

Item Analysis of English Summative Test at The Eleventh Grade Students of SMA Negeri 9 Kota Jambi City Academic Year 2018/2019

Ardhan Khariri, Suyadi

English Education Study Program, Faculty of Teacher Training and Education,
Batanghari University

ABSTRACT

Summative test is method of evaluation performed at the end of a unit that allows a teacher to measure a student's understanding, typically against standardized criteria. Therefore, to measure students' comprehension of the material learned, the item analysis of summative test is necessary. The research is aimed to find out the level of English Summative Test items for the eleventh grade students of SMAN 9 Kota Jambi 2018/2019. This research is categorized as quantitative method. The researcher took one class XI IPA 1 which consisted 32 members of student. The total number of the test is 40 items. This research is supported by ANATES 4.0 and Manual Counting to analyze item test. Based on item analysis and discussion, the reseacher found that the difficult item was six items (15%) with the test numbers are 3, 6, 8, 26, 27, 29. Secondly, medium item was nineteen items (47,5%) with the test numbers are 1, 4, 5, 7, 10, 11, 15, 20, 23, 25, 31, 32, 33, 34, 35, 37, 38, 39, 40. The last, the easy item was fifteen items (37,5%) with the test numbers are 2, 9, 12, 13, 14,16, 17, 18, 19, 21, 22, 24, 28, 30, 36. Most of the items were categorized as medium interpretation, and only a few in easy and difficult interpretation. It concluded that the total index level of the test is 0,59. Therefore, the summative test is categorized in medium level. The researcher concludes that the English summative test for eleventh grade at SMAN 9 Kota Jambi is not well enough yet and it has quite difficulties as well.

Keywords: *Item Analysis, English Summative Test.*

INTRODUCTION

Most teachers contend that testing is all too often viewed as a necessary challenge. Testing provides the teacher with clues and details related to the effectiveness and soundness of a specific teaching approach and method. It also gives an objective evaluation of learner's progress individually, his attainment of the set course objectives and his performance in relation to that of the other classmates. Valette (1977) adds, One should view testing as a bridge-building process between teaching and learning and classroom tests as mirrors in which teachers and students see their reflections clearly.

Additionally, Richards et al (1985) views, In educational terms, a test can be defined as any procedure for measuring ability, knowledge and performance. Meanwhile, Brown (1994) notes, A test is a method of measuring a

person's ability or knowledge in a given area. Then, Allan (1995) adds, in very practical terms, tests yield scores that mirror attributes or characteristics of individuals. Brown's definition seems to be more comprehensive in the sense that it covers all the main components of a test.

Test is often contributed to the process as the implementation of evaluation. Indeed, a test is one kind of evaluation instrument to collect data. A test is defined as a systematic procedure for observing and describing one or more characteristics of a person with the aid of either a numerical scale or category system" In other word, a test measures a person's ability or knowledge with a number of tasks or questions. Furthermore, there are some tests which can determine the student's competence grades in the past of the learning activities in the classroom, one

of the tests is achievement test. The type of the achievement test which often used by the teacher to evaluate the successfulness his or her teaching and learning in the classroom is summative test. In this research, the researcher will focus analyzing Summative test.

What makes summative test so invaluable is that each high-stakes test may result in educators using the data for decisions with significant long-term consequences affecting a student's future. Additionally, Moss (2013) defines summative test plays a role in improving future instruction by providing educators with data on the effectiveness of curriculum and instruction. Knowing what methods worked for a lesson or semester may not help current students, but it can provide educators with the necessary insights into how and where to redesign instructional practices to elevate next year's student scores. Further, he summative test is used in the end of the courses of instruction to know the students' performance grade whether they have already mastered all of the materials which they have reached while the teaching and learning process.

THEORIES OF SUMMATIVE TEST

Summative assessments are efforts to use information about students or programs after a set of instructional segments has occurred. Their purpose is to summarize how well a particular student, group of students, or teacher performed on a set of learning standards or objectives.

According to Başol & Johanson (2009), Summative assessment is a form of appraisal that occurs at the end of an instructional unit or at a specific point in time, such as the end of the school year. Meanwhile Tinambunan (1998) states, summative test is a test that usually administered at the end of the course test items on the syllabus.

