



**AN ANALYSIS OF CONTENT VALIDITY,  
READING COMPREHENSION ASPECTS, VOCABULARY,  
AND GRAMMAR COMPONENTS  
OF ENGLISH SUMMATIVE TEACHER-MADE TEST  
(The Case of Tenth Grade English Summative Teacher-Made Test  
of SMAN in Pekalongan in the Academic Year of 2015/2016)**

a final project  
submitted in partial fulfilment of the requirements  
for the degree of *Sarjana Pendidikan*  
in English

Nur Rohman  
2201413147

**ENGLISH DEPARTMENT  
FACULTY OF LANGUAGES AND ARTS  
UNIVERSITAS NEGERI SEMARANG  
2019**

## APPROVAL

This final project entitled *An Analysis of Content Validity, Reading Comprehension Aspects, Vocabulary, and Grammar Components of English Summative Teacher-Made Test (The Case of Tenth Grade English Summative Teacher-Made Test of SMAN in Pekalongan in the Academic Year of 2015/2016)* has been approved by the board of examiners and officially verified by the Dean of the English Department of Languages and Arts Faculty of Universitas Negeri Semarang on January 2019.

### Board of Examiners

**1. Chairperson**

Prof. Dr. Agus Nuryatin, M.Hum.

NIP. 196008031989011001

**2. Secretary**

Galuh Kirana Dwi Areni, S.S., M.Pd.

NIP. 197411042006042001

**3. First Examiner**

Puji Astuti, S.Pd., M.Pd., Ph.D.

NIP. 197806252008122001

**4. Second Examiner as Second Advisor**

Novia Trisanti, S.Pd, M.Pd.

NIP. 197611062005012002

**5. Third Examiner as First Advisor**

Drs. Amir Sisbiyanto, M.Hum.

NIP. 195407281983031002



Approved by

Dean of Faculty of Languages and Arts

Prof. Dr. Muhammad Jazuli, M. Hum.

NIP. 196107041988031003

## DECLARATION OF ORIGINALITY

I am Nur Rohman declare that this final project entitle *An Analysis of Content Validity, Reading Comprehension Aspects, Vocabulary, and Grammar Components of English Summative Teacher-Made Test (The Case of Tenth Grade English Summative Teacher-Made Test of SMAN in Pekalongan in the Academic Year of 2015/2016)* is my own work and has not been submitted in any work form for another degree or diploma at any university or other institute of tertiary education. Information derived from the published and unpublished works of others has been acknowledged in the text and a list of references is attached in the bibliography.

Semarang, January 2019



Nur Rohman  
2201413147

## MOTTO AND DEDICATION

*Live as you will but know that one day you are going to die.*

*Love whom you will but know that one day you will be separated from that person.*

*Do as you will and know that you will be compensated and rewarded accordingly.*

*Know that nobility of the believer is one standing up in prayer at night. And as for one's dignity, it's one being independent of people.*

*Jibreel AS*

This final project is dedicated to:

*My beloved parents and family, teachers and friends.*

## ACKNOWLEDGEMENT

In the name of Allah Swt., the Most gracious Most merciful. Complete blessings and salutations are sent upon Muhammad SAW. All praises are indeed due to Allah the Almighty, the Lord of the world and the master of the day after, for the blessing, graciousness, power, and ways of ease given to complete this final project entitled *An Analysis of Content Validity, Reading Comprehension Aspects, Vocabulary, and Grammar Components of English Summative Teacher-Made Test (The Case of Tenth Grade English Summative Teacher-Made Test of SMAN in Pekalongan in the Academic Year of 2015/2016)*. This is submitted as the final requirement in accomplishing undergraduate degree at English Education Study Program, English Department, Faculty of Literatures and Arts, Universitas Negeri Semarang.

In this opportunity, I would like to express my gratitude and appreciation to Drs. Amir Sisbiyanto, M.Hum., as my first advisor and Novia Trisanti, S.Pd., M.Pd. as my second advisor who have spent their valuable time to guide and give excellent suggestions, corrections, and invaluable criticism to me in preparing and completing this final project. I would also like to give my deep appreciation to all the lecturers of the English Department who have taught and guided me patiently during the years of my study.

My gratitude goes as well to my beloved parents and family who have prayed endlessly, encouraged sincerely, and reminded me to finish this final project. I also extend my deep appreciation to Hinsha Putri, S.Pd., M.Pd. and Enoni Alifiani, S.Pd., M.Pd. as my mentor for this final project.

To all my best friends in English Education Program, whom I have spent precious and memorable time for recent years with. Thank you for keeping me motivated and helping me during the process of completing this final project. Last, but not least, my very sincere thanks to anyone who are not mentioned personally here who teaches and inspires me a lot of things.

Finally, I hope that this final project may be useful for all the readers either for English teachers or for English students, and for anyone to improve educational quality.

Semarang, January 2019

A handwritten signature in black ink, appearing to read 'Nur Rohman', with a long horizontal stroke extending to the right.

Nur Rohman  
2201413147

## ABSTRACT

Rohman, Nur. *An Analysis of Content Validity, Reading Comprehension Aspects, Vocabulary, and Grammar Components of English Summative Teacher-Made Test (The Case of Tenth Grade English Summative Teacher-Made Test of SMAN in Pekalongan in the Academic Year of 2015/2016)*. Final Project. English Department. Universitas Negeri Semarang. First Advisor: Drs. Amir Sisbiyanto, M. Hum. Second Advisor: Novia Trisanti, S.Pd., M. Pd.

Key words: content validity, reading comprehension, vocabulary components, grammar components, summative teacher made-test.

This study investigates on how the content validity, reading comprehension aspects, vocabulary, and grammar components are realized in the English summative test prepared by the teacher for the second-semester of the tenth-grade students of SMAN in Pekalongan in the academic year 2015/2016. It is focused on the relevance of the summative teacher-made test items to the syllabus and the realization of the reading comprehension, vocabulary, and also grammar components in the test items. The total of 30 test items of the English summative test for second semester of tenth grade of SMAN in Pekalongan in the academic year 2015/2016 became the object of this study. The design of the research used in this study is a qualitative research design based on the process of qualitative data analysis according to Miles (1984), which qualitative-descriptive method is used for explaining the data analyzed and the descriptive data analysis result is presented. From the result of the analysis, it was found that the summative test only achieved an insufficient rate of content validity, in which the percentages between the relevant and the irrelevant items are split evenly 50-50 percents. It was also found that there were five materials that were relevant to the tenth-grade second semester syllabus of KTSP curriculum, and six irrelevant materials. Furthermore, the results of the reading comprehension, vocabulary, and also grammar components analysis showed that 23 test items (77%) were on reading comprehension, 7 test items (23%) were on grammar, and none of the test items were on vocabulary. In addition, there were also some errors in the construction of the test items and can be found in the text of the test items, in the questions or instructions of the items, and also in the options the items provided for the answers. Dealing with the result of the study, it is suggested that the test constructors pay attention to the syllabus and the curriculum in designing test items, so that the test items are relevant to the learning objectives and possess a high-level of content validity. Next, it is also better for the test designers to balance the main focus for the composition of the all test items.

## TABLE OF CONTENT

APPROVAL.....	ii
DECLARATION OF ORIGINALITY .....	iii
MOTTO AND DEDICATION .....	iv
ACKNOWLEDGEMENT .....	v
ABSTRACT.....	vii
TABLE OF CONTENT .....	viii
LIST OF TABLES .....	x
LIST OF FIGURES .....	xi
LIST OF APPENDICES .....	xii
CHAPTER I INTRODUCTION .....	1
1.1 Background of the Study .....	1
1.2 Reasons for Choosing the Topic .....	4
1.3 Research Questions .....	4
1.4 Objectives of the Study .....	5
1.5 Significance of the Study .....	5
1.6 Scope of the Study .....	6
1.7 Outline of the Final Project .....	7
CHAPTER II REVIEW OF RELATED LITERATURE .....	8
2.1 Previous Studies .....	8
2.2 Related Literature .....	26
2.2.1 Notions of Curriculum.....	26
2.2.1.1 The Main Competence of English for Senior High School.....	27
2.2.2 Notions of Syllabus .....	28
2.2.3 Language Assesment.....	30
2.2.4 Notions of Test .....	31
2.2.5 Types of Test .....	32
2.2.5.1 Language Proficiency Test .....	32
2.2.5.2 Language Aptitude Test .....	33
2.2.5.3 Achievement Test.....	35
2.2.6 Types of Test Items .....	36
2.2.6.1 The Advantages of Multiple-Choice Items .....	37
2.2.7 Characteristics of Good Test .....	38
2.2.8 Notions of Language Knowledge and Language Components .....	39
2.2.9 Language Testing .....	41
2.2.9.1 Language Components Testing .....	42



2.2.9.2 Language Skills Testing .....	44
2.2.10 Notions of Validity .....	45
2.2.11 Types of Validity .....	46
2.2.12 Content Validity .....	47
2.3 Framework of the Analysis .....	49
CHAPTER III METHODS OF INVESTIGATION .....	53
3.1 Research Design.....	53
3.2 Object of the Study.....	54
3.3 Types of Data .....	55
3.4 Unit of Analysis .....	55
3.5 Instruments for Collecting the Data .....	56
3.6 Procedure of Collecting the Data .....	63
3.7 Procedure of Analyzing the Data .....	64
3.8 Technique of Reporting Data .....	66
CHAPTER IV FINDINGS AND DISCUSSIONS .....	68
4.1 Findings.....	68
4.1.1 The Relevance of the Test Items Materials to the English Syllabus of KTSP Curriculum .....	68
4.1.2 The Reading Comprehension Aspects, Vocabulary, and Grammar Components .....	70
4.2 Discussions.....	71
4.2.1 The Analysis of the Valid Test Items.....	72
4.2.2 The Analysis of the Invalid Test Items .....	78
4.2.3 The Additional Comments of the Test Items Construction .....	88
4.2.4 Overall Interpretation of the Research Results .....	95
CHAPTER V CONCLUSIONS AND SUGGESTIONS.....	99
5.1 Conclusions.....	99
5.2 Suggestions .....	100
5.3 Limitations of the Study.....	101
BIBLIOGRAPHY .....	102
APPENDICES .....	107

