



International Journal of Humanities & Social Science Studies (IJHSSS)
A Peer-Reviewed Bi-monthly Bi-lingual Research Journal
ISSN: 2349-6959 (Online), ISSN: 2349-6711 (Print)
Volume-II, Issue-V, March 2016, Page No. 143-156
Published by Scholar Publications, Karimganj, Assam, India, 788711
Website: <http://www.ijhsss.com>

Improving EFL Achievement Tests at High Schools

Le Hai Duy

MA, Dong Thap University, Vietnam

Abstract

The paper addresses the investigation of EFL achievement tests used at Vietnamese high schools. The collected data were 3 in-class EFL tests from 3 high schools of Dong Thap Province, South of Vietnam. In addition, a questionnaire was completed by students of these 3 selected high schools. Then, based on the statistical analyses of the tests and questionnaire collected, the good and can-improved points are presented. It is thereby suggested that teachers pay more attention to the rules of designing tests, especially writing multiple-choice questions to improve test quality.

Key words: content, test, testing, reliability, validity.

Introduction: Heaton (1975) mentioned the relationship between testing and teaching. Both of them are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with the other. Tests can be used as devices to assess students' learning as well as the students' performance in the language classroom. Accordingly, teachers can make necessary changes in the way of teaching and what they teach. Similarly, In addition, tests have become a way of life in the educational world and tests are often used for pedagogical purposes, either as a means of motivating students to study, or as a means of reviewing material taught (Bachman, 1990).

Testing is one of the most important tools for both students and teachers to monitor the progress of teaching as well as learning. EFL at school is not an exception in this vein. Henning (1987, p.1) argued that the main functions of tests are to diagnose and provide feedback as "*Perhaps the most common use of language tests and educational tests in general is to pinpoint strengths and weaknesses in the learned ability of the students*". Moreover, the value of the tests can be said to "*provide critical information to the students, teachers, administrators that should make the learning process more efficient*" (p.2). Thus, how well do the students achieve EFL lessons? How effective our EFL teaching methods are? Tests are supposed to give us valid answers. However, there are still some problems in many EFL tests in high schools of present Vietnamese context, especially in the paper-based tests. It is therefore likely to result in that teachers can't assess their students properly with such kind of tests.

The present study has been designed to find out common issues in EFL tests for grade-10 schoolers and give recommendations to improve test quality. Accordingly, it seeks answers to the following questions: (1) *What are current EFL tests at high school like?* (2) *What are students' views on the tests?* (3) *What can improve test quality?*

Literature Review: *Testing* or administering a test is generally a method of measuring a person's ability or knowledge in a given area. It is a set of techniques, procedures and items that constitute an instrument of some sort that requires performance or activity on the part of the test-taker (and

sometimes on the part of the tester as well). Meanwhile, a *test* is an instrument or procedure designed to elicit performance from learners with the purpose of measuring their attainment of specified criteria. The method may be intuitive and informal or may be structured and explicit (Brown, 2001). Accordingly, language testing is the administration of tests in order to assess and measure a person's language competence and performance or testing language ability. It is an evaluation of an individual's language proficiency (Bachman, 1990).

Purposely, the tests are designed (Hughes, 1989): to (1) measure language proficiency regardless of any language courses that candidates may have followed; (2) discover how far students have achieved the objectives of a course of learning/study; (3) diagnose their students' strength and weakness, what they know and what they don't; (4) assist placement of students by identifying the stage or part of a teaching program that is suitable for their ability.

According to Heaton (1975, p.11-16), tests can be categorized according to the type of information they provide. They include: *achievement test* (to show mastery of a particular syllabus or language course), *proficiency test* (to measure people's ability in a language regardless of any training they may have had in that language), *diagnostic test* (to assess students' knowledge and skills in particular areas before a course of study is begun), and placement tests (to sort new students into teaching groups so that they are approximately at the same level as others when they start).

