THE CONTENT VALIDITTY OF THE SUMMATIVETEST OF IN THE SECOND SEMESTER OF THE FIRST YEAR STUDENTS
(A Case Study of SMAN 1 Cikarang Barat,- Bekasi)

Tita Puspitasari

Program Studi Pendidikan Bahasa Inggris

Fakultas Bahasa dan Seni, Universitas Indraprasta PGRI

Pos-el:*tita_ps@yahoo.com*

HP: 085213038230

ABSTRACT

This study aims to clarify the matter summative validitasisi created by a professional team for the first semester of the second year in SMAN 1 Cikarang, Bekasi. The number of items as many as 44 items, 40 multiple-choice and essay 6.Forty to about 4 questions for the reading and writing problems. The method in this study using a qualitative description that describes the suitability of matter summative SMAN 1 Cikarang, Bekasi with the syllabus and indicators. With a qualitative approach, researchers collected data by requesting a booklet summative English and in the school syllabus. In addition, the researchers also conducted interviews to a team of teachers who made a summative test questions. Furthermore, the researchers analyzed the items, whether the matter has valisitasisi summative. The results obtained from this study proves that the question summative English second semester of grade nine through percentage and analisisisi have validitasisi bad. In addition, about summative bahasaInggris not in accordance with the existing indicators in the syllabus. Problem grains in a summative test contained only two skills are measured. Two skills that are reading and writing. Based comprehension ideally weight to 25% listening skills, speaking skills 25%, 25% reading skills, and writing skills 25%. Thus, it can be concluded that the results of the study skills of reading reached 90.91% according to the syllabus of the test means can be said to be good. Meanwhile, for the listening skills pesrentasinya only 9, 01%. It shows that the test questions are weak or invalid.

Key words: validitasisi, summative tests, syllabi English

ABSTRAK

Peneliatian ini bertujuan untuk menjelaskan validitas isi pada soal sumatif yang dibuat oleh tim professional untuk semester ganjil tahun ke dua di SMAN 1 Cikarang, Bekasi. Jumlah butir soal sebanyak 44 butir soal, 40 pilihan ganda dan 6 esai.Empat puluh untuk soal reading dan 4 soal untuk soal writing.Metode dalam penelitian ini menggunakan menggunakan metode deskripsi kualitatif yakni mendeskripsikan kesesuaian soal sumatif SMAN 1 Cikarang, Bekasi dengan silabus dan indikator. Dengan pendekatan kualitatif, peneliti mengumpulkan data dengan meminta lembar soal sumatif bahasa Inggris dan silabus pada pihak sekolah. Selain itu, peneliti juga melakukan wawancara kepada tim guru yang membuat soal tes summative. Selanjutnya, peneliti manganalisis butir soal, apakah soal sumatif tersebut memiliki valisitas isi. Hasil yang diperoleh dari penelitian ini membuktikan bahwa soal sumatif bahasa Inggris semester dua tahun kelas sembilan melalui persentase dan analisis isi mempunyai validitas isi yang buruk. Selain iu, soal summative bahasa Inggris tidak sesuai dengan indikator yang ada di silabus. Soal butir dalam tes sumatif hanya berisi dua keterampilan yang diukur.Dua keterampilan itu adalah reading dan writing. Berdasarkan pemahanan, idealnya bobot untuk keterampilan menyimak 25%, keterampilan berbicara 25%, keterampilan membaca 25%, dan keterampilan menulis 25%. Dengan demikian, dapat disimpulkan bahwahasil penelitian keterampilan membaca mencapai 90,91 % sesuai dengan silabus berarti tes tersebut bisa dikatakan baik. Sementara itu, untuk keterampilan menyimak pesrentasinya hanya 9, 01 %.Hal itu menunjukkan bahwa tes soal tersebut lemah atau tidak valid.

Kata kunci: validitas isi, tes sumatif, silabus bahasa Inggris

INTRODUCTION

English is an international language used by many people in many countries in the world as a purpose of communication. It is generally a goal of either oral or written communication. People use English in order to make relationship among people in different countries in the world. English language becomes the first foreign language that should be taught to English students for every level of education in Indonesia. English is taught as a compulsory subject in elementary, junior and senior high school, and as a complementary subject in university.

The purpose of teaching English in Indonesia is to develop the communication skills especially in oral and written test skills (listening, speaking, reading, and writing). To reach the purpose the instructional activities, the teacher apply evaluation to measure how far the student understand about the material. Dickin and Germaib state that evaluation is an intrinsic part of teaching and learning. It is important for the teacher because it can provide a wealth of information to use for the future direction of classroom practice fort the planning of courses, and the management of learning task and students. In doing evaluation, especially in collecting the data, the teacher can use tool on instruments. They are test or non-test.

A test is a measuring tool of students' learning which is expected to be able for providing good information that it can be responsible. Thus, the teacher can know how far the students make progress that has been reached in learning process and to know whether teaching method is effective.

One of the forms to evaluate the student's ability is test. There are some types of test. They are placement test, achievement test, proficiency test and aptitude test. Test is which usually used by the teacher to know how far the students have mastered the lesson is the achievement test. The achievement test is intended to establish how successful individual students group of students of the courses themselves have been in achieving objectives of language courses

Teachers are those who know to construct a test to measure their students' achievement and it is not an easy job. Some teachers make the test carelessly. The test that is made accordingly can help the teacher increasing the teaching-learning process. High quality test can give the information about how we'll the students have comprehended the material, which have been taught by the teacher. By having a test, the improvement in the next lesson planning can be upgraded. So, teaching-learning process will be more effective without any overlapping. Evaluation can be done in form of test. This test can be a teacher-made test or standardized test. In the teacher-made test, the teacher who makes the test should know and master the principles and the steps that must be done in making the test. By this knowledge the teachers will get a clear figure about the general systematic framework of evaluation.

In other to measure accurately, the teachers should use a good test. It is not easy work for them to make it. There are some characteristics or requirement that must be fulfilled. Sudijono said the characteristic of a good test include validity, reliability, objectivity, and practicality.

Gronlund Norman E states that validity is the most important consideration in test evaluation. The concept refers to the appropriateness, meaning, and usefulness of specific inferences made from the score. The test validation is the process of accumulation evidence to support such inference the former types of validity (content, criterion related, and

construct) are simply considered to be convenient categories for accumulating evidence to support the validity of an interpretation.

DISCUSSION

Theoretical Review

Definition of Test

There are some definitions about test. Penny Ur (1996) said that tests are used as a mean to motivate students to learn or review specific material. It means that test is one motivation of students to learn or review material in their school. Fernandes (1984) states that a test as a systematic procedure for surveying a person's behavior and explaining it with the aid of a numerical scale or category system.

In addition, according Linn and Gronlund (1995) test is a particular type of assessment that typically consists of a set of question administered during a fixed period of time under reasonably, comparable conditions for all students.

Therefore, it can be concluded test is the particular types of assessment to reinforce learning and to motivate students by giving a task or a set of tasks. Through a test, the teacher do not only measure and motivate the students but also improve the lesson in teaching learning process. In order to make a proper decision, the teacher needs accurate data and to gain data, so a good instrument is needed.

The Type of Test

Teacher constructs a test not only to measure how far the testers' comprehension about the materials, but also to know in which part the testers get difficulty, to know the effectiveness of the teaching method that has been used in the class, etc. According to Wilmar Tinambunan there are two types of tests used in determining a person's ability: aptitude test and achievement test. An aptitude test is primarily designed to predict success in some future learning activity, whereas an achievement test is designed to indicate degree of success in some past learning activity. The major purpose of testing in the classroom is to measure student achievement as an indication of progress toward educational objectives set for the students. While according to Mary and Sake, there are four basic types of language tests: achievement tests, and proficiency tests, diagnostic tests, and aptitude tests.

Achievement Test

Achievement-type or attainment tests are widely employed in language teaching just as they are in most other subjects. They are used to measure the amount and degree of control of discrete language and cultural items and of integrated language skills acquired by the student within a specific period of instruction in a specific course.

Proficiency Test

Language proficiency tests are designed to measure control of language or cultural items and communication skills already present at the time of testing, irrespective of formal

training. Such tests are generally used for specific purposes; for example, to determine selection of students for a specific program, to ascertain appropriate placement levels in courses or in advanced training programs, and to judge the examinee's readiness to perform specific tasks in a work activity.

Diagnostic Test

Diagnostic language test seeks to identify the specific strengths and weaknesses of a foreign language student. It may contain the same types of test items which are used in the achievement or proficiency tests.

Aptitude Test

The aptitude test is conceived as a prognostic measure that indicates whether a student is likely to learn a second language readily. It is generally given before the student begins language study, and may be used to select students for a language course or to place students in section appropriate to their ability. Gronlund divided achievement test into four types of test based on each purpose; Summative test, Formative test, Diagnostic test, and Placement test.

Summative Test

Summative test, which designed to determine the extent to which the instructional objectives have been achieved, and the test also can be used to assign course grade for certifying student's mastery of intended learning outcomes.

Formative Test

Formative test is a kind of test which is administered by teacher to know how successful the learning and teaching process in the classroom in periodical meeting. Similar with this statement, Gronlund states that a formative test is given to monitoring learning progress during instruction.

Diagnostic Test

Diagnostic test is used for analyzing the difficulties of the students or the strength of the students in some aspect of language at the starting or during the implementation of instruction. According to Harmer, ―a diagnostic test is a test that is used to expose learners' difficulties during a course. Furthermore, Hughes said that ―Diagnostic test is used to identify learner's strength and weakness.

Placement test

Placement test is the test that is arranged for measuring students' ability inlanguage and give students major in an appropriate grade of educationalorganization. Moreover, Gronlund defines placement tests are ―pre-tests designedto measure (1) whether pupils possess the perquisite skills need to succeed in aunit or course or (2) to what extent pupils have already achieved the objectives ofthe planned instruction.Meanwhile, Hughes states ―placement tests areintended to provide information that will help to place students at the stage of theteaching program most appropriate to their abilities.

From all those types of the test above, it can be conclude that the tests are a problem –solving in the educational. The teacher or test maker can administer the test based on his or her objective. When the teacher want to determine the extent to which the instructional objectives have been achieved, she/ he can use the summative test, when teacher will diagnose the students' difficulties, she/he can use diagnostic test and when the teacher want to monitor learning progress during instruction, she /he can use the formative test, and when the teacher want to provide information that will help to place students at the stage of the teaching program most appropriate to their abilities as a problem-solving.

Validity

From the previous explanation, that one of characteristics of a good test is validity. According to Wilmar, validity refers to the extent to which the result of an evaluation procedure serve the particular uses for which they are intended if the results are to be used to describe pupil achievement . Thus, the validity of attest is the extent to which the test measure what is intended to measure

While Heaton states, the validity of a test is the extent to which it measure what it is supposed to measure and nothing else.The statements means the validity of a test must be considered in measurement in this case there must beseen whether the test used really measures what are supposed to measure, briefly. According to Charles Alderson, there are three types of validity; rational,empirical and construct validity. Rational validation depend on a logical analysisof the test's content to see whether the test contain a preventative

sample of therelevant language skill. Empirical validation, depend on and statistical evidence asto whether students' mark on the test similar to their scores on other tests, theirself - assessment or the teacher' rating of their ability. Construct validation referson what the test scores actually mean

Different statement comes from Hughes. He classifies validity into fourtypes: content validity, face validity, construct validity, and criterion-relatedvalidity The first content validity, Content validity is concerned with whether or not the content of the test is sufficiently and comprehensive for the test to be validmeasure of what it is supposed to measure. The objective of the course are basedon the curriculum made by government, thus to know whether the test has contentvalidity the tester should look the objective of the curriculum and the material oftest. The content validity is discussing more detail in the next subchapter. According to Norman E. Gronlund content validity is the process ofdetermining the extent to which a set of tasks provides a relevant andrepresentative sample of domain of tasks under consideration. The writer canconclude that content validity is the degree of whether the test content isrepresentative to the relevant subject matter, substance, or topics studied that havebeen designed to measure.

Second, Face validity. Face validity is a property of a test intended tomeasure something. It is the validity of a test at face value. In other words, a testcan be said to have face validity if it ―looks like‖ it is going to measure what it issupposed to measure. Face validity is more referring to the shape and appearance instruments.According to

Djamaludin Ancok in Arikunto, it is very important in measuringthe ability of individuals such as the measurement of honesty, intelligence, talentand skill.Moreover, Heaton defines ―If a test item looks right to other testers,teachers, moderators, and testers, it can be described as having at least facevalidity.

Therefore, can be assumed that face validity means the test appearance, which is readable, acceptable, and appropriate with whatsupposed to test. Third classifications of validity are construct validity. The difficulties of thestudents and the test must be qualified in construct validity. Arthur Hughes statesthat a test, part of test, or a testing technique is said to have construct validity if itcan be demonstrated that it measures just the ability or trait, which it is supposedto measure The word 'construct' above refers to any underlying ability or traits, Which is hypothesized in a theory of language ability. One might hypothesize, forexample, that the ability to read includes a number of sub- abilities, such as abilityto find out the main idea of a text. Construct validity is construction of a test itself from theoretical frameworkabout behavior derived from the empirical investigation that suggests positivelycorrelate with the scores of the other test designed to measure the samebehavior.

The last one classifications of validity are criterion validity. Criterion validity of a test involves the relationship or correlation between the test score and the scores on the some measure representing an identified criterion. For instance, the criterion measure may be another test. The correlation coefficient can be computed between the scores on the test

being validated and the scores on the criterion. A correlation coefficient so used is called a validity coefficient. There are two different types of criterion validity; concurrent validity and predictive validity. Concurrent validity applies if data on the two measures-tests (comparison of the test scores with some other measure for the same candidates taken at roughly the same time as the test) and criterion are collected at or about the same time. Predictive validity applies if there is an intervening period between the time of testing and the collection of data on the criterion.

Content Validity

We already know from previous explanation that one of classifies validity is content validity. Content validity is an important aspect in designing a test. The thing that must be concerned the most in preparing a test is content validity and then the sample of test items. It means that a test designer has to prepare content validity first before continuing to make test items. Content validity is also called curricular validity, because materials that will be tested are based on curriculum. As Grand explains that content validity concerned with how far an assessment is in line with the intended learning outcomes, standards, or objectives of an instructional unit. Therefore, If the test which is given to the students does not have a content validity, there will be consequences, there are: the students cannot demonstrate skills that they possess if they are not tested, and also that irrelevant items are presented that students will likely answer incorrectly only because the content was not taught. In addition, according to the book *Testing English as a Second Language* written by David P.

Harris, —If a test is design to measure mastery of a specific skill or the content of a particular course of study, we should expect the test to be based upon a careful analysis of the skill or outline of the course. According to *Evaluation of Instruction in Individually Guided Education*, there are two methods for demonstrating the content validity of a test. The first method is by listing all of the specific objectives that want to reach. After that they are matched with the items to see whether the items really conform to relevant objectives. The second method is by constructing a table to classify the items' content and taxonomic level, that is, student outcome required on the item. In addition, for making sure that a test has conformity between the subject matter with the learning outcomes or the test has good content validity, there are three steps which should be done. First, the test designer should identify the subject matter and outcomes of the students which is intended to measure. Then he/she should make a table of test specifications. Last, he/she should construct a test that conforms to the table. The process above is known as validation process. This content validation process based on language teaching method, material, given, and the goals. Furthermore, in establishing content validity in a test, the test designer should determine what she/he wants to be measured clearly. If the test does not have content validity, it will cause some problem. The first problem is the students cannot demonstrate their skills which they have but which are not tasted. The second problem is the students will not answer correctly the items because there is no relevancy between the items with the materials that have been taught.

Curriculum, KTSP and Syllabus

Curriculum

Curriculum means the basic standard or the objective which is used for guiding a success of teaching-learning process. As stated by Roger and Christopher that curriculum is as a written document or a plan which contains strategies in achieving desired goals. Besides, curriculum also contains scopes, sequences, and methods for conducting teaching-learning process. Moreover, ―a curriculum is the content, standards, or objectives from which schools hold students accountable. Others claim that a curriculum is the set of instructional strategies teachers plan to use. Curriculum has four aims based on Richards; they are ―to provide a clear definition of the purposes of a program, to provide guidelines for teachers, learners, and materials writers, to help provide a focus for instruction, to describe important and realizable changes in teaching.

Based on UU RI no 20 tahun 2003 stated that ―*kurikulum adalah seperangkat rencana dan pengaturan mengenai tujuan, isi, dan bahan pelajaranyang digunakan sebagai pedoman penyelenggaraan pembelajaran untukmencapai tujuan pendidikan itu* (curriculum is defined as all planed of learningincludes the objectives, content and materials used as a guide to achieve theobjective of the instruction of educational institution).

Therefore, it can be concluded hat curriculum is a written document which describes objectives, goals, scope and sequences, and systems as a guideline forimplementing the teaching-learning process.

KTSP

Team Pengembangan/ Advokasi Kurikulum DKI Jakarta gives the definition of KTSP "

> KTSP adalah kurikulum operasional yang disusun oleh dan dilaksanakan di masing-masing satuan pendidika. KTSP terdiri dari tujuan pendidikan tingkat satuan pendidikan, structural dan muatan kurikulum tingkat satuan pendidikan, kalender pendidikan, dan silabus".

The curriculum has experienced many changes such as 2004 curriculum based on competence, and *KTSP* (curriculum of educational level), its changes are conformed in development of times. Nowadays, *KTSP* is being used every school.

Syllabus

Syllabus is one of the curriculum developmental product and learning that it comprises material outline learning. It has main component inside. They are the purpose of curriculum, the learning material teaching-learning strategies, and the evaluation.

Tim pengembangan/advokat Kurikulum DKI Jakarta gives the definition of syllabus

> Silabus adalah rencanaya pempelajaran pada suatu dan atau kelompok mata pelajaran atau tema tertentuyang mencakup standar komeptens, kompetensidasar, materi pokok/ pembelajaran, kegiatan pembelajaran, indicator, penilaian, alokasi wkatu, dan sumber bahan/ alat belajar.

(Therefore, a syllabus is a plan or a reference that used by teachers in leading a teaching-learning process of a program and it includes eight elements, such as standard competence, minimum standard competence, materials, activities, indicator, scoring, time allocation, and media).

While Allison states that syllabus is an inventory of what is to be taught. The principle constitutes syllabus must be scientific; paying attention; and student's requirement systematically and also a learning material must be good and involve the scientific expert.

The component that the writer will be used in analyzing the content validity is the indicators from the syllabus, because it is the real form of minimum standard competence that is used for measuring student's behavior, skills, and knowledge. In addition, it has to be covered in teaching-learning process. The indicators of English subject for the second semester of the first year students of SMAN 1 Cikarang Barat.

Table 1.1

Reading Components Required by the School- Based Curriculum

No.	Standar Kompetensi	Kompetensi Dasar	Indikator
1.	**Membaca** 1.1.Memahami makna teks	1.1.1 Merespon makna	• Membaca nyaring

	fungsional pendek dan essai sederhana berbentuk narrative, descriptive, dan news item dalam konteks kehidupan sehari-hari dan untuk mengakses ilmu pengetahuan	teks fungsional pendek (misalnya pengumuman, iklan, undangan dll) resmi dan tak resmi secara akurat, lancer dan berterima yang menggunakan ragam bahasa tulis dalam konteks kehidupan sehari-hari	bermakna wacana ragam tulis yang dibahas dengan ucapan dan intonasi yang =- benar • Mengidentifikasi topic dari teks yang dibaca • Mengidentifikasi informasi tertentu
2.	**Menulis** 1.2 Mengungkapkan makna dalam teks tulis fungsional pendek dan essai sederhana berbentuk narrative, descriptive, dan news	1.2.1 Mengungkapkan makna dalam teks tulis fungsional pendek (misalnya pengumuman, iklan, undangan	• Menggunakan tata bahasa, kosa kata, tanda baca, ejaan, dan tata tulis dengan akurat • Menulis gagasan utama • Mengelaborasi

	item dalam konteks kehidupan sehari-hari	dll) resmi dan tak resmi secara akurat, lancer dan diterima yang menggunakan ragam bahasa tulis dalam konteks kehidupann sehari-hari	gagaasan utama • Membuat draft, merevisi, menyunting • Menghasilkan teks fungsional pendek

Based on the competence standard and basic competence, they were analyzed. The content of the test items were made by MGMP team at SMAN 1 Cikarang Barat.

METHODOLOGY

In this research basically aims at collecting data, analyzing the test items, weather each test items of summative test for the first grade students of senior high school represent the indicators as suggested its syllabus as an indicator of the content validity. The test is measured adaptable with the syllabus and indicators especially from reading and writing skill.

So based on the data and types of information needed of this research, the items used qualitative research. In addition, this study is categorized as descriptive analysis. It is described the conformity and inconformity of the Senior High School's summative test with the syllabus and indicators.

In analyzing the data, it is used descriptive analysis of the test itemsper number and the writer identified each test item related to the indicator of reading and writing. After analyzing the test items logically, the writer compared the data details from test items into the school syllabus then described the content validity of the English summative test qualitatively related to the reading and writing skill.

The writer used those comparative results as the source of quantitative approach by counting the conformity of the test items in percentage form through simple formula as follow:

P= *percentage*

F = *frequency*

$$P = \frac{F}{N} \text{ x } 100 \text{ \%}$$

N= *Number of sample*

The writer categorized the finding percentage of conformity test into the level criteria then revealed the criteria. There are four level criteria: based on Arikunto's opinion

Table 1.2

Achievement	Criteria
76% - 100 %	Good
56% - 75 %	Sufficient
40 % - 55 %	Less Bad
< 40 %	Bad

RESEARCH FINDING

Description of Data

In this research, the writer used English summative test at second semester or known as *Ulangan Semester Ganjil* at the first grade students of Senior High School, SMAN 1 Cikarang – Barat Bekasi. It was administrated on Friday in 2007. The time given for doing the test is 120 minutes. The test consists of 44 items. There are 40 multiple choices and 4 essay items. The test for reading is 40 items, 4 essay items for writing. To know if the English summative test has content validity or not, the writer compared the content of the test with the syllabus / indicator of KTSP. It has been checked each test item and given *checklist*(√) to record that are in line with the specification.

The following table describes the total frequency of the suitability of English summative test items with the indicators in the latest English syllabus. Based on the data of items analysis result, it can be seen as follow

Table 1.3

List of Question of the Learning Material Based on Syllabus

Reading Skill

Questions	Test		Total Item	Number
	Yes	No		
1. Do the texts contain types of text				
a. Narrative?	√		5	1,2,3,4,5,22,23,24,25,26
b. Descriptive?	√		5	6,7,8,9,10
c. Spoof/ Recount?	√		9	27,28,29,15,16,17,36,37
d. Report?	√		3	33,34,35
e. News item?	√		3	30,31,32
f. Procedure?		√	-	
Does the test contain that measure a short written functional text				
a. Announcement?	√			38,39,40
b. Advertisement?		√		
c. Invitation?		√		

Writing Skill

Questions	Test Yes	No	Total Item	Number
2. Do the texts contain types of text				
g. Narrative?	√		1	44 (a & b)
h. Descriptive?		√		
i. Spoof/ Recount?		√		
j. Report?		√		
k. News item?		√		
l. Procedure?	√		1	44 (a & b)
Does the test contain that measure a short written functional text				
d. Announcement?		√		
e. Advertisement?		√		
f. Invitation	√		1	43

Based on the specific Instructional Objective showed for reading skill as follow:

$$P = \frac{F}{N} \times 100\%$$ P= *percentage*

$$F = \textit{frequency}$$

$$N = \textit{Number of sample}$$

$$P = \frac{40}{44} \text{x } 100 \%$$

$$= 90, 91 \% \text{ (Reading Skill)}$$

$$P = \frac{4}{44} \text{x } 100 \%$$

$$= 9,09 \% \text{ (Writing Skill)}$$

CONCLUSION

Based on the writer's understanding generally, each skill must measure 25 % of Listening skill, 25 % of Speaking skill. 25 % of Reading skill, and 25 % of Writing skill. The item tests are consisted 100% for all skills. From the previous explanation, it can help to know whether the test items each skill is suitable or not.

From the percentage above, listening and Speaking did not have any percentage because it could not find out on the test for measuring listening skill. So the test items included to very poor test. It means that the test is not valid.

While reading skill got 90,91% for the suitability of the indicators in English syllabus. On reading skill, must be measured by around 25% of items. Based on the comparison between the indictors in English syllabus is reading skill included to be an excellent test based on Arikunto's opinion. It means that the test is valid. On writing skill,

the percentage got around 9, 09% in comparing the suitability of indicators in English syllabus could be categorized into poor test.

SUGGESTION

Based on the data analysis, there are some suggestions that might be useful in giving contribution to the teachers and the test makers. The suggestion as follow: The teacher has to understand the criteria of content validity in making the English summative test and the other test according to the syllabus components and of course they also have to develop the learning activities and the materials so that the students will learn what they have to be learn. Therefore, they will answer the test easily. The school has to make a preparation and give training for the teacher to make a preparation before constructing the test in order to construct a good and representative test. The students have to learn English more diligent in order to answer the test well.

BIBLIOGRAPHY

Airisan, Peter W. and Michael K. Russell. 2008. *Classroom Assessment and Evaluation*, Sixthedition. New York: McGRAW-Hill Companies Inc.

Alderson, Carolin Clapham and Dianne Wall, J. Charles. 1995.*Language Test Construction and Evaluation*.Cambridge : Cambridge   University Press.

Allison, Desmond.11999.*Language Testing and Evaluation (An Introductory Course)*. Kent Ridge Crescent: Singapore University Press.

Arikunto, Suharsini. 1992. *Prosedur Penelitian.* Jakarta: Rinekacipta.

Browers, Christopher Brumfit and Roger. 1991. *Applied Linguistice and English Language Teaching*. Modern English Publication and The British Council.

Brown, James Dean.1996.*Testing in Language Program*. New Jessey: Hall Regents.

Finnochiaro, Mary and Sydney Sako. 1983.*Foreign Language Testing*. New York: Regent Publishing Company, Inc.

Gareis, Christopher R and Leslie W. Grant. 2008. *Teacher-Made Assessments: How to Connect Curriculum, Instruction, and Student Learning*. New York: Eye OnEducation, Inc.

Genesee, F and A.U John. 1996. *Classroom Based Evaluation in Second Language Education*. New York: Cambridge University Press.

Groundlund, Norman E. 1985. *Measurement and Evaluation in Teaching*.USA: Mc. Millan Publishing Company.

Harmer, Jeremy. *The Practice of English Language Teaching*. (Essex: Person Education Limited), 4th Ed.

Harris, David P. *Testing English as a Second Language*. New York: McGraw-Hill Book Company.

Heaton, J.B. 1998.*Writing English Language Test*. London and New York: Longman.

Hughes, Arthur.1987.*Testing for English Teachers*. Cambridge: Cambridge University. Press.

Jay, Ronald and Mark E. Swerdlik. 1993. *Psychology Testing and Assessment: an introduction to Test and Measurement*. California: May Field PublishingCompany.

Nitko, Anthony J. 1983. *Educational Test and Measurement An Introduction*. NewYork: Harcourt Brace Jovanovic, Inc.

Pengembangan Silabusdan Rencana Pelaksanaan Pembelajaran dalam KTSP.2008. Jakarta: Direktorat Tenaga Kependidikan, Direktorat Jenderal, Peningkatan Mutu Pendidikdan Tenaga Kependidikan, dan Departemen Pendidikan Nasional.

Popham, W. James. 2003. *Educational Evaluation*. University of California. Los Angeles..

Sudijono, Anas. 1996. *EvaluasiPendidikan*. Jakarta: PT Raja Grafindo Persada.

Tinambunan, Wilmar. 1988. *Evaluation of Student Achievement*. Jakarta: Departemen Pendidikan dan Kebudayaan Direktorat Jendral Pendidikan Tinggi.

Wiersma, William. And Stephent G. Juts. *Educational Measurement and Testing*.