answer is B. The usage of 'and' in the options D and E are more appropriate.

Question 27-30: ‘Inul aims to drill after haj’. The title should use title case. There’s no problem in the question, key answer and distracters.

Conclusion

The review semester tests in Interlanguage: English for Senior High School Students X by Joko Priyana, Virga Renitasari, Arnys Rahayu Irjayanti Published by Pusat Perbukuan Departemen Pendidikan Nasional (2008) do not fulfill all the principle of language assessment. There are still many mistakes in validity such as the instructions are not clear. The test is less authentic because some of the items in the test do not use natural language. Then the items are isolated rather than contextualized. Next, there are no thematic organizations provided such as a story line or episode. The question and the distracters are not valid because they lead to ambiguity. The conversation in the questions is too formal. It is rare to happen in the real world. The test also has the wrong key answer. It is suggested for the writer, the editor and the publisher to check all the principle of language assessment before they print it in a text book.

References

Singapore.