The most important quality of a test is its *validity*, or the extent to which the inferences or decisions we make on the basis of test scores are meaningful, appropriate, and useful. For a test score to be a meaningful indicator of a particular individual's ability, we must be sure it measures that ability and very little else (Bachman, 1990). Meanwhile, *reliability* is a quality of test scores, and a perfectly reliable score would be one free from errors of measurement. There are many factors other than the ability being measured that can affect performance on tests and that constitute sources of measurement error. Individual's performance may be affected by differences in testing conditions, fatigue, and anxiety, and they may thus obtain scores that are inconsistent from one occasion to the next. If, for example, a student receives a low score on a test one day and a high score on the same test does not yield consistent results, and the scores can't be considered reliable indicators of the individual's ability (Bachman, 1990). Reliability thus has to do with the consistency of measures across different times, test forms, raters, and other characteristics of the measurement context.

Reliability and validity are the two most vital attributes that constitute a good test. Both together make a good test. The absence of either would result in a poor test. On the one hand, it is possible for a test to be reliable without being valid. It means that a test can give the same result time after time but not measure what it is intended to measure. For example, a Multiple-choice questions test could be highly reliable in the sense of testing individual vocabulary, but it would not be valid if it were taken to indicate the students' ability to use the words productively (i.e. in speaking skills). As Bachman (1990, p.25) says, "While reliability is a quality of test scores themselves, validity is a quality of test interpretation and use". On the other hand, if the test is not reliable, it cannot be valid at all. To be valid, "a test must provide consistently accurate measurements. It must therefore be reliable. A reliable test, however, may not be valid at all" (Hughes, 1989, p.42). For example, in a writing test, candidates may be required to translate a text of 500 words into their native language. This could well be a reliable test but it cannot be a valid test of writing.

Over the past decades, there have been a great number of studies on the tests design and test evaluation. The procedure and qualification of a test were mentioned in some of the studies by Nguyen (2006) in which he restated some theoretical background related to testing. Thus, the process of test analysis includes reliability, validity and item discrimination. It also suggests some useful considerations to improve the quality of the future tests.

Iseni (2011) looked at testing and evaluation in another aspect: *How the teachers deal with errors and mistakes collected from their students' tests*. The study is the result of continuous four-year work in the field of noting and reviewing of errors in learning English. In this work, the errors made by students during tests, interviews, dictations, writing paragraphs and different essays were collected and studied. The findings suggested that if the spoken language errors can be allowed without being corrected, as long as the message comes clear, it is understood that in the written work errors should be corrected more carefully, because if they are left without correction, these errors can become fossilized.

Lee and Winker (2013) conducted a study on the differences among three-, four-, and five-option-item formats in the context of a high-stakes English-language listening test. They found that the mean test scores of the three-option tests were significantly higher than those of four- and five-option tests. While no difference was found in mean item discriminations across the three different test formats, reliability coefficients showed inconsistent patterns depending on the number of options and test versions. Then, Lee suggested possible explanations for the results. One possible interpretation of the low correlations among the scores of three test formats is that items with different numbers of options tap into skills other than listening. The findings suggest that statistically, three options may or may not be optimal depending on the point of view taken – from that of the test score users, or from that of the test stakeholders. In addition, test developers must consider multiple statistical, affective, and contextual factors in determining the optimal number of options.

Thanyapa and Currie (2014) employed a study to closely examine the impacts of number of options in a multiple choice test. In the study, they gave stem-equivalent short answer and multiple choice tests of English structure and reading to students at Prince of Songkla University, Thailand. After that, they made a comparison of scores, facility, discrimination, reliability and validity from 3-, 4- and 5-option versions of the multiple choices test. The study results suggested that there was little variation in the performance of the three option formats although higher numbers of options tended to increase difficulty and to spread responses between the options. The researchers concluded that option number format is of little importance in the measurement of language ability compared with the much greater construct variation apparently induced by the use of multiple choice items as against constructed response formats.

In Pakistan, Sarwar et al. (2014) conducted a study about how assessments affect the learning style and focus of students in Pakistan. They assessed English speaking skills of prospective teachers at entry and graduation level in teacher education program. The results of this study reveal that there is no improvement in the English speaking skills after one complete academic year, despite the fact that the medium of instruction and assessment is English. The main reason was stated: Firstly, any language can be decomposed into four basic skills: reading, writing, listening and speaking. However, because in actual practice, only written English is assessed in Pakistan, students divert their effort towards writing skills only. So, the students get prepared for written examination according to the requirements of the assessment. They do not pay any heed to spoken skills which

are not assessed. After all, they concluded that assessment has much more influence on students' learning behavior as compared to that of the teaching. This is a reminder to the Pakistani universities where students pass the examinations without any improvement in oral skills. The pressure to maximize examination scores shifts the emphasis on those tasks only which are assessed.