## **LIST OF TABLES**

Table 2.1 List of English Materials in the X Grade of Senior High School.....	29
Table 3.1 The Relevance of the Materials to the KTSP Curriculum.....	56
Table 3.2 The Relevance of the Test Items Materials to the English Syllabus.....	59
Table 3.3 The Main Focus of the Test Items in the English Syllabus.....	61
Table 3.4 Criteria of Test Validity.....	66
Table 4.1 The Relevance of the Materials to the KTSP Curriculum.....	69
Table 4.2 The Items Relevant to the Syllabus.....	72
Table 4.3 The Items Irrelevant to the Syllabus.....	79
Table 4.4 The Analysis Results of the Content Validity, Reading Comprehension Aspects, Vocabulary, and Grammar Components of the Test Items.....	85
Table 4.5 The Corrections for the Construction Errors of the Test Items.....	89

## **LIST OF FIGURES**

Figure 2.1 Theoretical Framework of the Research.....	52
Figure 3.1 The Research Design.....	54
Figure 4.1 The Percentages of the Relevance of English Summative Test Items to the Tenth-grade English Syllabus of KTSP Curriculum for the Second Semester.....	86
Figure 4.2 The Percentages of the Reading Comprehension Aspects, Vocabulary, and Grammar Components Realized in the English Summative Test Items.....	88

## **LIST OF APPENDICES**

Appendix 1 Syllabus.....	108
Appendix 2 English Summative Teacher-Made Test.....	120
Appendix 3 Content Validity Analysis.....	129
Appendix 4 The Main Focus of the Test Items in the English Syllabus.....	130
Appendix 5 Complete Instrument of Analysis.....	132
Appendix 6 Letter of Testimony.....	141

# **CHAPTER I**

## **INTRODUCTION**

This chapter consists of seven subchapters. The first subchapter is background of the study which explains the brief overview of the concerned field related to this research. The second one is reasons for choosing the topic which explains the main situation underlies this research that has been examined. The third one is research questions which states the questions that can be answered by the research. Next is objectives of the study which explains the purpose of the research. After that is significance of the study which explains about the advantages provided by the research. Next is scope of the study which explains about the focus and scope of the research. Lastly, there is outline of the report as the seventh subchapter which describes the presentation of the research.

### **1.1 Background of the Study**

The significance of English communication in society is undoubtedly very apparent since English has become the most widely used international languages in the world. It is one reason why the Indonesian government has chosen to teach English as a necessary subject in the Junior/Senior High School. In Indonesian educational programs, English is presented and considered as the first foreign language to be taught in Junior/Senior High School.

In an educational process, students are expected to experience changes. Based on this view, we expect that every program course or unit of a subject will have the capacity to bring significant changes for the students. It is necessary for teachers to conduct a test or an examination as one of the evaluation instruments to find out whether the expected changes have taken place or not. It contributes directly to the teaching-learning process used in the classroom instruction, and it is useful in programmed instruction, curriculum development, marking, guidance and, counseling, school administration and research (Gronlund, 1981, p.7). Although English is taught through various approaches, in Indonesia, the outcome is still unacceptable.

Evaluation is one of the activities which is an essential part of every educational institution. It is one of the activities to stimulate the students, so they will learn and master the materials which have been taught. Indonesian government provides curriculum as guidance for teaching English. The curriculum which is used currently for teaching English is "Curriculum 2013". Even though many schools have already applied this new curriculum, there are also few schools that still apply the old KTSP (*Kurikulum Tingkat Satuan Pendidikan* or School-based Curriculum). This KTSP is still used in the certain SMAN in Pekalongan for 2015/2016 academic year. The curriculum has been intended to empower the understudies to finish the learning system, to study and read independently,

to build their aptitudes, and to represent considerable authority in many parts of English of their decision.

The KTSP divides the academic year into two semesters. At the end of each semester, there is a test to measure the level of accomplishment of the learning program during a semester. In the second semester, there is a summative test. It is usually administered at the end of a marking period and measures the total of material covered. Summative evaluation is directed toward a much more general assessment of the degree to which the larger outcomes have been attained over the entire or some substantial part of it (Bloom et al, 1981, p.71).

Furthermore, the reason for conducting this study is based on the lack of mastery from the teacher in constructing a good test based on the objectives of the syllabus and curriculum according to many previous studies. In addition, this is the last summative test in which the KTSP curriculum is applied and would be changed with the new Curriculum 2013 for the next test. If there is still a problem with the implementation of the old curriculum in term of validity of the test, it will be almost certain that the teacher will face many difficulties in implementing the new curriculum. Moreover, the reason of the researcher to focus on the validity, reading comprehension aspects, grammar, and vocabulary is that based on previous studies which have been reviewed, none were analyzing all of those aspects in one integrated study. Therefore, the researcher is suggested to conduct

the study on those objects. Further elaboration will be presented in the following chapter.

## **1.2 Reasons for Choosing the Topic**

The reasons for choosing this topic is that the summative test which administered to the second semester of the tenth-grade students of SMAN in Pekalongan in the academic year 2015/2016 has not been proved in terms of its content validity. Furthermore, according to the KTSP, the evaluation of the teaching-learning process is carried out at the end of each semester. If test constructors do not pay attention to the selection of the items and the main focus assessed in the test items regarding the reading comprehension aspects, vocabulary and also grammar components, the validity of each item will be less guaranteed. For this reason, every test constructor must be careful in constructing the test items so that the result will meet the desired goal.

## **1.3 Research Questions**

Through this study, the researcher would like to find the answer to the following question:

- (1) How is the content validity achieved in the English summative test of the tenth-grade students of SMAN in Pekalongan in the academic year 2015/2016?
- (2) How are the reading comprehension aspects, vocabulary, and grammar components realized in the test according to the KTSP curriculum?



#### **1.4 Objectives of the Study**

The general objective of this study is to obtain an objective description of the English summative test for the second semester of the tenth-grade students of SMAN in Pekalongan in the academic year 2015/2016. The objectives are then specified into the following aims:

- (1) To figure out the content validity achieved in the English summative test of the tenth-grade students of SMAN in Pekalongan in the academic year 2015/2016.
- (2) To figure out the reading comprehension aspects, vocabulary, and grammar components realized in the test according to the KTSP curriculum.

#### **1.5 Significance of the Study**

By doing this research, the writer hopes that the result of this study will hopefully provide some advantages as follows:

Theoretically, this study is expected to provide information about the content validity of the English summative teacher-made test so that the teachers can have an overview of the content validity and the English language component that will help them in constructing the English summative test and apply it in their teaching-learning and the assessment process respectively.

Practically, teachers can construct the assessment after overviewing this study and able to implement the aspect of content validity and the realization of the language components into the test correctly. The school also makes

the right decision in assessment to support the teaching-learning process, especially the teaching of English, in order to get a better output and better result of the test assessment and administration.

Pedagogically, the study gives a point of view about the content validity of the English summative teacher-made test and additional information in developing a system in assessment and pedagogic field. The information would be a reference for the education stakeholders to be able to develop a curriculum that meets the needs of the field and perform the evaluation at a more advanced level by taking into account the validity of the assessment.

### **1.6 Scope of the Study**

The writer would like to put forward the following main focus of the study:

As stated before, there are two summative tests in one year for students of SMA. In this study, the writer analyzed only the second semester teacher-made English summative test for the second-semester students of the tenth grade of SMAN in Pekalongan in the academic year 2015/2016.

Generally, the summative test consists of number of items in the form of multiple-choice items and some essay-type tasks in the form of listening and written test. However, the writer wants to analyze the written test of the multiple-choice items and essays only. The reason is that the written test items will be more apparent if there is any error. Thus the assessor will assess the items objectively.

In addition, by using this type of test item, the inaccurate part of the test in terms of construction, spelling mistakes and structure of the test could be practically determined.

### **1.7 Outline of the Report**

The result of this study are presented in five chapters. Chapter I presents the introduction, which consists of the background of the study, reasons for choosing the topic, statements of the problem, objectives of the study, significance of the study, scope of the study, and also the outline of the report. Chapter II discusses the review of related literature supporting the research including the review of previous studies, theoretical studies and theoretical background of the study dealing with the content validity and language components. Chapter III explains the research method, which discusses the design of the research, object of the study, source of data, instrument of the study, technique of collecting data, procedure of collecting data, and data analyzing. Chapter IV presents the data analysis and the discussion of the result, which discusses the relevance analysis of the test to the curriculum, and language components realizations. Chapter V presents the conclusions of the investigation together with the suggestions for the improvement of creating the English summative teacher-made test.

## **CHAPTER II**

### **REVIEW OF RELATED LITERATURE**

This chapter consists of three subchapters (Previous Studies, Related Literature, and Framework Analysis). The first subchapter is previous studies which explains about studies related to this research. The second one is related literature which explains about the main idea underlies this research, i.e curriculum, syllabus, assessment, test, types of test, and types of test items, language knowledge and skills, and also language components, English language testing, validity, types of validity, and content validity. Lastly, there is framework of the analysis as the third subchapter which describes the starting point and basic guidance for the implementation of the research.

#### **2.1 Previous Studies**

This subchapter presents several previous studies regarding the content validity of test, reading comprehension, vocabulary and grammar components of the English language components. Those studies provide some insights for the researcher in conducting this research in analyzing the reading comprehension aspects, vocabulary and grammar components in the summative test.