In addition, Hart et al (2015) explains, Summative assessment plays a pivotal role in education by troubleshooting weaknesses in the

system. It provides educators with valuable information to determine the effectiveness of instruction for a particular unit of study, to make high-stakes decisions, and to evaluate the effectiveness of schoolwide interventions. It works to improve overall instruction (1) by providing feedback on progress measured against benchmarks, (2) by helping teachers to improve, and (3) as an accountability instrument for continuous improvement of systems

1. Item Analysis in Summative Test

Item analysis uses statistics and expert judgment to evaluate tests based on the quality of individual items, item sets, and entire sets of items, as well as the relationship of each item to other items. It "investigates the performance of items considered individually either in relation to some external criterion or in relation to the remaining items on the test" (Thompson & Levitov, 1985). It uses this information to improve item and test quality. Item analysis concepts are similar for norm-referenced and criterion-referenced tests, but they differ in specific, significant ways, criterion-referenced tests, use norm-referenced statistics for pretest data and criterion-referenced statistics for posttest data.

According to Purwanto (2009), A good test item should have three criteria; moderate difficulty level, high discriminating power and distracter analysis which work effectively. Then, Sudijono (2008) also says, The best way to analyze test items is using difficulty level, high discriminating power, and distracter analysis.

Those suggestions above assume that untrained persons will know relatively little about pretest material, so the assumptions on which norm-referenced statistics are based are applicable. Once people are trained, a test is criterion-referenced, and criterion-referenced statistics must be used. In the same way, item analysis tells us basically how difficult each item.

2. Multiple-Choice Items in Summative Test

Gronlund and Linn (1990) state, a multiple-choice item is generally recognized as the most widely applicable and a useful type of objective test items. Meanwhile, Marshall and Hales (1972) describe, the strengths of this item. Firstly, Multiple-Choice item is flexible and versatile of all selection type examinations.

Gronlund and Linn (1990) also explain, versatility in multiple-choice means that multiple-choice can be used to measure a variety of learning outcomes from simple to complex. Secondly, Multiple-choice can be scored rapidly, accurately, and objectively by all individuals who are unqualified to teach the subject area examined. Thirdly, the scoring of multiple-choice is not influenced by all outside performances. It means that the multiple-choice items have range control of students' way of thinking through the provided alternative answers. Students have to know exactly what the items ask about then they will choose the correct answer to answer the items. A multiple-choice items are made up of an item stem, which present questions, and several alternative answers. The options usually of, *a*, *b*, *c* or *d*. that will be counted correct, and the distractors, which are those choices that will be counted as incorrect.

However, Multiple-choice item has some disadvantages. Firstly, Harris (1969) mentions, constructing multiple-choice item requires considerable skill and time to prepare. Secondly, Alderson (2000) adds, the tester does not know how the students' ways of thinking. The tester does not know whether the students really know the answer or just guess the answer. The students may guess at their choices or they may have totally different reasons in mind from that which the test-maker intends when they are writing the items. It is possible to get an item correct for 'wrong' reason or to get an item wrong for the 'right' reason. Thirdly, Anderson et al (2001)

states, Multiple-choice items cannot be used to measure all objectives at all levels of cognitive domain.

THEORIES OF TEACHING ENGLISH AS A FOREIGN LANGUAGE

1. Teaching English in Indonesia

In Indonesia, Teaching of English as a foreign language seems to be on the crossroads due to two different curricula implementation. Curriculum changes are mandatory; however, the changes should not be counter-productive to the attainment of expected learning competency.

The newly advocated curriculum is known as 'K-2013' or the 2013 Curriculum. K-2013 is so designed that reflects a scientific approach to learning. K-2013 focuses its learning on the attainment of Kompetensi Inti (Core Competence) and Kompetensi Dasar (Basic Competence). The core competence pivots around (1) religious belief and its application, (2) upholding good characters, (3) gaining factual, conceptual and procedural knowledge. (4) trying, processing and presenting in concrete and abstract domains. While the basic competence transforms the core competence into observable cognition, affection and action (Kementerian Pendidikan dan Kebudayaan, 2013).

When observing, language learners are encouraged to observe by reading, watching, listening to exemplars or models from different learning sources beyond and over the teacher. They have to focus on the social function, text structure, linguistic attribute of each text observed. When necessary, teachers provide examples and contexts of usage by which learners imitate, write and use them appropriately.