However, in the past few decades, only a limited number of studies were done related to EFL testing at high schools. In Dong Thap Province, no research has been found related to the investigation into the reliability and validity of grade 10 EFL tests at high schools. As a result, the present study hopes to make contributions to EFL test improvements.

Methodologies

Research Participants: The participants of this study were 210 students of grade 10 from 3 selected high schools in Dong Thap Province (one from a remote area, one from the urban and the other from sub-urban). The students were randomly chosen from the total of all students at these 3 high schools in Dong Thap Province. Most of them come from the districts and communes of the local province and all of them have learned English (as a foreign language and compulsory subject at lower secondary school) for at least 4 years.

Data Collection Instruments

(1) Achievement Tests Collected: Three achievement tests were collected in the second term of school year 2013-2014. The tests are coded A, B, C in which

- + Test A: a test from a high school in remote area (see Appendix),
- + Test B: refers to a test from a high school in urban area,
- + Test C: refers to a test from a high school in the sub-urban area.

The tests are coded to keep the confidential data about the schools participated in the study. During the study the participants were told that the data collected in the study would be used for reference only and all of the information, materials related were kept confidentially. Only the researcher himself can reach the raw materials collected in the study. All the three collected tests have 34 items with 5 to 6 separate parts.

(2) Questionnaires: In the present study, one questionnaire was employed. It was given to the students to collect information about:

- (i) How they feel about the test or the motivation of the tests to their study
- (ii) Their self-assessment on the test they have taken
- (iii) What they think about the ways they are tested.

The questionnaire is in Vietnamese so that all the students can understand it clearly. Using the students' first language means that the information participants receive has more chance of being accurate and complete. A total of 210 copies of the questionnaire for students were given at the 3 selected schools.

Data Collection Procedure:

- The researcher obtained the copies of the 3 relevant marked tests from 3 selected schools.
- The matrix of the tests was also collected during the design process of the tests. After that, the content of each test was compared with the matrix and syllabus to investigate the content and construct validity.
- After collecting the tests, a questionnaire was distributed to students. They were asked to complete the questionnaire at home and the questionnaire was collected the next day.
- After that, the software SPSS. Version 20 was used to analyze the data.

Results and Discussions

Test Findings

(1) Good points

- Most parts of the tests are employed in the multiple choice questions. With the choice of multiple choice formats, the tests are therefore said to somehow have the following strengths:

+ Ability to measure different thinking abilities with the integration of various answers (multiple-choice) to choose for each question.

+ The tests have the wide range of multiple choice options that cover much of the content of the whole course.

+ We can also analyze the properties of each item easily for further purposes. With items analysis, we can easily identify which question is too easy, too difficult, and ambiguous or even not correlated to the aims of the test. Then with the results from the analysis, teachers can improve the quality of the tests.

+ The convenience and practicality of the marking procedure: The tests can be marked by computer with appropriate program which can save time of the teachers.

- Also, all of the tests collected are well-typed with clear and easy-to-read format. Most of the parts in the tests have rubrics (in Vietnamese or English) which provide clear instructions for the tasks in the tests.

- The number of items and the allotted time seem to be appropriate to most of the students. The length of the tests is not too much which may make students tired. In addition, they are not too short so that it may hard to evaluate students' performance.

(2) Can-improve Points

(i) Testing two different types of knowledge at the same time (Test A)

For example:

16. *Gold near San Francisco in 1848, and the gold rush started the following year.*

A. discovered	B. was discovered
C. has been discovered	D. is discovered.

To answer this question successfully, the students should have some knowledge about the tenses. Clearly, options C and D of this item can be eliminated because of inappropriate tenses used. Moreover, the knowledge on Voices (Passive or active) is also tested. According to Henning (1987, p.44), such kind of item may “lack validity as measures of what they are purported to measure”.

(ii) The incomparable length of options

According to Bachman (1990), teachers should construct incorrect answers that are comparable in length, complexity, and grammatical form to those of the correct answer. It is important to avoid giving clues in the construction of response alternatives. The purpose of an item is to measure what a student knows and understands and not how clever a test taker or he or she is.