There are numerous studies have been conducted relating to the content validity. In 2013, Ozer, Fitzgerald, and Sulbaran held a research which investigated the reliability and content validity of an English as a

Foreign Language (EFL) grade-level test for Turkish primary grade students. The purpose of their study was to evaluate the quality of an EFL test for 3rd graders and to provide empirical evidence related to its use as a standardized test. The quality of this EFL test was examined using commonly applied techniques to establish validity and reliability.

As a result, they found that the EFL test results seemed to suggest a high degree of internal consistency for subsection items for having moderate to high correlations between different parts of the test. However, the Content Validity Index (CVI) of the test as a whole was considered to be low. Consequently, revision of the items (e.g., re-wording of items; changing pictures; edits) was recommended.

In 2002, a research about validation of the test of English conversation proficiency was also conducted by Moritoshi. This study validated one such test, the Test of English Conversation Proficiency (TECP), designed in-house for use at a Japanese university. It examined various aspects of the TECP's validity, reliability and 'usefulness' using students' end-of-semester tests as the raw data source. The evaluation found that the TECP has good-high validity.

Another study about content validity has also been done by Palacio, Gaviria, and Brown in 2016. The topic of their study is about the alignment of English language testing with the curriculum. It is a study of an item-by-item evaluation of a series of English exams for validity and reliability in an effort to guarantee the quality of the process of test design. The purposes of

the study are to examine how well the midterm and final examinations for Courses 5-8 in the *Idiomas-subchapters* English Language Assessment Program for Adults are functioning and suggest ways to improve individual items as well as the tests as a whole. The study included descriptive statistics, item analysis, correlational analyses, reliability estimates, and validity analyses. The results of the study show that the new tests are an excellent addition to the program and an improvement over traditional tests for having a good-rate of content validity.

Another study about content validity is conducted by Delgado (2012). The purpose of this study was to show how to conduct a study aimed at obtaining content validity evidence in the test construction/adaptation process. This study presented the content validity analysis of the Spanish adaptation of the State-Trait Cheerfulness Inventory trait form (STCI-T) and used an applied perspective. The study described the most important criteria to consider in the selection of experts, the procedure recommended to obtain judgments, the material to administer, aspects of items to assess, and the type of analyses that should be conducted. The result of the study showed a good content validity achievement for the Spanish version of the STCI-T.

In 2014, Sireci conducted a study about the logic and theory underlying evidence of the traditional and modern methods for gathering and analyzing content validity data. A comprehensive review of the literature and standard was presented. The result indicated that for education tests and other assessment targeting knowledge and skill possessed by

examinees, validity evidence based on test content was necessary for building a validity argument to support the use of a test for particular purpose. The conclusion was by following the methods described in this article, practitioners have a wide arsenal of tools available for determining how well the content of an assessment was congruent with and appropriate for the specific testing purposes.

There was also a study about content validity conducted by Lei (2015). It investigated the content validity of teacher-made assessment and teacher understanding of table of specification in Chinese Elementary Schools in Johor. A questionnaire with 10 items was distributed to 30 teachers in order to get the data on table of specification. Items 1 to 4 tested teacher's understanding of the table of specification while items 5-10 examined the content validity of teacher-made assessment. The result of the study showed that the teachers had a low understanding of the table of specification and the majority of them never attended courses concerning in table of specification. Another finding showed that teacher-made assessment was valid in terms of content validity. However, most of the teacher had lack basic knowledge in designing a standard table of specification.

Furthermore, a study was conducted by Masuwai (2016). The study was aimed to examine the reliability and validity of the Teaching and Learning Guiding Principles Instrument administered in the Malaysian teacher education programme institutions. Nine experts were invited to

review the instrument. All members of the panel were academicians who had worked in the field of education as educators/authors/researchers. The content validity was examined based on the magnitude of the Content Validity Index (CVI) values as related to the degree of agreement among the panelists. After they analyzed the instrument which consists of 67 items, all of the items were agreed by the panelists on the relevance to their thematic domains.

The next research of content validity was conducted by Gholamreza Jandaghi in 2010. The purpose of this research was to decide high school teachers' skill rate in designing exam questions in physics subject. They took 364 samples to be assessed. The result showed that the designed exams have suitable coefficients of validity and reliability. The level of difficulty of exams was high. No significant relationship was found between male and female teachers in terms of the coefficient of validity and reliability but a significant difference between the difficulty level in male and female teachers was found. It means that female teachers had designed more difficult questions.

A research of content validity was also done by Kahsay in 2014. The research aimed to assess whether or not the Ethiopian General Secondary Education Certificate (EGSEC) English exams administered by the National Educational Assessment and Examinations Agency fairly represent the content coverage and adequate sampling of the objectives stated in the syllabus. The mixed method research design used to attain the purpose of



this research where the qualitative data analysis technique and quantitative approach were used together. The result of the study showed that the degree of relationship and strength of association between the contents of the sample SATs and textbooks were divergent and weak in the strength of association.

In Indonesia, there were also several studies which had ever been conducted about content validity analysis of a test. First is by Prasetyo (2015). His object of the research was the test items of English summative test of tenth grade MA Al Ma'arif Kediri, which consists of 50 items. This research belongs to qualitative approach and uses content analysis as the design of the research. The result of this study stated the content validity of the test item of English final examination for the tenth-grade student that was used by MA Al Ma'arif Kediri in term of test item appropriate or not with standard competence are 94% valid. It means the test items are good.

Second, was conducted by Nofiyanti (2011). This research was aimed at analyzing the English summative test validity used by SMP N 87 Jakarta for even semester of first grade students, it consists of (1) the English material displayed within the English summative test used at SMP N 87 Jakarta Academic year 2009/2010, (2) the English material covered by English summative test appropriate with the content standard of the latest syllabus suggested, and (3) the validity of English summative test at SMP N 87 Jakarta. The findings of the research stated that the content validity of English summative test at SMP N 87 Jakarta for the even semester of

second-grade student reached good content validity. In conclusion, the researcher has already proved that from 50 questions altogether, the test has only 5 invalid items.

The third research was held by Kristiana (2014). The objective of this study is to describe whether each test item of the summative test for second-grade students of SMPTPP Daarul Amanah represented the indicators as suggested in the syllabus as an indicator of the content validity. The total number of the test items are 40 multiple choice and 6 essay items. The method of this research is qualitative research in which the writer collected the data by asking the English summative paper test and the syllabus to the school. Then the writer analyzed it whether the test has content validity. The result of the research showed that the test items of the English summative test have bad validity. Based on the percentage, the writer found that the English summative test was 58.7% valid in terms of its conformity to the indicators.

The fourth research was held by Fauzi (2011). The purpose of this study is to find empirical evidence if the test items of English summative test for the second-grade students of MTs. Salafiyah Bedahan have good content validity and in line with the English Syllabus. The object of this research is the test items of English summative test of second grade MTs. Salafiyah Bedahan Kota Bekasi, which consists of 45 items. The researcher used the Comparative Analytic Method to observe and analyze the English test items by comparing the content of the test with the English syllabus and

the English curriculum for the second grade of junior high school. The findings of this study proved that the English test item of MTs. Salafiyah Bedahan for the second-grade students is invalid because the items that appropriate with the recommended indicators in reading and writing skill in the English syllabus consist of 11 indicators and they cover only 48%. Therefore, the English summative test at even semester for the second grade of MTs. Salafiyah Bedahan Kota Depok is considered as an invalid test because of having bad content validity.

The fifth research was conducted by Siregar (2014). He analyzed the content validity of English summative test for eleventh-grade students in SMA Islam Nurul Amanah East Java academic year 2013/ 2014. The test consists of twenty-five multiple choices and five essay tests. To get the data, he took the summative test given and the syllabus of the even semester. After analyzing the writer concluded that the test designer seemed to design the test without considering the principles of constructing a good test and not following curriculum 2013. Finally, the overall conclusion of all analysis in this research showed that the test designed for students for SMA Islam Nurul Amanah Bangkalan, East Java was invalid.

A study about content validity is also conducted by Putri (2009). The study analyzed the content validity of teacher-made English final second semester test for the year eleven students of SMAN 1 Ambarawa in the academic year of 2008/2009 based on the representativeness of content standard. The test papers consist of 50 items in the form of multiple choices.

The result of the analysis showed that the questions of the test are related to the 2006 Curriculum, but the topics of the questions are not related to the students' study program. In this final test, it is clear that the test is not valid and need some revisions. Finally, the writer made a conclusion that the topic in the test should be related to the students' background knowledge or students' study program though the curriculum mentions the same standard competence for students of science and social. So, the content or the topic of the test items should be different though the standard competence stated in the curriculum is same. In general, the writer suggested to the teacher as the test makers to pay attention to the writing of multiple choice items and the characteristics of a good test.

The next research about content validity is conducted by Nawati (2009). This study aimed to find the level of the content validity of the English national writing test and to give the suggested test items that should meet the Graduate Competence Standard. To achieve the objective of the study, the writer analyzed the content validity of the English national writing test items and made the suggested test items based on the Graduate Competence Standard, and then found the reliabilities. The sources of the data were the English National Test for Junior High School in the Academic Year of 2007/ 2008 and the scores from 42 students of SMP 1 Semarang and 38 students of SMP Nusa Bhakti Semarang in the Academic Year of 2008/ 2009. In drawing conclusion about the content validity of the English national writing test for Junior High School based on the Graduate

Competence Standard of KTSP Curriculum, the analysis of the data is compared to the Graduate Competence Standard of KTSP Curriculum. After analyzing the data the writer found that there are no items which indicate that the student must possess the ability of writing so that the items do not meet the Graduate Competence Standard. Although there are indicators for writing, those are restricted only in the organizations and the arrangements of the texts. It can be concluded that the content validity of the English national writing test is not congruent with the Graduate Competence Standard.

In 2009, Haris Rizqi Arifin analyzed the content validity of test items in English final examination for the sixth grade students of elementary school in south Semarang Regency in the academic year 2008/2009. The data used in this study were taken from the test papers and students' answer sheets. The test papers consist of 50 items in the form of multiple choices. From the result of the analysis, the mean of validity level is 0.3250. Based on the result, the writer suggested to the teacher as the test makers to prepare test items far in advance before they give it to the students. They should also pay attention to the writing of multiple choice items and the characteristic of a good language. Finally, the writer draws a conclusion that the items in the English final test for the sixth grade students of elementary schools in South Semarang Regency could still be used as an instrument of evaluation with some revisions.

Another study about content validity was conducted by Ajeng Desy in 2009. The purpose of this study was to acquire an objective description of a good test item structure in the mid-term test for the eighth-grade students of SMP 33 Semarang. The result of the research revealed that there are 33 valid items and 17 invalid items. Similar study was also conducted by Widyaningrum (2014). The objective of the study was to examine the content validity and authenticity of the English test items of National Examination for Senior High School year 2012. The researcher employed a qualitative research with document analysis. The research objects were the listening items and the five types of reading test items of National Examination year 2012 which aimed to avoid students' fraudulence. The data of the analysis were obtained by using checklists and the data were employed to answer the research questions. There were two findings of this research. First, the content of the National Examination year 2012 was 98.8% valid since almost the contents were relevant to the test specifications. Second, the National Examination year 2012 met the criteria of authenticity with percentage 79.5% since some listening and reading test items were not qualified to authenticity criteria.

In 2018, Hanendra analyzed the English summative teacher-made test of SMAN in Batang City in the academic year of 2015/2016. The study was conducted to investigate how the content validity and the character building values achieved in the English summative teacher-made test. The test consists of 33 multiple-choice items and 10 essay items. There were 7

reading passages in the summative test. The multiple-choice and essay items were analyzed to find out the content validity. According to the research findings, 31 out of 43 test items of the English summative teacher-made test are relevant (72%), 12 items are irrelevant (28%), and none of the test items is partly relevant (0%). Furthermore, 5 out of 6 character building values that were implemented in the syllabus were found in the reading passages of the summative test. The study concluded that the test constructors are suggested to really refer to the syllabus in designing the summative test that all the test items materials would be relevant to the syllabus. Furthermore, it would be better if the test constructors were to apply the character building values in the reading passages since character building values are implemented in curriculum 2013.

Another related study was conducted by Mahardika (2018). The study aimed to investigate the content validity of the test items to prove that the test item is compatible with the syllabus. The English summative test consists of 40 multiple-choice items and 5 essay items. The research findings prove that 31 out of 40 test items of the English summative test for the even semester are relevant, 9 items are categorized as irrelevant items, and none of the test items are partly relevant. Those 9 items are said to be irrelevant because the material administered on each items belong to the first semester.

Similar study about analysis of content validity of English summative test was also done by Janet (2018). The study was conducted to

explore how the content validity and the authenticity achieved in the English summative test for tenth-grade students of a State Senior High School in Kendal in the academic year 2015/2016. The English summative test consists of 40 multiple-choice items (listening and reading section). However, the researcher only analysis 25 multiple-choice items of reading section and 5 reading passages. The result of the study shows that the content of English summative test was 96% valid since most of the test items are relevant to the syllabus of 2013 Curriculum.

The next study about content validity was done by Nugroho (2015). The writer wanted to find out the quality of the English school final exam for the ninth graders made by teachers in SMP N 17 Semarang and SMP Kesatrian 2 Semarang. From the result of the data analysis of validity of the test, the writer concluded that both tests were valid in the term of content validity. However, the both test were not valid in the term of criterion validity. From the result of the analysis, it was also found that the quality of both tests were not good enough. However, the English school final exam of SMP N 17 was better than SMP Kesatrian 2.

After reviewing some previous studies regarding the content validity level of the test, the researcher also reviewed some studies regarding the reading comprehension. There is a study that shows the correlation between the students' reading comprehensive ability to the result of the test. It was conducted by Iswati (2007). The object of this research is Reading Comprehension Ability and National State Exam. The data were collected



by using an instrument consisting of 65 items of multiple choice reading comprehension questions. After analyzing their validity, reliability, the difficulty level and the discriminating power, 45 items were said to be valid and could be used as a research instrument. To find out whether or not there is a correlation between reading comprehension ability as a dependent variable and national state exam as an independent variable, Pearson Product Moment Correlation was used. The result of the study showed that there is a correlation between the students' comprehension ability with the examination score.

The reading comprehension ability is said to be one of important aspects that KTSP curriculum underlined. The use of reading passage in the test items should be related to the syllabus of the curriculum. There is a study that analyzed such focus conducted by Izati (2010). The study was based on the research which tried to investigate the material used in the test according to the KTSP curriculum and also review the construction of multiple choice test items in the English summative test. The result of the study indicates that not all the reading passages found in the summative test are relevant to KTSP. The materials were two four types i.e. narrative (text 7) and recount (text 2, 3, and 6) and invitation (text 4) and short message (text 8). There was one short functional text, announcement, which is not available in the summative test. In addition, there were two genres of text, descriptive and greeting that were not included in the syllabus of KTSP. By

using Gronlund's criteria in analyzing the construction of the multiple-choice test, there are major factors which cause some of the items invalid.

Furthermore, the researcher also reviewed some studies for additional information in this research. A study about language components, which also become one of the main focus in this research, was conducted by Parsi (2018). The study intended to investigate the possible difference between EAP and EFL learners' beliefs concerning learning and teaching of language components, namely, vocabulary, pronunciation and grammar. This study examined the association between EAP and EFL learners' beliefs and their language components' development. Furthermore, for language components' development, the EAP learners' scores in English course and EFL learners' average scores in their Basic English courses were taken into account. The results of an Independent Samples t-test revealed that there existed a statistically significant difference between EAP and EFL learners' beliefs on learning and teaching language components. Furthermore, the results of Pearson correlation coefficients indicated a statistically significant positive association between EFL learners' beliefs and their language components' development, however no statistically significant correlation was found between EAP learners' beliefs and their language components' development.

A similar study about language components was also conducted by Al-Yaari et al, (2013). The study was conducted at the Departments of English both in the College of Education and College of Arts & Linguistics,

Thamar University, Yemen. Ten (10) final term tests administered by ten instructors within two colleges were evaluated. The tests were given to freshmen students to test their productive, receptive skills, as well as their language components. Results show that some tests lack validity, practicality, reliability, discrimination, accountability, and wash back effect. Findings also show that instructors' rubrics were neither clear nor brief. Questions also involved orthographic as well as grammatical mistakes. Finally, in comparison with instructors of College of Education, instructors of College of Arts are found to be better in administering tests.

In addition for the review of previous studies about language components, there was a study conducted by Rakhmah (2009). The general objectives of this study were to examine the validity of the language components covered by the test, test items constructed in term of its content, relevance of the test to the curriculum, levels of Bloom's cognitive taxonomy, and variety of test types. The writer focused on the content validity of the English reading test for the second semester of the eleventh grade of SMA in Brebes in the academic year 2007/2008. The method of this study was qualitative research. The result of the study showed that the validity of the tests was high and relevant to the KTSP curriculum.

In accordance with the main aspects of this research regarding the reading comprehension and one of the language components i.e. vocabulary components, there is a study which shows the correlation between those two aspects. Azkiya (2017) conducted a study which attempted to investigate

the effectiveness of contextual vocabulary teaching in improving year VIII students in one of the junior high schools in Ambal, Kabupaten Kebumen and to find out the significant difference of students' reading comprehension achievement between the students who were taught by using contextual vocabulary teaching and those who were not. By using purposive random sampling, one class was taken and was labeled as the experimental group, and another class was labeled as control group. The experimental group was taught by using contextual vocabulary teaching and the control group was taught using grammar translation method. The data were gained by administering pre-test and post-test of reading comprehension test. The results analyses of the quantitative data showed that the mean scores of the experimental group increased significantly. T-test results of experimental group after utilizing the treatment showed that Sig.2 tailed (0.000) were lower than p value (0.05). It proved that the experimental group's reading achievement improved more significant than the control group did. Consequently, the null hypothesis ( $H_0$ ) was rejected and the alternative hypothesis was accepted which states that there was a significant difference in students' reading comprehension achievement which were taught using contextual vocabulary teaching. Therefore, contextual vocabulary teaching was effective to improve students' reading comprehension achievement.

In addition to the vocabulary components discussion, there was also a study conducted by Al-Qahtani (2015). In his study, he summarized important researches on the importance of vocabulary and explaining many

techniques used by English teachers when teaching English, as well as his own personal view of these issues. His piece of work aims to highlight the importance of vocabulary learning as an essential part in foreign language learning. Although it has been neglected for many decades, during the 80s it became a 'hot' topic for researchers. Lexical knowledge is central to communicative competence and to the acquisition of a second/foreign language and a lack of vocabulary knowledge is an obstacle to learning. Moreover, he argued that in presenting one planned vocabulary item, the teacher usually combine more than one technique, instead of employing one single technique. Teachers are suggested to employ planned vocabulary presentations as various as possible.

All of the research studies above are mainly about content validity analyses, reading comprehension, and the language components. However, those all previous studies only analyzed the content validity of a test, the reading comprehension aspects, and also the language components separately. None of the studies above have integrated all of those three main aspects into one study. Therefore, the study presented in this final project is different from the previous studies. In this study, the writer conducted his research on analysis of content validity, reading comprehension, and also the language components. Hopefully, the result of this study will fill the gap that have been left by all of those previous studies. Furthermore, the results of this study also provide some new insights for the future researchers in conducting the study about content validity, reading comprehension, and

language components. Hopefully, the future studies can deal with the analysis of the content validity and also with both of the language skills and language components, which become the cores and bolts of language, embedded in the summative test balancedly.

## **2.2 Related Literature**

This subchapter discusses the notions of curriculum, syllabus, assessment, test, table of test specification, types of test, and types of test items, language knowledge and skills, and also language components, language testing, validity, types of validity, and content validity.

### **2.2.1 Notions of Curriculum**

Curriculum or educational module is the foundation for doing teaching and learning process. Educators ought to understand the curriculum before they show students, as the educational programs have been made on the premise of understudies' need. Considering the significance of educational modules, the author will talk further about its definition and KTSP educational programs.

According to Oxford Advanced Learners' Dictionary of Current English, a curriculum is the subjects included in a course of study or taught at a particular school, college, etc. (Hornby, 1995, p.287). From the previous explanation, it can be concluded that a study in schools, colleges, universities, and some other institutions should be relevant to the curriculum.

### **2.2.1.1 The Main Competence of English for Senior High School**

According to the BNSP's Standar Isi KTSP for Senior High School (2006, p.126), standard competences of English subject at SMA are to communicate the language in spoken and written form, using various texts fluently and accurately.

Furthermore, Standar Isi KTSP for Senior High School (2006: 126) shows that the scope of English in SMA involved:

- (1) Statement's ability: comprehensive ability and/or produce spoken text and/or written text which is realized together in four language skills, they are listening, speaking, reading and writing to reach informational literacy's level.
- (2) Comprehensive ability and creative various short functional texts, monolog and essay in form of procedure, recount, narrative, report, news item, analytical exposition, hortatory exposition, spoof, explanation, discussion, review, and public speaking. The use of vocabulary, grammar, and rhetorical way could be seen on the material's gradation.
- (3) Supporting competence: linguistics competence, sociocultural competence, strategy competence, and statement competence.

From the explanation above, the conclusion is that SMA graduates should master listening, speaking, reading, and writing text in form of descriptive, narrative, recount, procedure, report and anecdote text.

### **2.2.2 Notions of Syllabus**

Meanwhile, the curriculum is defined as the guideline of the chapters and academic content covered by an educational system while undergoing a particular course or program. It is well planned, guided and designed by the government or the educational institution and aimed at both physical and mental development of a student.

The syllabus is defined as the documents that consist of topics or portion covered in a particular subject. It is considered as a guide and helps the students to know about the subject in detail, why it is a part of their course of study, what are the expectations from students, consequences of failure, etc. It contains general rules, policies, instructions, topics covered, assignments, projects, test dates, and so on.

In other words, the syllabus is described as the summary of the topics covered or units to be taught in the particular subject, while curriculum refers to the overall content, taught in an educational system or a course. A subject syllabus is a unit of the curriculum. The two terms differ in a sense that curriculum is a combination of some factors which help in the planning of an educational program, whereas a syllabus covers the portion of what topics should be taught in a particular subject.

Based on the explanation above, it is crucial for a test to be constructed based on applied syllabus and curriculum. As stated by Brown (2004, p.22), if the test contents cover the test objectives, which in turn are representatives of the syllabus, it could be said that the test possesses content



validity. Therefore, the consistency between the syllabus content, the test objective, and the test contents must be maintained in constructing the test. In brief, this consistency is the main aspect the researcher will try to analyze to reveal the content validity of the summative-test. The table below presents the overall English materials for the senior high school students derived from the syllabus of tenth-grade in the KTSP Curriculum.

**Table 2.1**

**List of English Materials in the X Grade of Senior High School**

<b>Materials in the KTSP Curriculum</b>	
<b>A. Functional Skills</b>	a. Expression of acquainted
	b. Expression of greeting
	c. Expression of departing
	d. Expression of inviting
	e. Expression of accepting invitation/offer
	f. Expression of accepting appointment
	g. Expression of cancelling appointment
	h. Expression of happiness
	i. Expression of attention
	j. Expression of sympathy
	k. Expression of command
	l. Expression of complimenting
	m. Expression of gratitude
	n. Expression of congratulating
	o. Expression of surprise
	p. Expression of disbelief
<b>B. Genres</b>	<b>a. <i>Short Functional Text</i></b>
	1. Invitation
	2. Advertisement
	3. Announcement
	4. Short Message
	5. Notice
	<b>b. <i>Monologue</i></b>
	1. Recount
	2. Narrative
	3. Procedure
	4. Descriptive
	5. News Item
	<b>c. <i>Essay</i></b>
	1. Recount
	2. Narrative

<b>C. Grammar</b>	3. Procedure
	4. Descriptive
	5. News Item
	a. Present & Past Tenses
	b. Progressive Tenses
	c. Perfect & Future Tenses
	d. Passive Voice
	e. Reported Speech
	f. Adjectival and Adverbial Clauses

*(Standar Isi KTSP for Senior High School, 2006)*

### 2.2.3 Language Assessment

Assessment is a popular and sometimes misunderstood term in current educational practice. It is often thought that testing and assessing are synonymous terms, but they are not. According to Brown (2004, p.4), assessment is an ongoing process that encompasses a much wider domain. Whenever a student responds to a question, offers a comment, or tries out a new word or structure, the teacher subconsciously makes an assessment of the student's performance.

Meanwhile, tests are prepared administrative procedures that occur at identifiable times in a curriculum when learners muster all their faculties to offer peak performance, knowing that their responses are being measured and evaluated. Tests, then, are a subset of assessment. They are certainly not the only form of assessment that a teacher can make. Tests can be useful devices, but they are only one among many procedures and tasks that teachers can use to assess students (Brown, 2004, p.4).

For the optimal learning to take place, students in the classroom must have the freedom to experiment, to try out their own hypotheses about language without feeling that their overall competence is being judged in

terms of those trials and errors. Teaching can set up the practice games of language learning: the opportunities for learners to listen, think, take risks, set goals, and process feedback from the teacher and then recycle through the skills that they are trying to master.

However, before teachers try to design the assessment as their part of teaching, they have to know that to teach English in EFL context means that teachers need to prepare many things toward their teaching and learning process in the classroom. Trisanti (2017), said that in language teaching and learning, many things need to be prepared related to background, situation, and characters of the learners so that it will give impacts in the teaching and learning process. That is why need analysis need to be done to know the need of learners in learning the target language. Teachers also need to prepare themselves in improving their language skills of the target language since it may give impact on the way the learners get input in their learning.

#### **2.2.4 Notions of Test**

In way to know how well the learning procedure takes the progress, the educator must assess it. There are numerous ways to assess the teaching-learning process. One of which is a test. According to Penny Ur (1996, p.33), tests are used as a means to motivate students to learn or review specific material. It means students have to learn and review the material in order to comprehend it and pass the test.

Meanwhile, Linn & Gronlund (1995, p.26) stated that a test is a particular type of assessment that typically consists of a set of questions

administered during a fixed period of time under reasonably comparable conditions for all students. In addition, Brown (2004, p.3) explained that a test is a method of measuring a person's ability, knowledge or performance in a given domain. In this study, the given domain is related to the second language acquisition. Therefore, the test deals with the language ability, knowledge and performance in the language learning.

### **2.2.5 Types of Test**

The next segment will elaborate the divisions of basic types of language test.

#### **2.2.5.1 Language Proficiency Test**

According to Hughes (2003, p.9), proficiency tests are designed to measure learner's ability in a language, regardless of any training they may have in that language. The content of proficiency test is based on a specification of what candidates have to be able to do in the language in order to be considered proficient. This in line with Harris (1969, p.6) that stated the proficiency test measures what students have learned, but the aim of such test is to determine whether this language ability corresponds to specific language requirements. The proficiency tests usually report student language ability on a continuum that reflects a predetermined set of categories.

Furthermore, Brown (2004, p.44) mentioned that a proficiency test is not limited to any one course, curriculum, or single skill in the language; rather, it tests overall ability. Proficiency tests have traditionally consisted of standardized multiple-choice items on grammar, vocabulary, reading

comprehension, and aural comprehension. Sometimes a sample of writing is added, and more recent tests also include oral production performance. From this explanation, it can be inferred that the proficiency tests are typically standardized test types.

### **1. Standardized Test**

Brown (2004, p.67) stated that a standardized test presupposes certain standard objectives, or criteria, that are held constant across one form of the test to another. The criteria in large-scale standardized-tests are designed to apply to a broad band of competencies that are usually not exclusive to one particular curriculum. A good standardized test is the product of a thorough process of empirical research and development. It dictates standard procedures for administration and scoring. Some typical examples of the standardized proficiency test are the TOEFL, ESLPT, and IELTS.

#### **2.2.5.2 Language Aptitude Test**

Brown (2004, p.43) stated that language aptitude test is designed to measure general ability to learn a foreign language and the ultimate success in that undertaking. Language aptitude tests are apparently designed to apply to the classroom learning of any language.

In line with the previous statement, Harris (1969, p.5) explained that 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 students begin language study, and may be used to select students

for a language course or to place students in sections appropriate to their ability.

From previous explanation, it can be concluded that the language aptitude tests also include another types of tests which are placement tests and screening tests or diagnostic test. The elaboration of such tests are as follows:

### **(1) Placement Test**

Brown (2004, p.45) stated that the purpose of a placement test is to place a student into a particular level or section of a language curriculum or school. Such test usually, but not always, includes a sampling of the material to be covered in the various courses in a curriculum; a student's performance on the test should indicate the point at which the student will find material neither too easy nor too difficult but appropriately challenging.

### **(2) Diagnostic Test**

According to Brown (2004, p.46), a diagnostic test is designed to diagnose specified aspects of a language. For example, a test in pronunciation might diagnose the phonological features of English that are difficult for learners and should become part of a curriculum.

The diagnostic and placement test may sometimes indistinguishable from each other. However, those two tests are slightly differ on the purposes. Any placement test that offers information beyond simply designating a course level may also serve diagnostic purposes.

### **2.2.5.3 Achievement Test**

According to Brown (2004, p.47), an achievement test aims to assess the students' achievement in classroom instruction and it is an important procedure in individual and programmed instruction. Such test is related to classroom lessons, units, or even a total curriculum. Achievement tests are (or should be) limited to a particular material addressed in a curriculum within a particular time frame and are offered after a course has focused on the objectives in question.

From the explanation above, it can be concluded that the achievement tests also include another types of tests, as follows:

#### **1. Formative Test**

Formative test measures how much the student has learned in a specific course of instruction. It the test that the classroom teacher prepares for administration at the end of a unit. Brown (2004, p.6) also explained that formative test is evaluating students in the process of “forming” their competencies and skills with the goal of helping them to continue that growth process. The key to such formation is the delivery (by the teacher) and internalization (by the student) of appropriate feedback on performance, with an eye contact toward the future continuation (or formation) of learning.

#### **2. Summative Test**

This test is usually given at the end of a marking period and measures the ‘sum’ total of the material covered. It confirms the explanation from

Brown (2004, p.6), summative test aims to measure or summarize, what a student has grasped, and typically occurs at the end of a course or unit of instruction. Final exams in a course and general proficiency exams are examples of summative test. To conclude, summative test measures how much student has learned in the course of second-language instruction.

### **2.2.6 Types of Test Items**

There are 12 types of test items according to Penny Ur (1996, p.39). However, in this section, the author won't examine all those tests yet concentrates only on the types found in the summative test being analyzed. It is mainly just on the written, multiple-choice-type test. The elaboration for the types of test items is in the following:

#### **(1) Questions and Answers**

Simple questions, very often following reading, or as part of an interview; may require short or long answers.

#### **(2) Gap-filling and Completion**

The testee has to complete a sentence by filling a gap or adding something. A gap may or may not be signaled by a blank or dash; the word to be inserted may or may not be given or hinted at.

#### **(3) Rewriting**

A sentence is given; the tester rewrites it, incorporating a given change of expression, but preserving the basic meaning.



#### **(4) Multiple-Choice**

The question consists of a stem and a number of options (usually four), from which the tester has to select the right one. This particular test item type has some advantages in the administration and several suggestions for constructing such type.

##### **2.2.6.1 The Advantages of Multiple-Choice Items**

As indicated by Tinambunan (1988, p.75), there are six preferences of utilizing the multiple-choice form. The advantages are:

- (1) The multiple-choice item is adaptable to subject matter content areas as well as different levels of behavior. It can be used in assessing ability to reason, discriminate, interpret, analyze, make inferences and solve problems.
- (2) The structure of a premise with four or five alternatives provides less chance of guessing the correct answers from blind guessing: therefore, the higher-achieving students should get the item correct.
- (3) One advantage of the multiple-choice item over the true-false item is that pupils cannot receive credit for simply knowing that a statement is incorrect; they must also know what is meant by the correct one (Gronlund, 1981, p.173).
- (4) Four or five options in the multiple-choice test provide more incorrect choices for selections of responses by the student who does not know the best or correct answer.

- (5) The difficulty of each multiple-choice item can be controlled by changing the alternatives. The more homogeneous the alternatives, the more difficult it is to select the correct response from the given alternatives.
- (6) Multiple-choice items are agreeable to item analysis which enables the teacher to determine how well the items functioned with the students tested and how well each alternative functioned in discriminating between the higher achieving and lower achieving students.

### **2.2.7 Characteristics of Good Test**

According to Ahmed (2014), a good test should have a positive effect on learning, and teaching should result in improved learning habits. Such a test will aim at locating the specific and precise areas of difficulties experienced by the class or the individual student so that assistance in the form of additional practice and corrective exercises can be given.

The test should enable the teacher to find out which parts of the language program cause difficulty for the class. In this way, the teacher can evaluate the effectiveness of the syllabus as well as the methods and materials he or she is using. A good test should also motivate by measuring student performance without in any way setting "traps" for them. A well-developed test should provide an opportunity for students to show their ability to perform certain language tasks.

According to Bachman and Palmer (1996, p.17), a test's usefulness can be determined by considering the measurements qualities of the test

such as reliability, validity, practicality, discrimination and authenticity. These qualities can easily describe a good language test's usefulness. The test's usefulness is the most important quality or cornerstone of testing. They state that test usefulness provides a kind of metric by which we can evaluate not only the tests that we develop and use but also all aspects of test development and use.

### **2.2.8 Notions of Language Knowledge and Language Components**

Bachman (1990, p.253) suggested that language knowledge includes two types of knowledge that a second language learner must internalize:

- a) Organizational knowledge, that is knowing how to control the formal structure of a second language so as to produce correct sentences and organize these in texts. It subsumes grammatical and textual knowledge.
- b) Pragmatic knowledge, which involves knowing how words and utterances can be assigned specific meanings in context and function according to the user's intentions. This knowledge is also structured in lexical knowledge, functional knowledge, and sociolinguistic knowledge.

Later on, Bachman and Palmer (1996, p.61) explained language ability was viewed as a set of finite components-grammar, vocabulary, pronunciation, spelling-that were realized as four skills-listening, speaking, reading, and writing.

The language also has been analyzed as being composed of skills: listening, speaking, reading, writing and component: sound, vocabulary, structure (Harris, 1969, p.9). The language skills of listening contain the language components of vocabulary and structure; speaking--sound, vocabulary, and structure; reading--vocabulary and structure as writing does. Harris' view was a static one as it ignored the dynamic nature of language use which causes the skills to be interchangeable requiring participants in the interactional and transactional discourses to negotiate ideas, feelings and understanding (Widdowson, 1983, p.73).

Both skills and components are cores and bolts of language. The former may be compared to bone or skeletal structure of a human flesh body while the latter to the human flesh. They must be together to form a complete living human being. In language testing, to test language components alone void of skill is called discrete-point testing.

Brown (2004, p.3) has explained that a person's ability, knowledge or performance of second language acquisition is being measured in a test. Therefore, the test deals with the language ability, knowledge and performance in the language learning. Language learning comprises of language skills and language components in different structures. So does language testing, it can be developed in various organizations. This will be discussed later.

### **2.2.9 Language Testing**

The KTSP curriculum of SMA emphasizes communicative competence and has applied the communicative approach to the teaching of English as a first foreign language in Indonesia. Students are expected to have the language ability such as the ability to know about language (language competence) and the ability to use language (language performance).

A language test is often interpreted as an activity where students in a foreign language class are obliged to do a written examination. In fact, there are numerous examination frameworks which can be administered to students. Written and oral examinations are parts of them. However, the written examination has more benefits than oral examination. It has been obvious that an oral examination takes time since there can be no more than two or three students sitting in such an examination.

Hence, the assessments of the students' skill are less objective. For this reason, the written examination turned out to be more normally utilized. The advantage of the written against oral examination is that the existence of students' performance records, so the judgment of any examiner can be checked against that of another.

There are some ways to assess performance of major skills. Language skills incorporate the test of listening, speaking, reading and writing, while language components oblige students to have adequate knowledge about phonology/sound system, vocabulary and grammar. The

discussion for the testing of language components and language skills will be presented in the following subchapters.

### **2.2.9.1 Language Components Testing**

Language components testing belongs to language learning. There are three main aspects in the language components. They are phonology/sound system, vocabulary, and grammar. The following sections will elaborate tests of the language components mentioned above.

#### **(1) Test of Phonology/Sound System**

Test items intended to test phonology may endeavor to assess several sub-skills. According to Heaton (1975, p.120), test of phonology can be used to evaluate the students' ability in recognizing and pronouncing the significant sound contrasts, in recognizing and using the stress and the melody/the patterns of the tunes of a language. It is clear that test of phonology is blended in oral and aural abilities.

However, such sort of test is uncommon today. This condition is caused by some factors. From what can be seen nowadays, there are instructors who totally couldn't care less about the testing of phonology, in this case, the test of pronunciation, since they do not comprehend the role of pronunciation in language. Additionally, there are instructors who don't give such a test despite the fact that they understand its significance. It is probably caused by their lack of knowledge of how to test it in a productive and straightforward way.

## **(2) Test of Vocabulary**

Vocabulary tests have been prepared as measures of general ability or intelligence and as measures of achievement in special fields. The first task for the administrator of a vocabulary test is to determine the level to which he wishes to concentrate on testing the students' active or passive vocabulary. His next task is then to decide whether the lexical items in the test should be taken.

Madsen (1983) said further that the purpose of a vocabulary test is to measure the comprehension and production of words used in speaking and writing. Deciding how to test vocabulary is related to how to teach it. Most English as second language teachers today do not recommend having students simply memorize lists of words. Instead, they teach students to find the meaning of words through the context of sentence, and they help increase comprehension by teaching important affixes.

## **(3) Test of Grammar**

Grammatical structure is the patterns of arrangement of words in sentences and the patterns of arrangement of parts of words into words. Meanwhile, the test content of grammar is used to measure the students' ability to manipulate structures and to distinguish appropriate grammatical forms from inappropriate ones (Brown, 2004, p.226).

Selecting grammar subjects to test is usually rather easy; simply figure out what structures the instructor has educated since the last test.

The result of quizzes or homework assignments can show those things that students have learned well and which that need reviewing. The subjects they know well can be ignored. However, a few of those could be included at the beginning of the test to encourage students.

Grammar tests appear to be the most famous among the other test of language components. Tests of grammar are often presented in the form of multiple-choice items. This form is considered to be an efficient way. The other forms, however, sometimes are also used for testing grammar such as simple competition and cloze test.

#### **2.2.9.2 Language Skills Testing**

Four major skills in communicating through language are frequently comprehensively characterized as listening, speaking, reading, and writing. Furthermore, Harris (1969) also stated that speaking and writing are the decoding processes by which we communicate our ideas, thoughts, or feelings through a language, and listening and reading are decoding processes by which we ‘understand’ either a spoken or a written message.

However, communication skill is judged to be successful if a person is able to handle and combine the four major skills in his real interaction. In other words, a successful communication needs a person’s ability to express facts, feelings, ideas, and attitudes clearly and effortlessly, either spoken or in written, and his ability to comprehend what he hears or reads. These abilities are often called fluency, and according to Heaton (1975), it can best be measured by tests which evaluate performance in the language skills.



The aspect of the summative test being analyzed belongs to the domain of reading skill. So, the following section will only deal with test of the language skill being analyzed.

### **(1) Reading Test**

Reading skill in the second or foreign language can be developed well if that of the first language has been well mastered. In accordance with this, Heaton (1975) has stated that before reading tests in the second or foreign language can be successfully constructed, the first language reading skills of the testers must be determined.

Tests of reading come in a wide variety of forms and evaluate a broad field of reading activities. However, reading activity to which students are usually exposed is reading comprehension. According to Madsen (1983), tests of reading comprehension include tests of sentence comprehension and passage comprehension.

### **2.2.10 Notions of Validity**

Validity is the degree to which a test measures what it cases to gauge. It is important for a test to be legitimate all together for the outcomes to be precisely connected and deciphered. Bachman in Salwa (2012, p.47) stated that validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and action based on test scores or other modes of assessment.

According to Bloom et al (1981, p.39), validity is concerned with the specific use to be made of the result. It must be emphasized that the validity of such inferences depends on an accurate knowledge of the background-at least the background in the school's experience-which the student brings to the item. This is in accordance with Brown's (2004, p.22) statement about validity..

In some cases, it may be appropriate to examine the extent to which a test calls for performance that matches that of the course or unit of study being tested. In other cases, it may be concerned with how well a test determines whether or not students have reached an established set of goals or level of competence.

### **2.2.11 Types of Validity**

Johnson (2001, p.301) mentions there are five types of validity:

#### **(1)Content Validity**

Content validity is about what actually goes into the test. To have content validity, a test's content must be seen as representative of the subject area being covered. The content validity is concerned with how the test measures the subject matter and behavior under consideration. The definition of content validity states that the test items must be a representative sample of the domain of possible content or behavior.

#### **(2)Face Validity**

Face validity, on the other hand, is to do with what the world thinks of the test. It relates again to that all-important figure in the testing world,

the outsider. People must be able to look at a test and, as a layperson, be convinced that it is a test giving valid information about language use.

### **(3) Construct Validity**

Construct validity deals with the relationship between a test and a particular view of language and language learning.

### **(4) Empirical Validity**

Empirical validity deals with how the test relates to other testing measures. The test should not yield results that are dramatically at odds with the results of other forms of assessment.

### **(5) Predictive Validity**

Predictive validity is where its results yield some information about the future. It may be particularly useful to find out whether placement tests are valid or not.

From the explanation above, it can be concluded that each type of validity is significant. However, one of the most complex and important criteria of effective test goes to the content validity, in which the compatibility of the test contents and the test objectives with the curriculum and syllabus applied in the course becomes the main concern.

### **2.2.12 Content Validity**

The principal validity for achievement tests is content validity, sometimes also called content relevance. Content validity discusses about content of test. Wiersma and Stephen (1990, p.184) said that content validity is concerned with the extent to which the test is representative of a defined

body of content consisting of topics and processes. Therefore, the test should reflect instructional objectives or subject matters. But it is not expected that every knowledge or skills will always appear in the test; there may simply be too many things for all of them to appear in a single test.

The test constructor should look into whether the test content is representative of the skills that are supposed to be measured. This involves looking into the consistency between the syllabus content, the test objective and the test contents. If the test contents cover the test objectives, which in turn are representatives of the syllabus, it could be said that the test possesses content validity (Brown, 2004, p.22).

The content of a test should be decided by considering the purposes of the assessment, and then set a list known as a content specification. The content specification is important since it ensures as far as possible that the test reflects all the areas to be assessed in suitable proportions and also because it represents a balanced sample, without bias towards the kinds of items which are easiest to write or towards the test material which happens to be available.

Hence, content validity is considered with how well the sample of test tasks represent the domain of task to be measured. Content validity is also called curricular validity, because materials that will be tested are based on the applied curriculum. Therefore, the way to know the content validity is by looking the objective of curriculum and the material of the test. The measurement of content validity is important in employment setting, where

tests are used to hire and promote people are carefully scrutinized for their relevance to the job. So, the test makers must be careful to make a test to be good test based on the syllabus.

The content validity is concerned with how the test measures the subject matter and behaviour under consideration. The short definition of content validity states that the test items must be a representative sample of the domain of possible content or behaviour. Though not the only types of validity, content validity is the most appropriate method for evaluating the validity of achievement tests. The previous elaboration is the requirements of content validity which must be included in constructing the test in order to get a high quality test in terms of the content. In conclusion, then, content validity is concerned with whether or not the content of the test is sufficiently representative and comprehensive for the test to be a valid measure of what it is supposed to measure.

### **2.3 Framework of the Analysis**

One of the most important aspects of the teaching-learning process is evaluation, as a systematic gathering of information. The instrument that is needed to collect information of student is a test. Djiwandono (1996, p.1) states that test is a tool or product used to measure the student language of proficiency. However, the fact nowadays there are many teachers that do not care about the content of the test given to the student. The teacher took very limited time to prepare a test. The limited time of teacher on preparing the test can make the test does not have good quality. When teachers make

a test, they must concern about five principles of test that use as assessments. The five principles are reliability, validity, practicality, authenticity, and wash back. It is why the researcher decides to conduct the research about content validity.

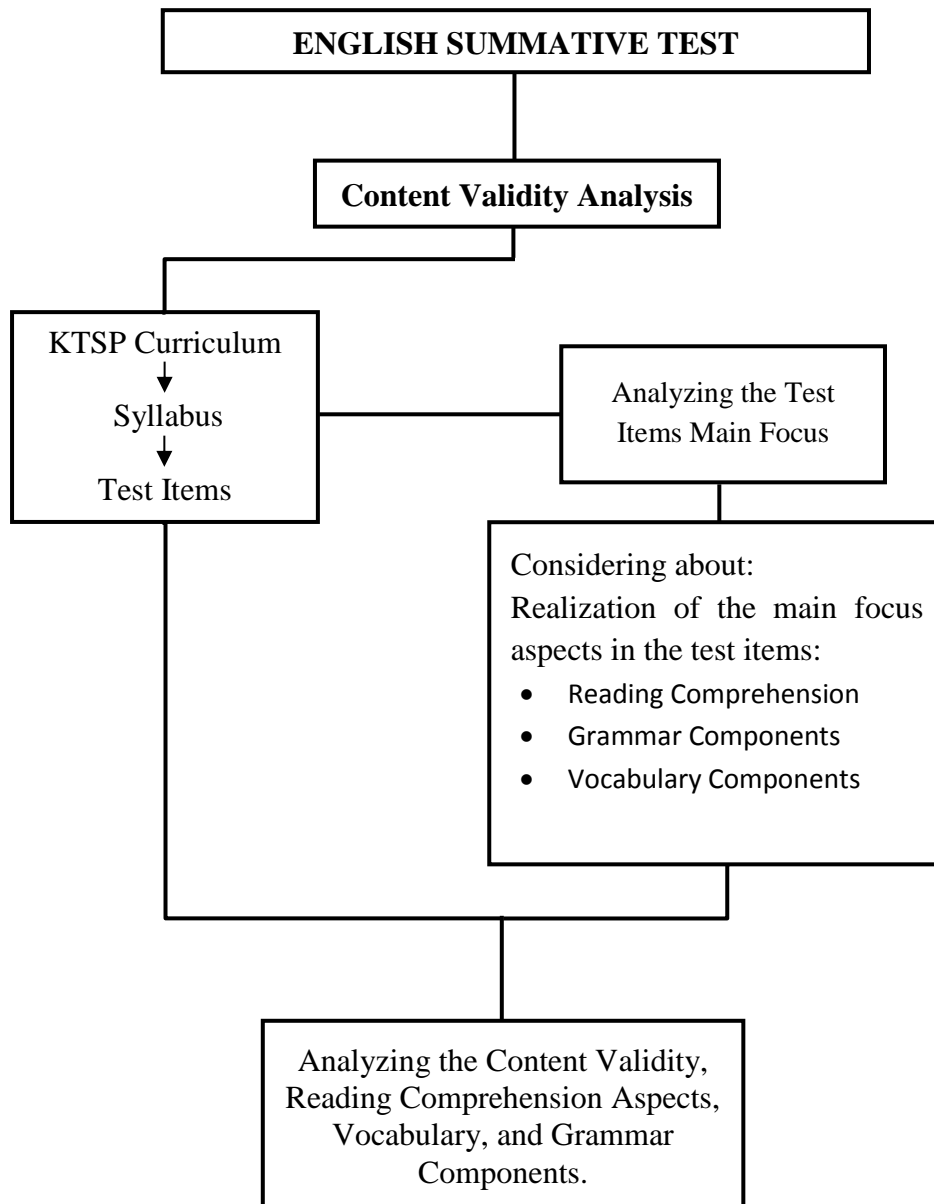
In order to have a content validity, the test content should be a representative of the skills that are supposed to be measured. Hughes (2005, p.26), stated that a test is said to have content validity if its content constitutes a representative sample of the language skills, materials, structures, etc. in which it is meant to be concerned. Those set of skills, structures, and materials are already proposed by the syllabus and curriculum.

Furthermore, the language components being realized in the test is also analyzed. In line with Harris' explanation (1969), Penny Ur (1996) also said that linguists have broken language down into three main components: the phonology, or the sound system of the language; the lexis, or the words or phrases which express concepts; the structure, or the way words or bits of words are strung together to make acceptable sentences or phrases. Language teachers define these more conveniently as pronunciation, vocabulary, and grammar.

However, since the test does not assess the pronunciation components, the reading comprehensive ability will replace the pronunciation place. Considering it is included in the scope of English in SMA as explained earlier. The aspects regarding reading comprehensive

being analyzed are also based on the syllabus and KTSP curriculum for Senior High School. The main aspects proposed for analyzing the main focus of the test items regarding the reading comprehensive, vocabulary, and grammar components are derived from the indicators stated in the syllabus. The operational verbs used in the indicators of the syllabus then are used for constructing the criteria points being applied in analyzing the test items. Therefore, this study focuses on the discussion about analysis of content validity and the main focus assessed in the test items regarding the reading comprehension aspects, vocabulary and also grammar components in the second-semester English summative teacher-made test for the tenth grade students. The validity is the important principles of language testing. It is the extent to which inferences made from assessment results are appropriate, meaningful, and useful in term of the purpose of the assessment.

The framework analysis of the research is described in the picture below:



**Figure 2.1 Theoretical Framework**

A further explanation about the subject of the study, research instruments, and research procedures will be elaborated in the next chapter.



## **CHAPTER V**

### **CONCLUSIONS AND SUGGESTIONS**

Chapter 5 presents the conclusions and the suggestions. The conclusion contains the result of the study. While the suggestion presents some suggestions for those who concerned in educational field. The researcher would also like to state the limitation of this study in the last subchapter.

#### **5.1 Conclusions**

The objective of this study is to investigate whether the English school final test items administered to the tenth grade of senior high schools in Pekalongan in the Academic Year 2015/2016 are relevant to the syllabus of KTSP curriculum or not, including the realization of its main focus in terms of reading comprehension aspects, vocabulary and also grammar components. Finally, some conclusions for the analysis can be derived.

Viewed from the relevance of the test items materials to the KTSP curriculum, the level of validity was insufficient. It can be seen from the fact that the number of test items which have relevant materials to the KTSP curriculum was the same (50%) with the irrelevant ones (50%). Meanwhile the minimum percentage for the good criteria must be above 76% and above 56% for the sufficient criteria.

From thirty test items, twenty-three of them, or 77 per cent, consist of test on reading comprehension aspects. It means that most of the test items in the multiple-choice type test in this English summative test

administered to second semester students of the tenth grade SMA in Pekalongan in the academic year 2015/2016 are mainly test of reading comprehension.

## **5.2 Suggestions**

Constructing test items, especially multiple-choice test is not an easy job. The test markers should know the rules of constructing good test items. In accordance with the conclusions above, the writer would like to offer the following suggestions.

Theoretically, the implementation of the curriculum must be adequately introduced to all the teachers and test makers. Therefore; the test makers, especially the English teachers in the MGMP in the Pekalongan regency, can know and be familiar with the curriculum and understand the rules of constructing a good test in order to achieve a high-rate of content validity.

Practically, I suggest the test makers to really refer to the syllabus in constructing the test items, so that all the learning materials are relevant to the syllabus in terms of the time the test is administered, whether it belongs to the first or second semester. Furthermore, the test items would be better if they are constructed based on the elaborate table of specification. In addition, it is also better for the test designers to balance all the composition of the test items main focus. Lastly, the test makers, especially the English teachers should know or be familiar with the rules of constructing a good

test and they should check and proofread the draft to avoid errors in the test items construction.

Pedagogically, it is suggested that the teachers really pay attention the syllabus and the curriculum in designing test items, so that the test items are really relevant to the learning objectives and possess a high-level of content validity. It is also important for education stakeholders to be able to develop a curriculum that meets the needs of the field and perform the evaluation at a more advanced level by taking into account the validity of the assessment. Finally, the writer suggests that this test must be revised.

### **5.3 Limitations of the Study**

The objective of this study is to investigate the content validity and the realization of language components of the English school final test items administered to the tenth grade of senior high schools in Pekalongan in the Academic Year 2015/2016. Unfortunately, the result for the realization of the test items main focus in terms of reading comprehension aspects, vocabulary and also grammar components in the second-semester English summative teacher-made test for the tenth-grade students of SMAN in Pekalongan, especially the language components aspect, cannot be compared with another study. The researcher was not able to find such closely-related study since that studies focusing on the all particular feature of language components, like grammar, pronunciation or vocabulary are scarce. Therefore, the researcher suggests that other future researcher should be able to conduct more study about it.

## BIBLIOGRAPHY

- Ahmed, H. (2014). Five Characteristics of a Good Language Test. *National Journal of Extensive Education and Interdisciplinary Research*, Vol. 2, Issue 4 (pp 61-66). Retrieved from [https://www.academia.edu/13648955/FIVE\\_CHARACTERISTICS\\_OF\\_A\\_GOOD\\_LANGUAGE\\_TEST?auto=download](https://www.academia.edu/13648955/FIVE_CHARACTERISTICS_OF_A_GOOD_LANGUAGE_TEST?auto=download).
- Al-Qahtani, M. (2015). The Importance of Vocabulary in Language Learning and How to be Taught. *International Journal of Teaching and Education*, 3(3), 21-34. 10.20472/TE.2015.3.3.002.
- Al-Yaari, S. A., Al-Hammadi, F. S., & Alyami, S. A. (2013). Evaluation of Language Testing: An Applied Linguistic Perspective. *International Journal of English Language Education*, 1(2), 27-54. 10.5296/ijele.v1i2.3059. Retrieved from <http://dx.doi.org/10.5296/ijele.v1i2.3059>
- Arifin, H. R. (2009). *Analyzing of the Test Items in English Final Examination for the Sixth Grade Students of Elementary School in South Semarang Regency in the Academic Year 2008/2009*. Faculty of Languages and Arts. Semarang: Unnes. Retrieved from <https://lib.unnes.ac.id/118/>
- Arikunto, S. (2006). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: PT. Rineka Cipta.
- Azkiya, A. (2017). *Improving the Students' Reading Comprehension Achievement through Contextual Vocabulary Teaching. A Case Of Year Viii Students Of Junior High School In Ambal, Kabupaten Kebumen*. Faculty of Languages and Arts. Semarang: Unnes. Retrieved from <https://lib.unnes.ac.id/30548/>
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. New York: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*, New York: Oxford University Press.
- Bloom, B. S., George, F., & Hastings, J. T. (1981). *Evaluation to Improve Learning*. New York: McGraw Hill.
- BNSP. (2006). *Standar Isi KTSP for Senior High School*. Jakarta: Badan Nasional Standar Pendidikan.
- Brown, D. H. (2004). *Language Assessment: Principles and Classroom Practices*. (4th ed.). New York: Longman.

- Delgado, R. E. (2012). Content Validity Evidences in Test Development. *International Journal of Clinical and Health Psychology*, 12(3), 449-460. Retrieved from <http://www.zora.uzh.ch/64551/1/ijchp-421.pdf>
- Desy, A. (2009). *The Analysis of Validity, Reliability, Discrimination Power and Level of Difficulty of First Mid-Term Test in the Case of the Eighth Grade Students of SMP 33 Semarang*. Faculty of Languages and Arts. Semarang: Unnes. Retrieved from <http://lib.unnes.ac.id/2261/1/6444/>
- Djiwandono, M. S. (1996). *Bahasa Tes Pengukuran*. Bandung: Penerbit ITB.
- Fauzi, S. (2011). *Analysis of the Content Validity of the English Summative Test (A Case Study at the Second Grade of MTs. Salafiyah Bedahan Kota Depok)*. Faculty of Tarbiyah and Teachers. Jakarta: State Islamic University Syarif Hidayatulloh. Retrieved from [repository.uinjkt.ac.id/.../1/salman%20fauzi-fitk.pdf](http://repository.uinjkt.ac.id/.../1/salman%20fauzi-fitk.pdf).
- Gronlund, N. E. (1981). *Measurement and Evaluation in Teaching*. Fourth Edition. London: McMillan.
- Hanendra, I. (2018). *Investigating the Content Validity and Character Building Values of English Summative Teacher-Made Test (The Case of Tenth Grade Students of SMA Negeri in Batang in the Academic Year of 2015/2016)*. Faculty of Languages and Arts. Semarang: Unnes.
- Harris, D. P. (1969). *Testing English as a Second Language*. New York: McGraw Hill.
- Heaton, J. B. (1975). *Writing English Language Tests*. Essex. England: Longman Group Limited.
- Hornby, A. S. (1995). *Oxford Advanced Learner's Dictionary of Current English*. (5th ed.). London: Oxford University Press.
- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Iswati, T. (2007). *The Ability in Comprehending Reading Materials in Relation to the Final Achievement of the Third-Grade Students of Mts. Al-Uswah Bergas Semarang Regency*. Faculty of Languages and Arts. Semarang: Unnes. Retrieved from <https://lib.unnes.ac.id/3099/>
- Izati, N. (2010). *The Construction of English Multiple Choice Summative Test Items of KTSP (A Test for the Eight Grade Students in the Second Semester of SMP in Batang in the Academic Year Of 2009/2010)*. Faculty of Languages and Arts. Semarang: Unnes. Retrieved from <https://lib.unnes.ac.id/5024/>

- Jandaghi, G. (2010). Assessment of Validity, Reliability and Difficulty Indices for Teacher-Built Physics Exam Questions in First Year High School. *Educational Research and Review*, 5(11), 651-654. Retrieved from <http://www.academicjournals.org/journal/ERR/article-abstract/2AA93A14201>
- Janet, A. (2018). *Exploring Content Validity and Authenticity of English Summative Test (A Case of Tenth Graders of SMA Negeri 1 Kendal in the Academic Year of 2015/2016)*. Faculty of Languages and Arts. