

ANALYSIS OF ENGLISH ENTRANCE TESTS AT SENIOR HIGH SCHOOLS

By
Wildanur*
Sofyan A. Gani
Muslem Daud

Syiah Kuala University, Banda Aceh

ABSTRACT

This study was aimed to find out how English teachers understand the concept of analysis of test items and apply it in constructing test items in Senior High Schools Harapan Persada and Tunas Bangsa Abdya. The targets were two sets of entrance test documents for 2016 and 2017. This research was conducted from 22 to 28 of July 2017. Descriptive analysis design was used as the research method. Interview and checklist were used as the research technique. The results showed that Senior High School Harapan Persada for the year 2016 test has 20 test items: 14 (revised), 6 (utilized), and 0 (changed). For 2017 test it has 30 test items: 3 (revised), 27 (utilized), and 0 (changed). Senior high school Tunas Bangsa for the year 2016 has 10 test items: 0 (revised,) 10 (utilized), and 0 (changed). For the 2017 test, it has 15 test items: 15 (utilized), 0 (revised), and 0 (changed). In assessing students at both schools, the maximum score for multiple-choice section is 100; as well as for interview. The final result was gained by calculating both of the sections, divided by two. Meanwhile, in interview test, personal introduction, school domain, daily activities, hobbies and interests, and purpose/decision were used as the measurement. There were differences and similarities in constructing test items of both schools. Senior high school Harapan Persada for the test of 2016 only implemented vocabulary and communicative skills, without including genres and short functional texts. However, for the test of 2017 it has implemented the integrated unit: transactional, interpersonal, short functional texts and genres. Meanwhile, Senior High School Tunas Bangsa (in 2016 and 2017) has implemented integrated unit in

*Corresponding author: nwildanur@yahoo.com

arranging English entrance test. It is suggested that the appropriate test construction procedure before providing tests be highly considered by the test developers, particularly English teachers.

Keywords: *test items, English entrance test, validity, reliability, students' ability*

INTRODUCTION

A learners' ability in any skill can be measured by using tests. This is because testing is recognized as one of the most powerful tools for measuring a students' ability and finding out whether they have already mastered certain learning materials or not. As stated by Brown (2004), a test is defined as a method of measuring a person's ability, knowledge, or performance in a given domain.

Darling-Hammond and Mc-Closkey (2008) state that testing can evaluate students' progress and develop standards for teaching-learning. According to Cronback in Azwar (2012), a test is a systematic procedure for observing someone's behavior, skills or ability and describing it with the aid of a numerical scale or category system. Based on Alias (2005), a test must possess two important characteristics, viz.: validity and reliability. However, tests can lead to a mismatch which then causes the test to fail to provide evidence from which teachers can make valid judgments about students' progress (Brookhart, 2003). In attempting accurate measurements of results, a good quality test has to be provided since the test not only influences the students' learning but also affects improvements in the teaching-learning processes followed by the teachers (McNamara, 1994). Teachers should not design test items carelessly, nor disregard any instructional objectives in the syllabus. A test can be a primary device to reinforce learning and motivate the performance of students in language learning (Heaton, 1989). Educators are supposed to always consider tests as the best tool for measurement possible (Bachman, 1990).

However, a problem appeared when the preliminary research was done (by the researcher herself) at two senior high schools (SMA) in South West Aceh (Abdya), namely SMA Harapan Persada and SMA Tunas Bangsa. It was found that the teachers of English at these two schools had never done any analysis of the English entrance tests that they had made and used. This information was obtained through

interviewing the English teachers who designed the English section of the tests (one teacher from each school). Basically, item analysis is a re-examination of the test items to find out the strengths and the flaws of each item (Bachman, 2004). This is one of the very important stages of creating a good test, which these two test-makers missed out on. Therefore, the writer was interested to investigate this issue further.

Several studies have been conducted related to the analysis of test items. One was done by Sahardin (2010). That research was done in order to find out the level of proficiency of the students doing National, English Final Examinations. That study aimed to find out which questions were the most difficult or easiest to answer for the students. The population was 42 SMA schools in the City of Banda Aceh and the District of Aceh Besar. The survey was done in 13 of them, selected at random, as the sample. The results showed that the mean total score of the students was 57 and it was found that 91% of the students could not answer all questions correctly.

A second study done by Suzanna (2010) found that most students could answer nearly all items correctly. 146 out of 180 participants could answer more than 60% of the items. Only 10% or 5 items could be classified as difficult. Meanwhile, the rest of the items, 90% or 45 items could be classified as average to easy items. Most of the students found that vocabulary/word power and factual fact finding in texts were the items that they had problems with. To help students solve such problems, teachers should train students to practice how to find factual information in texts and how to practice using new words which will help them increase their working vocabulary.

An earlier study was done by Brown (1991) in 16 high schools in total in Hawaii. Out of 10,858 students, who took an earlier test (Form G of the HSTEC), 300 were selected at random as an experimental group and 318 other students were selected as a normal sample or comparative group representing the high school students in Hawaii as a whole. The Form G contained 140 items, covering 14 Essential Competencies (EC). The test results showed that some students clearly passed all of the sub-tests, indicating that not all of the students were in need of help with the HSTEC.

Other relevant studies regarding entrance tests have also been done, one done in Turkey by Hatipoglu (2016) showed that the University Entrance Exams had both positive and negative effects. Then, a study was done by Salehi et al. (2011) in Iran where the entrance test was seen as a means to motivate the students and

encourage disciplined teaching-learning. This included providing the learning materials effectively and also emphasizing how to answer and solve problems during a test. Thus more effective teaching could lead to positive effects for the students in the entrance exams.

Another study done by Razmjoo and Madani (2013) was a content analysis of the English section of a university entrance exam based on Bloom's Revised Taxonomy. It was carried out to examine the test items in the English section of a university entrance exam by referring to the standards set out in the revised version of Bloom's Taxonomy. The findings showed that in 2011 the English section was focused on HOTS (Higher-Order Thinking Skills) while LOTS (Lower-Order Thinking Skills) still dominated in prior teaching-learning. Nevertheless, in Azad University the test items were more likely to be HOTS.

The analyses of the studies above were done quantitatively, meaning that the test items were analyzed after the students had taken the tests. However, in this present study the writer wanted to do the analysis qualitatively, that is, to analyze the test items before they were distributed. Besides, those earlier studies were conducted on the learners at university level or at the end stage of senior high school, but this study was carried out on learners at the end of junior high school entering senior high school level. As mentioned, both SMAs, Harapan Persada and Tunas Bangsa Abdya were taken as the subjects for the study as the English teachers at both those schools had prepared the entrance tests but had never done any analysis of their English entrance test items.

Research Questions

- 1.What are the results of the analysis of English entrance test at Senior High SchoolsHarapanPersada and Tunas Bangsa?
- 2.What is the procedure in scoring the English entrance test at Senior High Schools Harapan Persada and Tunas Bangsa?
- 3.What are the differences and similarities between English items for entrance test at Senior High SchoolsHarapan Persada and Tunas Bangsa?

LITERATURE REVIEW

Definition and Purposes of a Test

A test is one of the powerful tools used for measuring whether students have mastered learning materials. This is supported by Brown (2004) who has written that a test is a method of measuring a person's ability, knowledge, or performance. Also Arifin (2011, p. 118) has written, "a test is a technique or method used in order to carry out activities of measurement, in which there is a wide range of questions and statements, or a series of tasks". This is supported by Riduwan (2006) who has defined a test as an instrument for data collection formed with sets of questions or exercises which are used to measure the skills, knowledge, intelligence, aptitude and/or ability of an individual or group. Besides, Arikunto (2010) has stated that a test refers to a set of questions or other practices or devices used to measure the skills, intelligence, ability and/or talent of an individual or a group. Based on Azwar (2012), a test is a systematic procedure for observing someone's behavior and describing it. A good test is needed to get good data about student comprehension and ability (Cizek & Bunch, 2007). Sudjiono (2007, p. 99) has referred to a test as "a measure of development and progress of learners". Generally, tests are divided into two types, viz.: subjective and objective tests.

Bachman (1990) has stated that educators always want their tests to be the best measure possible. This is shown in the implementation of entrance tests. According to Alias (2005), a classroom test must possess two important characteristics: validity and reliability. Concerning the last statement, Brookhart (2003) has said that there can be both a real and a perceived mismatch between the content examined in class and the material assessed on an end of chapter/unit/term test. Moreover, a test can affect many aspects, depending on its purposes as Kappe and Flier (2012) have explained concerning high school grades and entrance exams such as standard admission tests and admission criteria tests. Besides, Rothstein (2004) has said that results hint at the importance of considering contextual factors that reflect the differences in students' opportunities to learn. Conversely, Clercq et al. (2013) have argued that tests can result in disadvantaging students because of the impact of using achievement measures for admission. From the explanations above it can be seen that a test that is good has to be well constructed and tested for its validity and reliability.

Item Analysis

Basically, *item analysis* examines the test items to find out the strengths and flaws of each item in a test (Bachman, 2004). Aiken (1994) describes item analysis as an important activity in the preparation of items in order to obtain well qualified question items. Besides, Arikunto (2012) argues that the purpose of analyzing test items is to get information about good or bad items. There are several benefits of reviewing or analysis of items. Anastasi and Urbina (1997) mention that “the main objective of test item analysis in a teacher-made test is to identify shortages in the test or in learning”. Besides, other benefits as stated by Arikunto (2012) include: a) to determine which questions are not useful; and b) to improve the test items through analysis of three features, namely (i) the level of difficulty, (ii) distinguishing any redundancy or duplication in questions, as well as (iii) improving learning through eliminating ambiguity and double entendre in items and certain skills which can lead to learners finding it difficult to answer a question. Furthermore, based on Purwanto (2010), “by doing the item analysis, we can obtain the important points from each question or item, that is: the extent of the rate or level of difficulty of the question and whether it has distinguishing features. These analyses will evaluate the quality of the test, enable it to be revised and improve it as a whole”.