In the tests, some of the options are longer in length in comparison with others.

27. Chris: Where do you work?
 Evan: In a studio
 Chris:
 A. That must be interesting
 B. What for?
 C. It is so...
 D. When?

From the above extract from the test A, it is obvious that option A is much longer than the other options in the item. This may cause extra clue for the students and maybe, the test-wise examinees can inflate score from such poorly written items like this.

Another example for this issue can be seen in test C. Item “Question 16” was designed in this way in which option B is much longer than the others.

Question 16: “I find horror films really interesting”
 “_____ . I find them very interesting”
 A. That’s right
 B. I don’t quite agree with you
 C. I agree
 D. I think so

Henning (1987, p.44) also stated “Frequently, the longest and most explicit option available is the obvious answer”. Therefore, teachers should avoid using test’s options that are too different in length.

(iii) The confidentiality of the tests: Most of the content of the tests can be easily found on the internet. One of the most popular sources for those materials is violet.vn (which is an open forum for teachers to share their tests, exercises, and experience). Although the chance for the students to find the content in prior to testing is extremely low, teachers should care more about the confidentiality of the tests. Only with good confidentiality, the tests can be said to have fair and reliable assessment to all students.

(iv) Using ambiguous questions: In the reading comprehension section of the test B, some of the questions are employed using pronoun “they”.

Question 4: Where can they be found?
 A. In rivers
 B. In the Pacific Ocean
 C. In all oceans in the world
 D. In the Arctic Ocean

appropriate. It is certain that the aims of the tests’ designers are to evaluate the text reading comprehension ability, not on how well they understand the question. Therefore, in this case, it should be clear from the question that it was used to ask information about dolphins so that students don’t need to spend their valuable time to infer what “They” is in the passage.

(v) Using answer choices that include other choices

An extract from the reading comprehension section of test B contains this problem.

Question 1: Where do they prefer to live?

- | | |
|---------------------|-------------------------------|
| A. In lakes | B. In bays |
| C. In coastal water | D. In coastal waters and bays |

In this item, it is obvious that option D is designed to consist of option B and C. According to Henning (1987), options such as “D” in the example “require the examinee to review the previous options and form conclusions about the interrelatedness or co-occurrence” and “such items may be more useful as measures of concentration and reasoning power than of reading comprehension or other purely linguistic skills”.

We also find the problem in item 2 (Test A):

Question 2: What do traffic policemen have to do, in one well-known country?

- | | |
|-----------------------------------|------------------------------------|
| A. They have to pollute the air | B. They have to cut down the trees |
| C. They have to wear oxygen masks | D. A and C are correct |

Therefore, such kind of options should be eliminated from not only reading comprehension but also other skills tests.

(vi) Testing common knowledge (Test B)

Question 2: Dolphins are _____.

- | | |
|------------------------------|------------|
| A. insects | B. fish |
| C. the animals which lay egg | D. mammals |

In this reading comprehension section, the aim of the test is to evaluate how well the students comprehend the reading passages. However, it is obvious that if students have knowledge on what kind of creature dolphins are, the correct response may be chosen without comprehension of the reading passage.

Test Descriptive Statistics

- Test A

First of all, the mean score of the test gained by the middle testees in the order of merits is **5.58**. Also, the mode of this test is **6.5**. It refers to the list of numbers that occur most frequently. Thus, in this test, **six students** have **the grade of 6.5**. In addition, we can see the range of this test is **5.0**. This term refers to the difference between the highest and the lowest scores. In this case, the highest score is **8.0** and the lowest is **3.0**.

Another important index in this test is **Standard deviation (Std.Dev) = 1.427**. In fact, the higher the value of Std.Dev, the better the test is discriminating among student performance levels. In this case we have a rather large standard deviation (**1.427**), so the test can be inferred to have the following characteristics:

- + The score distribution of the test was wide.
- + The test has spread the students out.
- + There was a wide range of ability among the students

From the final reliability statistics of 34 items, the average reliability coefficient of the test is:

Reliability Statistics

Cronbach's Alpha	N of Items
.670	34

The obtained reliability index of this test is **0.670**. It means that nearly **67%** of the test items are reliable and only **33%** may lack reliability. As such, this value indicates acceptable reliability of the test. This test could have been even better reliability if those items of low reliability were modified or removed. Some of the items that should be improved include “Q5” (Question 5), “Q28” which have poor correlation with the others. By improving those questions, the reliability of this test can be improved up to 70.9 percent. If item “Q5” is removed from the test, the reliability of the tests can be up to **6.99**. Then, the reliability can be increased to **7.09** when we delete/modify item “Q28”.