The next phase is questioning between teacher-students, students-teacher or students-students. Questions may pivot around social function, text structure, linguistic attribute of the text itself. Through questioning, learners will

gain factual, conceptual and procedural knowledge about English in its socio-cultural contexts. Exploration and experimentation are the next phase of learning. During this phase, learners will read, listen to, and watch videos from different sources other than presented in class. Association and analysis will engage learners to work collaboratively in small groups. Learners analyze the content, structure, tone, meaning of word, phrase and sentence of the text. Learners learn in small groups and get feedbacks from their peers and the teacher as well.

2. Factor Governing Learners' Success in EFL

Reports on a multivariate study that examined the contribution of two independent variables—cognitive / academic proficiency in first language and attitudes and motivation toward English as a foreign language. Second language acquisition research has shown that the level of proficiency in the first language has a direct influence on the development of proficiency in the second language. The lack of continuing first language development has been found to inhibit the levels of second language proficiency and cognitive academic growth. Saville - Troike (1984) reports, "in almost all cases, the bilingual instructors' judgments of students' relative competence in native language studies coincided with the same students' relative achievement in English. Hakuta (1990) views, native language proficiency as a strong indicator of second language development.

Larsari (2011) has noticed, the failure of form focusing approach in developing learners' communicative competence in real-life situations and shifted to adopt the communicative language teaching (CLT). The CLT approach highlights learners' communicative competence, which is defined as learners' ability to efficiently express what they mean in the target language and successfully, achieve

communications in real-life situations. In order to do so, learners not only need to acquire the linguistic but pragmatic knowledge of the target language. It is suggested that competence, both linguistic and pragmatic, is the knowledge developed and acquired through exposure and use. In other words, without sufficient exposure needed for learners to notice and acquire the language input and chances to use the knowledge, communicative competence is not likely to be promoted.

Meanwhile, Ahmadi & Gilakjani (2011) focused, his research on listening comprehension. He reasoned that listening plays a significant role in daily communication and educational process. In spite of its importance, listening has long been the neglected skill in second language acquisition, research, teaching, and assessment. However, in recent years there has been an increased focus on L2 listening ability because of its perceived importance in language learning and teaching. The study tries to find the factors influencing English listening comprehension and the strategies to be taken that might improve students' listening comprehension. The study focused on four main issues. First, it discusses the definition of listening, significance of listening. Second, it reviews the process of listening comprehension, strategies of listening comprehension. Third, analysis of listening comprehension problems is reviewed. Fourth, teaching methods for listening comprehension will be discussed. Fifth, researchers review teaching listening activities. Sixth, general principles in teaching listening comprehension are discussed. Findings based on the review of the literature along with analysis of the data are of great significance and can be advantageous to improve EFL learners' English listening comprehension skill.

Therefore, Summative assessments can serve as a guide to improving teaching methods in Teaching English as a Foreign

Language. Many English teachers use different teaching methodology within the classroom. Summative assessments can help them collaborate and improve teaching English methods from year to year.

RESEARCH METHODOLOGY

This research is categorized as quantitative method with descriptive approach. According to Ary et al (2010), Quantitative research uses objective measurement to gather numeric data that are used to answer question or test predetermined hypothesis. Meanwhile, Morrell and Carroll (2010) state, Quantitative studies are linked to statistics. Most statistical studies phrase their question in the form of a hypothesis. According to Smith et al (1975), Quantitative is based on testing a theory, measured with numbers, and analysed using statistical techniques and it particularly emphasizes objectivity and reproducibility. Then, Fraenkel & Wallen (2003) argues that the goal of quantitative methods is to determine whether the predictive generalizations of a theory hold true.

Additionally, Fraenkel & Wallen (2003) explain, Quantitative research can be classified as either descriptive or experimental research. The purpose of descriptive research is to become more familiar with phenomena, to gain new insight, and to formulate a more specific research problem or hypothesis. In contrast, experimental research is to test cause and affect relationships among variables. In descriptive research, researchers do not have direct control over independent variables because their manifestations have already occurred or because they are inherently not manipulable (Kerlinger & Lee, 2000).

According to Polit and Beck (2003), Population refers to the aggregate or totality of those conforming to a set of specifications. Then, Ary et al (2010) describes, a population is defined as all members of any well-defined class of people, events,

or objects. Meanwhile, Sugiyono (2010) says, Population is defined as an area of generalization consisting of: objects or subjects that have certain qualities and characteristics set by researchers to be studied and then drawn conclusions. In addition, Netra (1974) says, Population is a whole individual that is general and has the same characteristics.

Based on explanations above, it can be concluded that population is the number of living things that live together in the same place. This research is conducted at eleventh grade students of SMAN 9 Kota Jambi

Table 1. Total of the Students at SMAN 9 Kota Jambi

Tahun Pelajaran	Jumlah Siswa			Jumlah
	Kelas X	Kelas XI	Kelas XII	
2015-2016	320	266	232	818
2016-2017	300	308	278	886
2017-2018	243	278	301	822
2018-2019	248	238	269	755

Sample is a representative selection of the population and using the data collected as research information, as Lohr (1999) says, Sample should be representative in the sense that each sample unit will represent the characteristics of known number of units in the population. In addition, The researcher uses purposive sampling as sample methods of this research. According to Maxwell (1996), Purposive sampling is a strategy in which particular settings persons or events are selected deliberately in order to provide important information that cannot be obtained from other choices. Additionally, Sugiyono (2010) describes, Purposive sampling is a technique for determining research samples with certain considerations that aim to make the data obtained more representative. Using purposive sampling means choosing selectively

with certain considerations as sample to be included in the research.

To collect data, the researcher needs some instruments, the kind of instrument is documentation. According to Arikunto (2013), There are some objects are considered in obtaining information and one of them is paper or document. In this research, some documents will be collected and analyzed. They are questions test paper, answer sheet and answer key. The explanation of these instruments can be seen as follows: 1) The researcher applies English summative test first semester at eleventh grade students of SMAN 9 Kota Jambi 2017/2018 academic year. It consists of 40 items in multiple-choice form. The data source was taken from the English teacher. 2) Answer sheets. This answer sheets used to know the answer distribution. 3) Answer key. This answer key used as a valid guide in scoring each item.

The researcher uses quantitative with descriptive approach, the researcher focuses on analyzing level of English Summative test, the researcher calculates the data by using some formulas, the analyzing is conducted in the same time and procedure.

The steps of analyzing difficulty level of the English Summative Test for eleventh grade students at SMAN 9 Kota Jambi 2017/2018 Academic Year. According to J.B Heaton (1975), The index of difficulty (FV) is generally expressed as the fraction (or percentage) of the students who answered the item correctly. It's calculated by using the formula:

$$FV = \frac{R}{N}$$

FV = index of difficulty

R = represents of the number of correct answer

N = represents of the number of students taking the test

Additionally, the researcher uses Anates 4.0 application as well to get accurate results, and save time. Steps of

analyzing index difficulty in Anates Application as follows:

1. Open Anates Application,
2. Click on "Jalankan Anates Pilihan Ganda"
3. Select "Buat File Baru" → Fill in the blank column with number of students & number of items that u will be analyzed → OK
4. Next, Put all data into table → Back to Main Menu → Select Menu "Olah Data" → Click "Tingkat Kesukaran" → Result

The Classification of Index Difficulty Level (Adopted from Thorndike & Hagen, 1977) as follows:

Table 2. Classification of Interpretation

FV	Interpretation
$0.00 < P \leq 0.30$	Difficult
$0.30 < P \leq 0.70$	Medium
$0.70 < P \leq 1.00$	Easy

Source: Theory of Thorndike & Hagen (1977)

Information:

FV = Index difficulty
 $0.00 < P \leq 0.30$ = Score less than 0.30
 $0.30 < P \leq 0.70$ = Score between 0.30 – 0.70
 $0.70 < P \leq 1.00$ = Score more than 0.70

FINDINGS

In this research, the researcher took the English Summative Test at the eleventh grade students of SMAN 9 Kota Jambi. The researcher took one class which consists 32 students. The

number of test is 40 multiple-choice tests. This research was began on 11th February 2019 to 19th February 2019.

1. Analysis of Difficulty Level

Sumadi Suryabrata (2006) says, more educate level of difficulty is a price index of difficult is the transformation Z, the proportion of correct answer to the raw score. Meanwhile, Heaton (1988) explains, the index difficulty of an item simply knows how easy or difficult the particular item proved in the test.

For manual calculation, the researcher used the Theory of Thorndike et al (1991), it explains that measuring the difficulty item interpretation, it states, An item with a p value of .0 or a p value of 1.0 does not contribute to measuring individual differences, and this is almost certain to be useless. Item difficulty has a profound effect on both the variability of test scores and the precision with which test scores discriminate among different groups of examinees. The level can be divided as follows:

Table 3. Level of Interpretation

FV	Interpretation
$0.00 < P \leq 0.30$	Difficult
$0.30 < P \leq 0.70$	Medium
$0.70 < P \leq 1.00$	Easy

Source: Theory of Thorndike & Hagen (1977)

Information:
 FV = Index difficulty
 $0.00 < P \leq 0.30$ = Score less than 0.30
 $0.30 < P \leq 0.70$ = Score between 0.30 – 0.70
 $0.70 < P \leq 1.00$ = Score more than 0.70

The result analysis of the calculations using the formulas already available, it can be seen in table 4 and its explanation, as follows

Table 4. Result of Difficulty Item

The number is correct	Number of item	Number of tests	Score	Interpretation
30	12, 22	2	0,93	Easy
29	24, 36	2	0,90	Easy
28	17	1	0,87	Easy
27	19	1	0,84	Easy
26	2	1	0,81	Easy
25	30	1	0,78	Easy
24	16, 28	2	0,75	Easy
23	9, 13, 14, 18, 21	5	0,71	Easy
22	23, 34, 37	3	0,68	Medium
21	5, 15, 25, 39	4	0,65	Medium
20	20, 33	2	0,62	Medium
19	1, 10, 40	3	0,59	Medium
18	38	1	0,56	Medium
17	7	1	0,53	Medium
16	11, 32	2	0,50	Medium
13	31, 35	2	0,40	Medium
10	4	1	0,31	Medium
9	8, 27	2	0,28	Difficult
6	29	1	0,18	Difficult
3	3	1	0,09	Difficult

0	6,26	2	0	Difficult
---	------	---	---	-----------

From the table above, it can be explained that :

- a) It found that 30 students answered correctly on the number 12 & 22. The score of interpretation on each number is 0,93 and categorized with easy interpretation.
- b) 29 students answered correctly on the number 24 & 36. The score of interpretation on each number is 0,90 and categorized with easy interpretation.
- c) 28 students answered correctly on the number 17. The score of interpretation on that number is 0,87 and categorized with easy interpretation.
- d) 27 students answered correctly on the number 19. The score of interpretation on that number is 0,84 and categorized with easy interpretation.
- e) 26 students answered correctly on the number 2. The score of interpretation on that number is 0,81 and categorized with easy interpretation.
- f) 25 students answered correctly on the number 30. The score of interpretation on that number is 0,78 and categorized with easy interpretation.
- g) 24 students answered correctly on the number 16 & 28. The score of interpretation on that number is 0,75 and categorized with easy interpretation.
- h) 23 students answered correctly on the number 9, 13, 14, 18, 21. The score of interpretation on each number is 0,71 and categorized with easy interpretation.
- i) 22 students answered correctly on the number 23, 34, 37. The score of interpretation on each number is 0,68 and categorized with medium interpretation.
- j) 21 students answered correctly on the number 5, 15, 25, 39. The score of interpretation on each number is 0,65 and categorized with medium interpretation.
- k) 20 students answered correctly on the number 20 & 33. The score of interpretation on each number is 0,62 and categorized with medium interpretation.
- l) 19 students answered correctly on the number 1, 10, 40. The score of interpretation on each number is 0,59 and categorized with medium interpretation.
- m) 18 students answered correctly on the number 38. The score of interpretation on that number is 0,56 and categorized with medium interpretation.
- n) 17 students answered correctly on the number 7. The score of interpretation on that number is 0,53 and categorized with medium interpretation.
- o) 16 students answered correctly on the number 11 & 32. The score of interpretation on each number is 0,50 and categorized with medium interpretation.
- p) 13 students answered correctly on the number 31 & 35. The score of interpretation on each number is 0,40 and categorized with medium interpretation.
- q) 10 students answered correctly on the number 4. The score of interpretation on that number is 0,31 and categorized with medium interpretation.
- r) 9 students answered correctly on the number 8 & 27. The score of interpretation on each number is 0,28 and categorized with difficult interpretation.
- s) 6 students answered correctly on the number 9. The score of interpretation on that number is 0,18 and categorized with difficult interpretation.
- t) 3 students answered correctly on the number 3. The score of interpretation on that number is

- 0,09 and categorized with difficult interpretation.
- u) none of students answered correctly on the number 6 & 26. The score of interpretation on each number is 0 and categorized with difficult interpretation.

Based on explanations above. It can be concluded that the difficult item was six items (15%), medium items was nineteen items (47,5%) and easy item was fifteen items (37,5%). Most of the items were categorized as medium interpretation, and a few in easy and difficult interpretation.

DISCUSSIONS

Based on the item analysis data of the difficulty level that the researcher gets, it can be seen that from 40 items. Firstly, the difficult item was six items (15%) with the test numbers are 3, 6, 8, 26, 27, 29. Secondly, medium item was nineteen items (47,5%) with the test numbers are 1, 4, 5, 7, 10, 11, 15, 20, 23, 25, 31, 32, 33, 34, 35, 37, 38, 39, 40. The last, the easy item was fifteen items (37,5%) with the test numbers are 2, 9, 12, 13, 14,16, 17, 18, 19, 21, 22, 24, 28, 30, 36. Most of the items were categorized as medium interpretation, and only a few in easy and difficult interpretation.

4.2.1 Analysis of Items

The researcher explains by details the result of some items which are categorized in easy, medium and difficult level. The explanations can be seen as follows:

Analysis on Item 8

Our classroom is very dirty. How about. . . . it together?

- Clean
- Cleans
- Cleaned
- Cleaning**
- To clean

Item number 8 shows the index of difficulty was in level 0,28 which mean

only a few of students are able to answer this question. **Discussion:** If you're asking someone their opinion on something or if they would like to do something. You must put $V_1 + ING$ after (how about. . . / what about. . .). The right answer is How about **Cleaning** it together.

Analysis on Item 26

Dear John,

On behalf of your brother. John. And all the people of Alberta. I would like to personally invite you to visit Alberta in 2016 to help mark the province's bicentennial year.

A number of activities have been planned in the province throughout the year. And now John would love to have your visit to help mark his important milestone in Alberta's history.

To help you plan your travel, details of some of the events can be found on our website at www.albertabicentennial.ca, and additional information can be found at travel Alberta's website at www.travelalberta.com. I am enclosing a special gift for you. So, when you do visit Alberta, you can celebrate our bicentennial year for learning more about our past at one of Alberta's museums. Historic sites or interpretive canters.

"Please come join us, we celebrate our past and imagine our future". (paragraph 4)

The sentences above means. . . .

- To do together and play together
- To combine together and celebrate it
- To join together and do something right
- To merge together and hope to rise the product
- Let's commemorate the bicentennial year together**

Item number 26 shows the index of difficulty was in level 0,00 which mean none of students can answer this question.

Discussion: This sentence means an invitation to attend an event. Because this letter contains an invitation to attend the Bicentennial year, the sentence **"Please come join us . . ."** is the sentence of an invitation to jointly celebrate the Bicentennial year. So, the right answer is **Let's commemorate the bicentennial year together**.

Analysis on Item 5

Dear Mr.Tandrio

We appreciate your insightful suggestions regarding our front office staff. I myself have observed how they do their daily jobs. Especially, when they have to welcome customers. I do agree that their service has deteriorated. We have to improve it soon. We will design a training programme to enhance their sense of hospitably. If the situation does not change. I will take serious action against them.

Please come to my office on Monday training. Let us discuss this situation further.

Sincerely yours,

Pamela Wilson

“We have to improve it soon” (par.1)
What does the underlined pronoun refer to?

- Service
- Situation
- Serious Action
- Sense of Hospitably
- A Training Programme

Item number 5 shows the index of medium was in level 0,65 which mean more than an half number of students are able to answer this question.

Discussion: This letter tells us about criticism for the staff service in a hospital. And the underlined pronoun refer to “We have to improve it soon” (par.1). it refers to staff service. So, the right answer is Service.

Analysis on Item 11

What about. . . . to the bus stop?

- Walking on
- Walking**
- To walk
- Walks

Item number 11 shows the index of medium was in level 0,50 which mean only a few of students are able to answer this question.

Discussion: If you're asking someone their opinion on something or if they would like to do something. You must put V₁ + ING after (how about. . . . /

what about. . . .). The right answer is What about **Walking** to the bus stop?

Analysis on Item 12

Putra : You look real bad. What's wrong with you?

David : I have got influenza for a week. I always study late at night.

Putra :

...

- Are you ok?
- Are you all right?
- You should not take a rest
- What seems to be the problem?
- You should go to a doctor and have a rest**

Item number 12 shows the index of easy was in level 0,93 which mean almost a whole students are able to answer this question.

Discussion: The sentence above is a brief conversation between Putra and David. The right sentence to continue the empty dialogue is **You should go to a doctor and have a rest**, because David indirectly asked Putra for advice when he told him over his influenza.

Analysis on Item 19

Donal : Would you like me to carry these books to your room?

Ferbria : No, thanks.

Donal : Let me bring your bag then,

Ferbria :

What does the woman probably answer?

- It is a handmade leather bag?
- My bag is black and big
- I'd go to teacher's room
- Thanks you very much**
- The bag belongs to me

from Donal to Ferbria. The right probably answer to fill the empty dialogue is **Thank you very much**, because Ferbria accepted Donal's offering help.

Based on the analysis' above, the researcher concludes that the level of English summative test for the eleventh grade at SMAN 9 Kota Jambi based on

the calculating in the appendix of the total as a whole of the difficulty index is 0,59 in level medium.

CONCLUSIONS AND SUGGESTIONS

Based on the result of this research, it can be concluded that:

1. Based on analysis and the discussion on chapter 4, it found that the total of index difficulty level of the test is 0,59. It means the summative test is categorized in medium level. So, the researcher concludes that the English summative test for the eleventh grade students at SMAN 9 Kota Jambi has not good enough yet. In other word, the test needs few revisions and enrichment to the test. the characteristics of a good test that should be used as a reference by an evaluator in compiling a measure test which includes validity, reliability, objectivity, practicality. By referring to good test criteria, the teacher is expected to be able to know the effectiveness and efficiency of the learning system. So that it can improve the quality of learning from time to time so that it can contribute to improving the quality of education.
2. There are significant results after analyzing student's score of the test as follows, 28 students got score under 75 minimum criteria of mastery learning (KKM in SMAN 9 Kota Jambi is 75), in percentage form approximately 87,5% of students face difficulties in answering the test. Then, only 4 students got score above 75 minimum criteria of mastery learning, in percentage form approximately 12,5% of students face easy in answering

the test. It concludes that, the test has quite difficulties.

Based on the conclusions above, the following points are suggested to the following parties, namely: 1) The teacher should analyze the test that it has been tested to the students to know whether the test good or not. Additionally, the test maker should be creative in giving the test items, based on the material which is taught especially. 2) The researcher suggests the students to practice more often of having test related to subject matter being tested with teacher's guidance. Therefore, students' learning ability in answering test can improve. 3) It is suggested to the future researchers who are interested in analyzing item of English summative test to conduct a further research at different school by using similar or different methods.

REFERENCES

- Ahmadi, M.R. & Gilakjani, A.P. (2011). *A Study of Factors Affecting EFL Learners' English Listening Comprehension and the Strategies for Improvement*. Journal of Language Teaching and Research. Vol. 2, No. 5, pp. 977-988, September 2011
- Alderson, J. C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press
- Allan, A. (1992). *Development and Validation of a Scale to Measure Test-wiseness in EFL/ESL Reading Test Takers*. Language Testing. 9. 101-122
- Amalia, Rika. (2010). *An Analysis of the English Summative Test Items in terms of Difficulty Level for the Second Year Students of MTs. Darul Ma'arif Jakarta*. Jakarta: Post Graduate of English Education Department Islamic Education and Teacher Training Faculty at University of Syarif Hidayatullah State Islamic Jakarta

- Anas, Sudijono. (2008). *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A Taxonomy for Learning, Teaching, & Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Complete Edition*. Newyork: Longman
- Arikunto, Suharsimi. (2013). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta
- Ary, D, *et al.* (2002). *Intoduction to Research in Education (Sixth Edition)*. Stamford : Wadsworth Group
- Başol, G., & Johanson, G. (2009). *Effectiveness of Frequent Testing Over Achievement : A Meta Analysis Study*. International Journal of Human Sciences, 6 (2), 99 – 121
- Brown, H. Douglas. (2004). *Language Assessment, Principles, and Classroom Practice*. New York : Addison Wesley Longman Inc.
- Brown, James Dean. (2006). *Testing in Language Programs*. New Jersey : Prentice Hall Regents
- Fraenkel, J. R. & Wallen, N. E. (2003). *How to Design and Evaluate Research in Education (5th Edition)*. New York: McGraw-Hill.
- Geiser, S., & Santelices, M. V. (2007). *Validity of high-school Grades in Predicting Students Succes Beyond The Freshman Year: High-school Record vs. Standardized Tests as Indicators of four-year College Outcomes*. (Research and Occasional Paper Series CSHE. 6.07). Berkeley, CA: University of California, Berkeley, Center for Studies in Higher Education.
- Gronlund, N. E. & Linn, R. L. (1990). *Measurement and Evaluation in Teaching (6th Edition)*. New York : Macmillan Publishing Company
- Hakuta, K. (1990). *Language and Cognition in Bilingual Children*. In A. Padilla, C. Valdez & H. Fairchild (Eds.), *Bilingual Education: Issues and Strategies*. (Pp. 47-59). Newbury Park, California: Sage Publications
- Hart, R., Casserly, M., Uzzell, R., Palacios, M., Corcoran, A., & Spurgeon, L. (2015). *Student Testing in America's Great City Schools : An Inventory and Preliminary Analysis*. Washington, DC : Council of the Great City Schools.
- Heaton, J.B. (1988). *Writing English Language Tests*. USA: Longman Inc.
- Netra, I. B. (1974). *Statistik Infrensial*. Surabaya : Usaha Nasional
- Kerlinger, Fred N. & Howard B. Lee. (2000). *Foundations of Behavioral Research (4th Edition)*. Florida: Harcourt Inc.
- Kumari, Amrita. & Bhattacharya, S.B, Prof. (2016). *Item Analysis of Diagnostic Test in English Language Skills of Secondary School Students*. Banaras Hindu University, Uttar Pradesh: IOSR Journal of Research & Method in Education (IOSR-JRME) e-ISSN: 2320-7388, p-ISSN: 2320-737X Volume 6, Issue 4 Ver. I (Jul. - Aug. 2016), PP 01-05.
- Larsari, V.N. (2011). *Learners' Communicative Competence in English as a Foreign Language*. Iran: Islamic Azad University
- Lohr, L. S. (1999). *Sampling: Design & Analysis*. Albany: Duxbury Press

- Marshall, J. C. & Hales, L. W (1972). *Essentials of Testing, Reading*, MA: Addison – Wesley, 45 – 67
- Maxwell, J. A. (1996). *Qualitative Research Design: An Intractive Approach London*. Applied Social Research Methods Series.
- Morrell, P.D. & Carroll, J.B., (2010). *Conducting Educational Research: A Primer for Teachers and Administrators*. Rotterdam: Sense Publishers
- Moss, C. M. (2013). *Research on Classroom Summative Assessment*. In J. H. McMillan (Ed.), *Handbook of Research on Classroom Assessment*. Los Angeles, CA: Sage.
- Patel, R. M. (2017). *Use of Item analysis to improve quality of Multiple-Choice Questions in Second Year Bachelor of Medicine and Bachelor of Surgery*. Pramukhswami Medical College, Karamsad: Journal of Education Technology in Health Sciences, January - April, 2017;4(1): 22-29.
- Permendikbud. *Tujuan dan Landasan Hukum Kurikulum 2013*. Accessed online on <https://www.kemdikbud.go.id/>.
- Polit, F. D & Beck, T. C (2003). *Nursing Research: Principles and Methods*. Philadelphia: Lippincott Williams & Wilkins
- Purwanto. (2009). *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar
- Richards, Jack. (2001). *Curriculum Development in Language Teaching*. Cambridge : Cambridge University Press
- Smith, H. W. (1975). *Strategies of Social Research: The Methodological Imagination*. Englewood Cliffs, NJ: Prentice Hal
- Sugiyono, Dr. (2010). *Metode Penelitian Pendidikan (Pendekatan Kuantitatif, Kualitatif, dan R & D)*. Bandung : Alfabeta
- Suryabrata, Sumadi. (2006). *Metode Penelitian*, Jakarta: PT. Raja Grafindo Persada
- Thompson, B. & Levitov, J. E. (1985). *Using Microcomputers to Score and Evaluate Test Items*. Collegiate Microcomputer, 3, 163 – 168
- Thorndike, E.L., & H.P. Hagen, (1977). *Measurement and Evaluation in Psychology and Education (Fourth Edition)*. New York: John Wiley
- Valette, R. (1977). *Modern Language Testing (2nd Edition)*. San Diego, CA: Harcourt Brace Jovanovic