According to Yusrizal (2015), in analyzing test items, there are two kinds of analysis which can be used, namely - qualitative and quantitative analysis. According to Anastasi and Urbina (1997), item analysis can be conducted qualitatively (related to content and form) and quantitatively (related to statistical features). Qualitative analysis includes the consideration of content and construction validity, while quantitative analysis includes the measurement of item validity and reliability, test item difficulty, and item discrimination. Each of these techniques has both advantages and disadvantages, and therefore the best technique is to employ or combine them both.

Qualitative Analysis

Principally, the item analysis is qualitatively executed based on the rules of item writing for written tests. This review is usually done before the question is used or tested. What should be qualitatively considered is the material, the construction, the language and/or the culture, and the answer key or scoring guide for each of the questions/items. There are two techniques that can be employed to

analyze the items qualitatively, viz: the moderator and the panel technique.

Criteria of A Good Test

Arikunto (2012, p. 72) has stated that “a good test as an instrument of measurement has some requirements, viz: validity, reliability, objectivity, practicability and also is economical”.

Validity

Linn and Gronlund (2000) as cited by Brown (2004) defines test validity as the extent to which inferences made from the results of the assessment are appropriate, meaningful and useful in terms of the purpose of the assessment. Whilst Hughes (2013) has said that validity refers to whether the test measures accurately what it is intended to measure. In short, to be valid, a test must consistently provide accurate measurements in order to be an appropriate, meaningful and useful assessment. According to Alias (2005), to be of real value in decision-making, the test must possess two important characteristics, i.e. validity and reliability. This means that a test has to be well constructed by considering its validity and reliability to provide an accurate measurement of the test-taker’s ability in a particular domain and to avoid inappropriate tests of what was taught-learnt in the class/course. The teacher should not design the test items carelessly and leave out some of the instructional objectives that are stated in the syllabus.

In the encyclopediac book of Educational Evaluation written by Anderson et al. cited in Arikunto (2012, p. 80), there is a definition viz: “A test is valid if it measures what it aims to measure, this means that the results of the evaluation must correspond with the facts”.

Arikunto (2012, p. 82) has said that there are four kinds of validity: content validity, construct validity, concurrent validity and predictive validity. Content validity, also called curricular validity, is related to the material from the curriculum that will be measured in the test. Purwanto (2010) has said that a test has content validity if there is relevance between the test item with the goal and the learning materials. Moreover, a test is said to have construct validity if the test items measure every thinking aspect such as are included in the goals specifically for the instructions and the syllabus. Construct validity can be determined by pairing every test item with aspects in the goals and the instructions from the syllabus. It is done based on logic, not experience (Arikunto, 2012, p. 83). Furthermore, checking the

concurrent validity is done empirically. This means that it can only be done after the data has been collected in the field. The testing of validity can be done in two ways from predictive validity and concurrent validity (Azwar, 2012).

Reliability

Reliability involves the consistency or reproducibility, of test scores. That is the degree to which one (e.g. a teacher) can get relatively constant scores from individuals across various testing situations with the same, or parallel, testing instruments. Hughes (2013) has said out that if a test is not reliable, it cannot be valid. However, Alias (2005) argues that a test may be highly reliable but not necessarily valid, but a highly valid test is usually reliable. To determine the co-efficient for this type of reliability, the same test is given to a group of subjects on at least two separate occasions. In designing tests, teachers must apply good criteria for tests when creating multiple-choice questions and interview-test topics. Good criteria for multiple-choice tests are set out below.

Purwanto (2010) has defined reliability as the “accuracy or carefulness of a tool of evaluation”. Then, Arikunto (2012, p. 75) contends that “a reliable measure is one that provides consistent and stable indicators of the characteristic being investigated”. In short, a test is reliable when it gives the same results when done repeatedly. According to Purwanto (2010, p. 141), the “Objectivity of a test is influenced by the level or similarity of quality of the scores obtained with the test even though the results from the test are judged by different assessors.” In other words, the structure of a test is called objective only if in conducting the test it is not interfered with by any subjective factors. The relation with reliability is that objectivity emphasizes the consistency of the scoring system, whereas reliability compares the results from repeated tests.

Practicability

According to Arikunto (2012, p. 77), “A test is said to have high practicability if the test is practical and easy to administer.” A test is said to have high practicability if it is practicable to administer easily, to complete and to score with clear marking criteria. The characteristics of the practicability of a test are pointed out by Arikunto (2012) as follows:

(a) easy to be done, for example, it does not need many tools and gives freedom to students to do the first part; (b) easy to do investigation, which means that it is complete with answer keys as well as guidance for scoring. For the objectivity of a test, the investigations are easily done if done by students and easy to fill in the answer sheet(s); (c) complete with clear instructions that can be given to many.

Economical

Economical requirements mean that implementation does not require much expense, much energy or much time.