- Test B

From the test’s analysis, it is clear that the final result of the test is really low while the mean of this test is only 4.3. Also, the diagram shows that there is a huge dispersion between the best student and the less able ones. In this case, the range of the highest and lowest score is 5.75. This indicates great distance between the best student in class and the less able ones.

Reliability Statistics

Cronbach's Alpha	N of Items
.715	34

As the result of the reliability statistics shown above, the obtained reliability index of this test is **0.715**. It means that nearly **72%** of the test items are reliable and only **18%** may lack reliability. As such, this value indicates high reliability of the test. Although the estimated reliability of this test is already high, the reliability can be higher with some changes in the test’s items. Some of the items that should be improved include “Q11”, “Q18”. By improving those questions, the reliability of this test can be improved up to nearly 76% percent.

- + By modifying item “Q11”, teacher can increase the reliability of this test to 7.43
- + The reliability can be up to 7.58 if item “Q18” is modified or removed.

- Test C

First of all, the mean score of the test gained by the middle testees in the order of merits is **5.98**. Remarkably, there are several modes for this test including **5, 6, and 7**. It refers to the list of numbers that occur most frequently. Thus, in this test, four students have the **grades of 5, 6, and 7**. In addition, we can see the range of this test is **7.0**. This term refers to the difference between the highest and the lowest scores. In this case, the highest score is 9.0 and the lowest is 2.0. This indicates a great gap between the good and less able students in the class.

Reliability Statistics

Cronbach's Alpha	N of Items
.863	34

The obtained reliability index of this test is **0.863**. It means that over **86%** of the test items are reliable and only 14% may lack of reliability. As such, this value indicates high reliability of the

test. Because the reliability of this test is already high, there is no need to change or modify the test's items except for those that have poor correlation with other items.

Validity

Although the tests have rather high reliability, some of the tests' items still lack of validity. Firstly, *the content validity* of the tests seems to be satisfied when most of the contents of the tests are consistent with learning objectives and syllabus. Thus, for *Writing section*: The "Sentence Transformation" and "Sentence Building" cannot be considered as real writing since they do not reflect the real tasks in practice.

Secondly, the construct validity of the test should be taken into consideration. In most parts of the tests, the difficulty of each of them seems to be inconsistent. In the test, basing on the items difficulty, the most difficult items are mostly in the writing section while very few students could answer those questions successfully. It should be taken into consideration the reasons why those items are too difficult with students. Whether the structure in those items are too strange to the students or they don't know how to deal with such kinds of questions, that difficulty affect directly to the validity of the whole test. On the other hand, the tests are defined to have a "communicative competent" construct. However, most parts of the tests are in multiple choice forms, and there are few communicative skills tested.

In general, the 45-minute progress achievements tests collected at some high schools in Dong Thap seem to have adequate reliability with acceptable reliability coefficient (0.67) and even high reliability (0.71 and 0.86). However, they have some low difficulty items and poor discrimination ones. This may affect the validity of the tests.

Questionnaire Findings: To enhance our understanding on the students' motivation, and attitudes over the tests they have taken, a 10-item questionnaire was employed. In the study, 210 questionnaires were given to the students in different classes at the three selected high school in Dong Thap Province. The students were allowed to take the questionnaire home and returned them the following day. However, at the end, only 198 questionnaires were returned and used for analysis.

After calculating, the results of the questionnaire given to students can be reported in the following table:

Table 1: Questionnaire Results

ITEMS	RESULTS							
1. How long does each test take you to finish?	Half of the time		2/3 time		All of the time		Not enough time	
	43	21.7%	27.8%		65	32.8%	37	18.7%
2. How do you feel about your tests?	Very difficult		Difficult		Normal		Easy	
	56	28.2%	130	65.6%	10	5.0%	2	1.2%
3. The average score of your test is...	Under 5		5- 7		7-8		Above 8	
	57	28.8%	88	44.4%	6	3%	47	23.8%
4. How often are you tested?	After every lesson		Once a week		Once a month		Not fixed	
	0	0%	0	0%	68	34.3%	130	65.7%
5. Does your teacher limit the part of the lesson that you'll be tested?	Usually		Often		Rarely		Never	
	23	11.6%	143	72.2%	7	3.5%	25	12.6%
6. How many parts are there in your tests?	3		4		5		More than 5	
	0	0%	54	27.3%	115	59.0%	29	14.7%
7. What part of the test is the most difficult?	Reading		Grammar		Writing		Vocabulary	
	25	12.6%	17	8.6%	107	54%	47	23.8%
8. Do you think that your tests are important?	Very important		Rather important		Normal		Useless	
	167	84.3%	15	7.5%	13	6.5%	4	1.7%
9. Do you agree with your teachers' remarks?	Usually		Often		Rarely		Never	
	178	89.9%	20	10.1%	0	0%	0	0%
10. Where do you prepare for your tests?	Textbook		Exercise book		Reference book		Other	
	47	23.8%	61	30.8%	72	36.4%	18	9%

For the first question in the questionnaire, 21.7% of the students say it took them only half of the allotted time to finish the test. This suggests that the test was either too difficult for them or too easy that the students didn't need much time to finish it. The other 27.8% just had nearly enough time to finish it while over 50% could finish the test on time (32.8%) or failed to finish all parts of the test (18.7 %).

From the questionnaire, it can be inferred that most of the students have clear understanding on the tests they took while nearly half (59%) of the students still remember how many parts there are in their test.

Another point that proves the importance of tests in schools is students' attitudes towards their tests. More than 80% of the students think that the tests are very important when asked. Some (9.6%) keep neutral stands when they say that the tests are of normal or little importance.

Also, according to the students, the most difficult part in their tests is the Writing sections. This is consistent with the results from the descriptive statistics (mentioned below) in which the items belonging to the writing section have the highest difficulty index and discrimination. This may lower the reliability of the tests. Furthermore, this may discourage the students from writing skill.

Next, because most parts of the tests are designed under the multiple-choice format, it is obvious that most of the students (nearly 90%) agreed with their teachers' remarks. The objectiveness of the multiple-choice questions makes it easier for the teachers to have a fair and precise evaluation over the performance of their students.

Finally, the results collected from question 10 suggest that the textbook is still the major source when students prepare for their tests while 28.3% of the students just focus on their textbook and the knowledge they learned at school. Some decided to focus on the exercise book. One of the most amazing findings from the questionnaire is that when students are asked where they prepare for the exam, many of them decide to work on the reference book. It can be inferred that some of the content in their tests are available in reference book or the materials designed by their teachers. The last 9% of them choose to practice the exercises from the internet and/or other sources.

Conclusion: Generally, the test-analysis results show the good points: (1) Most parts of the tests are employed in the multiple choice questions; (2) All of the tests collected are well-typed with clear and easy-to-read format; (3) Most of the parts in the tests have rubrics (in Vietnamese or English) with clear instructions for the tasks in the tests; (4) The number of items and the allotted time seem to be appropriate to most of the students.

Yet, there are some issues related to the content, and validity of the tests which need to be improved. They include: (i) Invalid content: testing two different types of knowledge at the same time; the incomparable length of options; the confidentiality of the tests; using answer choices that include other choices; testing common knowledge; (ii) Low discrimination items.

In the light of the findings, it is recommended that teachers should take care of the following advice to improve the reliability and validity of their tests:

Remember the rules of multiple-choice tests: Although it is quite excessive to mention the rules again, they are very important as most of the tests in the recent years are designed in the multiple-choice format. In fact, some criterion for a good Multiple-Choice test can be summarized as follows:

- consider the kinds of mistakes a student is likely to make and use them as a basis for writing the incorrect response choices;
- construct incorrect answer's that are comparable in length, complexity, and grammatical form to the correct answer;
- write the questions and choices in language that your students can understand;
- state the items so that there can be only one interpretation of their meaning.
- not provide extra clues to correct answer within the item statement itself (for example, a or an, capitalized letter...);

- not test more than one points in a single item, however, multiple points can be made in a single item, where appropriate, by providing choices that specifically include two or more of the choices already given;

- write options in a correct grammar, except course in the case of specific grammar test items.

Test Content: One of the most serious problem that affect the validity of the tests is the tests' content. Some of the tests items seem to be too difficult or too easy to the students. This calls for revising those items so that they can be more valid in comparison with other items in the tests. Consequently, this should be improved by choosing the sentences, structures and vocabulary wisely and carefully which are more common for students. The content of the tests can be improved as follows:

- Replacing weak items in multiple-choice and reading comprehension (the items that are too ambiguous or much difficult for students);
- Using other more common structures for the five writing items which are too difficult for the students;
- Integrating/ modifying speaking items and listening items into the tests to fulfill the requirement of communicative competent construct.

References

1. Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: OUP.
2. Brown, H. D. (2001). *Teaching by principles: an interactive approach to language pedagogy* (2nd Ed). White Plains, New York: Longman.
3. Heaton, J. B. (1975). *Writing English language tests: a practical guide for teachers of English as a second or foreign language*. London: Longman.
4. Henning, G. (1987). *A guide to language testing: development, evaluation, research*. Rowley, Massachusetts: Newbury House.
5. Hughes, A. (1989). *Testing for language teachers*. Cambridge, New York: CUP.
6. Iseni, A. (2011). Assessment, testing and correcting students' errors and mistakes. *Language Testing in Asia*, 1(3), 60-90.
7. Lee, H. & Winke, P. (2013). The differences among three-, four-, and five-option-item formats in the context of a high-stakes English-language listening test. *Language Testing*, 30, 99 -123.
8. Nguyễn Anh Tuấn. (2009). *Đo lường và đánh giá trong giáo dục*. Đại học Quốc gia Hà Nội, Vietnam.
9. Sarwar, M., Alam, M., Hussain, S., Shah, A, & Jabeen, M. (2014). Developing and evaluating a dynamic assessment of listening comprehension in an EFL context *Language Testing in Asia*, 4(8), 1-19.
10. Thanyapa, I., & Currie, M. (2014). The number of options in multiple choice items in language tests: does it make any difference? Evidence from Thailand. *Language Testing in Asia*, 4(8), 1-21.

Trường THPT

Mã Phách (HD chấm ghi)

KIỂM TRA GIỮA HỌC KỲ II
MÔN: TIẾNG ANH 10
THỜI GIẠN: 45'

Mã đề : 103

Điểm

3,5

3
27

Read the following passage and choose one correct answer A, B, C, or D for each question

We are all slowly destroying the earth. The seas and rivers are too dirty to swim in. There is so much smoke in the air that it is unhealthy to live in many of the world's cities. In one well-known city, for example, poisonous gases from cars pollute the air so much that traffic policemen have to wear oxygen masks. We have cut down so many trees that there are now vast areas of wasteland all over the world. As a result, farmers in parts of Africa cannot grow enough to eat. In certain countries in Asia there is too little rice. Moreover, we do not take enough care of the countryside. Wild animals are quickly disappearing. For instant, tigers are rare in India now because we have killed too many for them to survive. However, it isn't enough simply to talk about the problem. We must act now before it is too late to do anything about it. Join us now. Save the Earth. This is too important to ignore.

- How are the seas and rivers nowadays ?
- | | |
|--------------------------------|----------------------------------------|
| A. are contaminated | B. are less dirty than they used to be |
| C. are dirty enough to swim in | D. cannot be swum in |
 - | | |
|-----------------------------------|--------------------------------------|
| A. They have to pollute the air | B. They have to cut down many trees. |
| C. They have to wear oxygen masks | D. A and C are correct |
 - | | |
|------------------------------------------|-------------------------------------------------------------|
| A. Because many trees have been polluted | B. Because people cut down many trees |
| C. Because there is too little rice | D. Because there are large areas of land that can't be used |
 - | | |
|--------------------------------------------------------|-------------------------------------|
| A. being protected from natural environment | B. in danger from extinction. |
| C. killed so many that they can't live in the forests. | D. So rare that they can't survive. |
 - | | |
|---------------------|------------------------|
| A. Save the Earth. | B. Protect the nature. |
| C. The Environment. | D. Conservation. |

Chọn từ gạch dưới có cách phát âm khác với từ còn lại :

- | | | | |
|----------|----------|------------|----------|
| A. plays | B. stops | C. listens | D. reads |
|----------|----------|------------|----------|
- | | | | |
|-------------|---------|-----------|--------------|
| A. thriller | B. film | C. silent | D. situation |
|-------------|---------|-----------|--------------|

Tim lỗi sai trong từng câu :

- Have the students be told about the changes of their timetable yet ?
A B C D
- If Mr. Black weren't late, he wouldn't have lost his job.
A B C D
- What would happen if the temperature is warmer ?
A B C D

Chọn một từ có dấu nhấn khác với từ còn lại :

- | | | | |
|-----------|------------|------------|-----------|
| A. listen | B. control | C. receive | D. divide |
|-----------|------------|------------|-----------|
- | | | | |
|--------------|----------------|------------|------------|
| A. different | B. interesting | C. unusual | D. capital |
|--------------|----------------|------------|------------|

Chọn đáp án đúng cho từng câu :

- Folk music _____ babies to sleep.
A. makes B. lulls C. lets D. wakes
- I wonder when the world came into.....
A. exist B. existential C. existent D. existence
- Have they tested all the machines?
A. Have all the machines been tested? B. Have all the machines been testing?
C. Have all the machines be tested? D. Have all the machines been being testing?
- Gold near San Francisco in 1848, and the *gold rush* started the following year.
A. discovered B. has been discovered C. was discovered D. is discovered
- The boy waswhen he was 2 years old.

- A. wasted B. increased C. abandoned D. gone
- 18 The lifeof sperm whale can be up to seventy years.
- A. distance B. span C. length D. period
19. If I..... hungry yesterday, I would have bought something to eat.
- A. feel B. felt C. have felt D. had felt
20. If She studies hard, shethe next exam.
- A. will pass B. passed C. would pass D. passes
21. I was tired because I stayed up late last night.
- A. If I hadn't stayed up late last night, I wouldn't have been tired.
- B. I wouldn't be tired if I didn't stay up late last night.
- C. If I had stayed up late last night, I would have been tired.
- D. I hadn't been tired if I wouldn't have stayed up late last night.
22. She is a biologist. She spends all his life studying sea plants and animals.
- A. ocean B. botany C. marine D. sea
23. Peter: How often do you go to school? Harry:
- A. I go there early. B. Every day except Sunday. C. I don't think so. D. I go there by bus.
24. They are building a new zoo near my village.
- A. A new zoo is being building near my village. B. A new zoo will be built near my village.
- C. A new zoo is built near my village. D. A new zoo is being built near my village.
25. A: Do you mind if I borrow a chair?
B: Do you need only one?
- A. I'm sorry. B. Yes, I do. C. Not at all. D. Yes, I would.
26. That sounds like a good offer. I.....it if I.....you.
- A. had accepted/ were B. would accept/ were C. will accept/ am D. accepted/ were
27. Chris: Where do you work ?
- Evan: In a studio.
- Chris:
- A. When? B. What for? C. It is so... D. That must be interesting
28. He is really delighted with his new CD player.
- A. entertained B. angry C. pleased D. annoyed

1	2	3	4	5	6	7	8	9	10	11	12	13	14
C	D	B	C	A	C	D	D	C	C	D	B	D	C
15	16	17	18	19	20	21	22	23	24	25	26	27	28
A	B	A	B	D	A	A	C	A	C	D	B	D	B

Phân tư luận:

Viết lại câu theo gợi ý :

29. They are repairing the streets this month. (passive voice)
 The streets are repaired this month. 0,25
30. We didn't go to the concert last night because we were busy. (If sentences)
 If we weren't busy, we would go to the concert last night. If we had been busy, we would have gone to the concert last night.
31. They had finished the preparations by the time the guests arrived. (passive voice)
 The guests were greeted by the time they arrived.
32. I can't look the word up because I don't have a dictionary. (Conditional sentence)
 If I had a dictionary, I could look the word up.
33. She left her relatives five million pound. (passive voice)
 Five million pounds were left by her relatives.
34. He goes to school late because the traffic is heavy. (Conditional sentence)
 If the traffic is heavy, he won't go to school late.