Developing A Standardized Test

Brown (2004) has stated that a designer or a design team for a test should apply a standard for a test before designing it. Kappe and Flier (2012) have noted that high school grading and entrance exams such as standard admission tests and admission criteria tests are widely used in higher education around the world. Besides, Rothstein (2004) has said that the results hint at the importance of considering contextual factors that reflect students' differences in opportunities to learn. Conversely, Clercq et al. (2013) have argued that tests can result in disadvantaging some students because of the impact of using achievement measures for admission. Also, Geiser and Santelices (2007) have emphasised the simple predictive validity of standardized tests for an individual student.

RESEARCH METHODOLOGY

Cresswell (2008) has stated "the basic idea of a case study is that one case or a small number of cases will be studied in detail using whatever methods that seem appropriate to understand the case in depth". Thus, the descriptive qualitative method was utilized for the specific design of this study. According to Moleong (2009, p. 3), "qualitative research is research which produces descriptive data in the form of written or oral words from observing people and behavior".

For this research the writer selected two credible senior high schools (SMA) in South-West Aceh = Aceh Barat Daya (Abdya), namely SMA Tunas Bangsa and SMA Harapan Persada. Actually, entrance tests have not been used by any other SMAs in Abdya. In fact, only those two SMAs, where this research was done, used an entrance

test. The participants were the two English teachers who developed the English entrance tests at each of these schools plus several experienced teachers from various backgrounds who analyzed or reviewed the English test items.

In collecting the data for the analysis of the English entrance tests, two instruments were used, viz: a checklist and an interview guide. Data collection can be defined as identifying and selecting individuals for a study, obtaining their permission to conduct a study on them, and gathering information through questions or observation of their behavior (Creswell, 2008). In this study, the data on the entrance tests used at both schools was collected through interviews and the study of documents. The writer used an interview guide for a semi-structured interview as a checklist of the topics to be covered, followed up with unplanned questions asked to expand on what the interviewer needed to explore more (Creswell, 2008). Besides, triangulation of data was used to check the reliability of the data collected from the different sources. Meanwhile, for the study of documents, the entrance tests from the last two years were examined using a checklist.

Qualitative analysis was applied to describe the consistency of the entrance test items from 2016 and 2017 of the English tests made by the two teachers using the qualitative multiple-choice analysis and the Panel technique. The Panel technique was carried out by examining each item based on the rule of the test item writing, i.e. regarding the material aspect, construction aspect, language/culture aspect, and the correctness of the answer key or scoring guidelines which was done by several reviewers (Yusrizal, 2016). The test items that were analyzed were the test items from the English entrance tests for the two SMAs in 2016 and 2017. All of the data was obtained from the English teachers at the two SMAs used for this research. In the analysis, the scoring ranged from 1 to 3, in which 1 meant the test item should be changed i.e. deleted; 2 meant the test item should be revised; and 3 meant that the test item should be used. More specifically, according to the Panel technique, for items to be changed the scores ranged from 0 to 7, the items to be revised scored from 8 to 14, and the good items scored from 15 to 21 (Yusrizal, 2016).

RESULTS AND DISCUSSIONS

Results

Analysis of Entrance Test Items for SMA Harapan Persada in 2016

Table 1. Scores for Test Items for 2016 at SMA Harapan Persada, Abdya

Item	A	B	C	D	E	F	G	Total	Conclusion
11	2	2	2	2	2	2	2	14	Revised
12	2	2	2	2	2	2	2	14	Revised
13	2	2	2	2	2	2	2	14	Revised
14	1	1	1	1	2	2	1	9	Revised
15	2	2	1	1	1	1	1	9	Revised
16	2	1	1	2	2	1	1	10	Revised
17	2	2	2	2	1	1	1	11	Revised
18	2	2	2	2	2	2	2	14	Revised
19	2	2	2	2	1	1	1	11	Revised
20	1	1	2	2	2	1	1	10	Revised
21	1	1	1	2	2	2	1	10	Revised
22	2	2	1	1	1	2	1	10	Revised
23	2	2	2	2	2	1	1	12	Revised
24	3	3	3	2	2	2	2	17	Good
25	2	2	3	3	3	3	3	20	Good
26	3	3	2	2	2	2	2	16	Good
27	3	2	2	2	3	3	2	17	Good
28	2	2	3	3	2	2	2	16	Good
29	2	2	3	3	2	2	2	16	Good
30	2	2	2	2	2	2	2	14	Revised

Explanation: Teacher experience:

A = teaching junior high school students,

B = junior high school learning materials,

C = curriculum development,

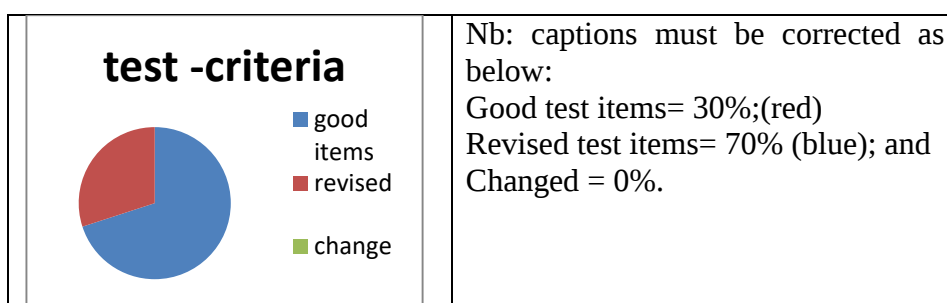
D = latticework of tests,

E = test construction,

F = grammar, and

G = English diction

The result from the analysis of the data showed that 14 of the test-items should be revised and 6 should be used while no test-items should be changed. To sum up the analysis from the table above, is shown in the pie chart that follows:

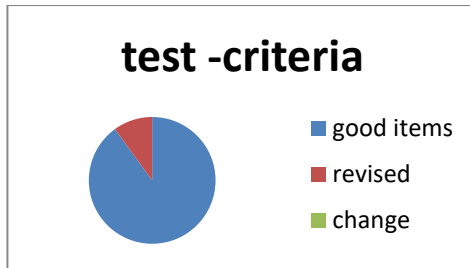


Analysis of Entrance Test Items for 2017 from SMA Harapan Persada

Table 2. Scores for Test Items for 2017

Item	A	B	C	D	E	F	G	Total	Conclusion
1	3	2	2	2	2	2	2	15	Good
2	3	3	3	3	3	2	2	19	Good
3	2	2	2	2	2	1	1	12	Revised
4	2	2	2	2	3	3	3	17	Good
5	3	3	3	2	2	3	2	18	Good
6	2	2	2	2	2	3	3	16	Good
7	3	3	3	3	2	2	2	18	Good
8	2	2	2	2	2	1	1	12	Revised
9	3	3	3	3	3	3	3	21	Good
10	2	2	2	2	2	2	2	14	Revised
11	3	3	3	3	3	3	3	21	Good
12	3	3	3	3	3	3	3	21	Good
13	3	3	3	3	3	3	3	21	Good
14	3	3	3	3	3	3	3	21	Good
15	3	3	3		3	3	3	21	Good
16	3	3	3		3	3	3	21	Good
17	3	3	3	3	3	3	3	21	Good
18	3	3	3	3	3	3	3	21	Good
19	3	3	3	3	3	3	3	21	Good
20	3	3	3	3	3	3	3	21	Good
21	3	3	3	3	3	3	3	21	Good
22	3	3	3	3	3	3	3	21	Good
23	3	3	3	3	3	3	3	21	Good
24	3	3	3	3	3	3	3	21	Good
25	3	3	3	3	3	3	3	21	Good
26	3	3	3	3	3	3	3	21	Good
27	3	3	3	3	3	3	3	21	Good
28	3	3	3	3	3	3	3	21	Good
29	3	3	3	3	3	3	3	21	Good
30	3	3	3	3	3	3	3	21	Good

The data above showed that: 3 (10%) of the items should be revised and 27 (90%) can be used while none (0%) should be changed as shown in the pie chart below.



Test Items from 2016 for SMA Tunas Bangsa

Table 3. Scores for Items on English Entrance Test in 2016

Item	A	B	C	D	E	F	G	Total	Conclusion
1	3	3	3	3	3	3	3	21	Good
2	3	3	3	3	3	3	3	21	Good
3	3	3	3	3	3	3	3	21	Good
4	3	3	3	3	3	3	3	21	Good
5	2	3	2	2	2	3	3	17	Good
6	3	3	2	2	2	3	3	18	Good
7	3	3	3	2	2	2	2	17	Good
8	3	3	3	2	2	2	2	17	Good
9	3	3	3	3	3	3	3	21	Good
10	3	3	3	3	3	3	3	21	Good

The results from the data analysis showed that: none of the test-items should be revised or changed and all 10 can be used.

Analysis of the Entrance Test Items for 2017 from MA Tunas Bangsa

Table 4. Scores for Items on English Entrance Test for 2017

Item	A	B	C	D	E	F	G	Total	Conclusion
1	3	3	3	2	2	2	2	17	Good
2	3	3	3	2	2	2	2	17	Good
3	3	3	3	2	2	2	2	17	Good
4	3	3	3	2	2	2	2	17	Good
5	3	3	3	2	2	2	2	17	Good
6	3	3	3	2	2	2	2	17	Good
7	3	3	3	3	3	3	3	21	Good
8	3	3	3	2	2	2	2	17	Good

9	3	3	3	2	2	2	2	17	Good
10	3	3	3	2	2	2	2	17	Good
11	3	3	3	2	2	2	2	17	Good
12	3	3	3	2	2	2	2	17	Good
13	3	3	3	2	2	2	2	17	Good
14	3	3	3	2	2	2	2	17	Good
15	3	3	3	2	2	2	2	17	Good

The results from the analysis of the data showed that all 15 test-items can be used and none should be revised or changed.

Interview Topics

Table 5. Topics Covered in Interview

No.	Aspect	Description
1	Personal introduction	Self introduction starts from the name, age, coming from where, living where and so on of the interviewee.
2	School domain	School of origin, ie. the school that the interviewee comes from is an important part of the usual questions in the interview.
3	Daily activities	What do you do every day, don't mention ugly habits such as lazy or always play. Try to mention useful activities.
4	Hobbies and interests	Talk about hobbies that you do and sport(s) you play regularly, like reading and football or.
5	Purpose/decision	Provide your reasons and purpose for wanting to enter this SMA? For example, answer if the school is a favourite one for you.

Interview for senior high school entrance test (directive test for high school level).

Scoring

Multiple choice = Maximum score 100

Interview = Maximum score 100

Differences and Similarities between English Items for Entrance Test

In general the test items constructed by the two schools were the same. The differences and similarities between English items for entrance test between Senior High Schools Harapan Persada and Tunas Bangsa can be seen in the test construction. For the multiple-choice items of 2016 for Senior High School Harapan Persada, it only implemented vocabulary and language aspects, which did not include

genres and short functional text. Contextual Learning Materials of English textbook (student's book grade IX 2008) had been arranged based on Standard of Content (SI) of the 2006 curriculum. Its aims to help increase students' communicative competencies. The communicative competencies were implemented through four language skills, namely listening, speaking, reading and writing, which are developed integrately. The integrated unit is implemented into unit development, based on *transactional, interpersonal, short functional texts* and *genres*. These are some efforts to fulfil the source learning material relevantly. Meanwhile, regarding the test construction of Senior High School Tunas Bangsa, this school had implemented integrated units in arranging English entrance test. It is can be seen in the test analysis in the table above.

Discussion

Senior High School Harapan Persada

Most of the mistakes in designing the test items of entrance exam were found in the language aspect – Point C (*language*) of item analysis by Balitbang-Depdiknas 2007, especially No. 14 “Using the grammatically appropriate language” and No. 15 “Using the communicative language”. The language issues that are detected include:

Punctuation. In item **12**, **14**, **15**, and **23** of 2016, the interrogative sentences are not ended by the question mark (?), but by full stop (.). This can make the testee confused to interpret the question content. Besides, the meaning of the sentences changes semantically as well. Item **9** of 2017 covers a sentence which is quoted from the provided *Announcement* text, i.e. “Preparation for school *anniversary*”, but the sentence is not enclosed by the quotation marks (“...”), while this is crucial to clarify to the students that the particular sentence is part of the whole passage described before.

Grammar or Structure. Item **18** of 2016 and items **4** and **6** of 2017 ask for the underlined word/sentence. However, the writing of question items is not in accordance with the English grammar and does not follow the consensus or general agreement of English question item writing. Item 18 states “The underline is about...” and items 4 and 6 state “The underline word has the same meaning with...”. Grammatically and habitually, it should be: “The underlined word/sentence is about...” or “the underlined word/sentence has the

same meaning with..." or "The underlined word/sentence can be replaced by..."

In item **19** (2016) the sentence "There is a Phiton in my bedroom last night." It is clearly mistaken. First of all, the adverb of time *last night* indicates that the sentence is considered *past tense*, and thus the sentence should start with *there was* instead of *there is*. Secondly, the word *Phiton* in English is spelled *Phyton*. Item **20** (2016) contains sentences which, in my opinion, have no *sense of English*. Rather than writing "Look at that! A new apartment", the writer would prefer "Look at that new apartment!" or "Look at that! It's a new apartment." or "Look at that! It's my new apartment". This will sound more *native like* and *natural*. Moreover, the sentence "My parents give me as a special gift" is ambiguous because logically, according to the tenses (time), if A talks about the apartment at that time, that means his/her parents gave him/her in the past. Thus, it should be "My parent gave me as a special gift" or "My parent gave it to me as a special gift".

Item **21** (2016) looks ambiguous in the sentence "Is it delicious, isn't it?". If test makers intend to make an ordinary interrogative sentence, they are supposed to write "Is it delicious?". If they want to make a sentence with a *question tag*, they should write "*It is* delicious, isn't it?". In item **22** (2016), "I waited you" is a preposition error. Indonesians are used to producing such an erroneous sentence due to the influence of the first language: i.e. "saya menunggu kamu". Meanwhile, in English we know that certain verbs are frequently paired with certain prepositions. Therefore, the sentence should be "I waited for you". Even though it seems trivial, this mistake must be avoided for the sake of communicative test items and professional test item makers.

For items **19–24** (2016), test instructions are missed. Of course, from the sequence/arrangement of questions it can be guessed that the instructions for these items are the same as the prior item (no. **18**): asking for the meaning of the underlined sentence. Yet, it is feared that not all examinees have a good reasoning/interpreting ability. To be more effective and efficient, the similar items are better grouped into and referred to as one instruction, as for items **27–28** or **29–30**. For example, "For questions 18 to 24, choose the answer that represents the meaning of the underlined sentence(s)!". Item **30** (2016) contains an uncommunicative redaction. Of all options/choices provided, we know that students are required to identify what kind of tense the sentence is included to. Unfortunately, the question redaction "From the sentence we know that ..." is too broad. Rather, it needs to be directly focused or

lead to one specific issue, such as “The sentence is included to ... tense.” or “what kind of tense(s) is it?”.

Redaction of item 3 (2017) is also grammatically wrong. The sentence “Who are probably come ...” should be written “Who probably come(s) ...” or “Who will probably come ...”. Also for item 8 (2017), “Who are come ...” should be “who come(s) ...” or “Who will come ...”. In terms of topic matter, the test items have overall been fitted with indicators and competencies for junior high school (SMP) level students. However, in some parts of the whole items, a few mistakes that can reduce the value of test item construction were found – Point B (*construction*) of good multiple-choice criteria by Yusrizal 2015, that is:

In item 12 (2016) two mistakes were detected. First, *typo* on the multiple choices: instead of *a-b-c-d*, it is typed *a-b-b-d*. The thing is that the correct answer is in the two b provided, which is b. plane. Secondly, the options a (bus), b (plane), and b (train) have been appropriate and homogeneous (all are transportation), but option d seems to be made up with the word pelene, while there is no such a word in English. It would be better to include English vocabulary that belongs to other means of transportation as well. The options in item 16 (2016) are considered as unchallenging or have no significant distractors. In other words, this item highlights the correct answer too much, which is a. dislikes lying; while the other options show that the subject likes/loves lying. These would be better to be replaced with words meaning do not like lying too, but made wrong on the grammar aspect, such as dislike (without +s) or don't like (should be doesn't like).

Senior High School Tunas Bangsa

Generally, the items of this entrance exam have no problem with material and language aspects. Still, some mistakes were detected in terms of technical and item construction (typing), that is: In the cloze texts such as for items 15, 16, and 17 (2016), it is written number 14, 15, and 16 on the passage instead. Although it is just a small technical mistake, this at least can also make the testee confused and can ruin their concentration during the test. Also for the next item (no. 18), the first sentence starts with no numbering (number 1).

In most of the test items of 2017, mistakes are clearly seen in the missing of test instructions, especially for groups of questions which refer to a certain passage/text described before them, such as

announcement, message, chart or advertisement. These groups of question items are supposed to begin with a clear and bold instruction, for instance, "Read the following text for questions 1 to 3."

To strengthen the result from the data analysis for the multiple-choice and interview test, the writer has to describe the result of data obtained from English teachers of both schools. The use of interview in this study was to make the researcher discover the reasons behind the test construction made by Senior High Schools Harapan Persada and Tunas Bangsa.

The Use of interview helped the researcher to identify some pertinent points relevant to the study and would not have been obtained through the other methods (Creswell, 2008). Besides, Arifin (2011) says that interview is a way in collecting some information deeply. The information was gathered orally and directly between the researcher and the interviewer. The questions were as in the following: how are English interview and multiple choice items for entrance test constructed at Senior High Schools Harapan Persada and Tunas Bangsa? the answers were that the tests were constructed by English teachers. Is there any training received in order to construct entrance test? The answers were that they have never involved in any training regarding entrance test. What is the procedure in scoring the English entrance test at Senior High Schools Harapan Persada and Tunas Bangsa? The writer found that the total multiple choice scores and interview scores were divided by two. Is there any the procedure in scoring the English entrance test; the procedure were a total of 100% for multiple choice plus 100% for interview. What are the criteria used in scoring the English entrance test; is there any other consideration of added values on the scoring (such as interview score and the percentage); there were no other considerations besides *written and interview scores*. Do you think your English entrance test is valid; do you think your English entrance test is reliable; for these questions both of English teachers explained that they had never analyzed the entrance test yet.

The differences and similarities between English items for entrance test at Senior High Schools Harapan Persada and Tunas Bangsa are as follows; Do you think if there are differences between English items for entrance test at your school and other schools; do you think if there are similarities between English items for entrance test at your school and other schools; in general they were the same but the data showed that senior high school Harapan Persada had 50% integrated unit, Tunas

Bangsa had interrupted integrated unit and genre based 100% of their entrance test. How can English items for entrance test at your school be improved? For the last interview they answered both English teachers needed teacher training, especially related to the constructed English entrance test.

CONCLUSIONS AND SUGGESTIONS

The results from the analysis of the Multile-Choice portion of the entrance tests are summarised in the table below:

Table 5. Results from Analysis of English Multiple-Choice Questions in Entrance Tests

SMA	Year	Questions			
		Change	Revise	Good	Total
Harapan Persada	2016	0	14	6	20
ditto	2017	0	3	27	30
Tunas Bangsa	2016	0	0	10	10
ditto	2017	0	0	15	15

The data from Senior High School Harapan Persada showed that for the interview section, 10 items were tested, including personal introduction, school domain, daily activities, hobbies and interests and purpose/decision. Meanwhile, Senior High School Tunas Bangsa had different items. There were 5 items for the interview test. The questions were adjusted to the junior high school (SMP) level, which basically expects achievement oriented competences. The differences and similarities between the English entrance test items from SMA Harapan Persada (HP) and SMA Tunas Bangsa (TB) were that, in 2016, HP still had not combined the communicative competencies into the multiple-choice construction; it had not implemented them as suggested by the Standard of Content (Standar Isi), which includes genres and short functional texts, but HP then included them in the 2017 entrance tests. Meanwhile, TB had implemented integrated units in constructing their English entrance tests in both 2016 and 2017. This can be seen from the analysis of the multiple choice test items, the results of which were 100% good.

Some suggestions from the researcher are as follows. Firstly, it is highly recommended to use English entrance tests when entering senior high school (SMA) level as it has been proven to raise students' achievements. Secondly, it would be better to use English in-service

teacher training for constructing better tests and analyzing them, working collaboratively instead of individually in preparing the tests or considering which test is the best one to use. Lastly, since studies on English entrance tests at SMA and Tertiary education levels are still fairly limited, it is essential that more related studies are done at other schools/institutions in the future.

REFERENCES

- Aiken, L. R. (1994). *Psychological testing and assessment*, (Eight edition), Boston: Allyn and Bacon.
- Alias, M. (2005). Assessment of learning outcomes: Validity and reliability of classroom tests. *World Transactions on Engineering and Technology Education*, 4(2), 235-238.
- Anastasi A., & Urbina, S. (1997). *Psychological testing*. (Seventh Edition). New Jersey: Prentice-Hall, Inc.
- Arifin, Z. (2011). *Evaluasi pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Arikunto, S. (2012). *Dasar-dasar evaluasi pendidikan*, (Edisi Dua). Jakarta: Bumi Aksara.
- Arikunto, S. (2010). *Prosedur penelitian suatu pendekatan praktik*. Jakarta: Rineka Cipta
- Azwar, S. (2012). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar.
- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Balitbang-Depdiknas. (2007). *Panduan penulisan soal pilihan Ganda*. Jakarta.
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. educational measurement. *Issues and Practice*, 22(4), 5–12. doi:10.1111/j.1745-3992.2003.tb00139.x
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.
- Brown, J. D. (1991). Performance of ESL students on a state minimal competency test. *Issues in Applied Linguistics*, 2(1), 31-47.

- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.
- Clercq, M. D., Galand, B., Dupont, S., & Frenay, M. (2013). Achievement among first-year university students: An integrated and contextualized approach. *European Journal of Psychology of Education*, 28(3), 641-662.
- Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, Merrill: Pearson Education, Inc.
- Darling-Hammond, L., & Mc-Closkey, L. (2008) Assessment for learning around the world: What would it mean to be internationally competitive?. *The Phi Delta Kappan*, 90(4), 263-272.
- Geiser, S., & Santelices, M. V. (2007, June). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes* CSHE Research & Occasional Paper Series. Bereley: University of California. Retrieved from <https://cshe.berkeley.edu/publications/validity-high-school-grades-predicting-student-success-beyond-freshman-yearhigh-school>
- Hatipoglu, C. (2016). The impact of university entrance exam on EFL education in Turkey: Pre-service English language teachers' perspective. *Social and Behavioral Sciences*, 232, 136 – 144.
- Heaton, J. B. (1989). *Writing English language test*. New York: Longman.
- hughes, a. (2013). *testing for language Teachers*. Cambridge: Cambridge University Press.
- Kappe, R., & Flier, H. V. D. (2012). Predicting academic success in higher education: What's more important than being smart? *European Journal of Psychology of Education*, 27(4), 605-619.
- McNamara, T. F. (1994). *Measuring second language performance*. New York: Longman.
- Moleong, L. J. (2009). *Metodologi penelitian kualitatif*, Edisi Revisi. Bandung: Remaja Rosdakarya.
- Purwanto, M. N. (2010). *Prinsip-prinsip dan teknik evaluasi pengajaran*. Bandung: Remaja Rosdakarya.
- Razmjoo, S. A., & Madani, H. (2013). A content analysis of the English section of university entrance exams based on Bloom's

- Revised Taxonomy. *International Journal of language Learning and Applied Linguistics World*, 4(3), 105-129.
- Riduwan. (2006). *Metode dan teknik menyusun thesis*. Bandung: Alfabeta.
- Rothstein, J. (2004). College performance predictions and the SAT. *Journal of Econometrics*, 121, 97-317.
- Sahardin, R. (2010). *An analysis of some question items of the English section of the UAN examination for senior high school students of 2006/ 2007 academic year*. (Unpublished master's thesis). Universitas Syiah Kuala, Banda Aceh.
- Salehi, H., Yunus, M. M., & Salehi, Z. (2012). Teachers' perceptions of high-stakes tests: A washback study. *International Journal of Social Science and Humanity*, 2(1), 70.
- Sudjiono, A. (2007). *Pengantar evaluasi pendidikan*. Jakarta: PT Raja Grafindo Persada.
- Suzanna. (2010). *Analyzing students' ability in answering English test of National Final Examination (UAN) 2007/2008 Academic Year (A case study at junior high school students in Kec. Rantau, Kab. Aceh Tamiang)*. (Unpublished master's thesis). Universitas Syiah Kuala, Banda Aceh.
- Yusrizal. (2015). *Tanya jawab seputar pengukuran, penilaian, dan evaluasi pendidikan*. Banda Aceh: Syiah Kuala University Press.
- Yusrizal. (2016, October). Lecture notes distributed in the course *Analisis butir* Program Studi Magister Pendidikan Bahasa Inggris. Universitas Syiah Kuala Banda Aceh.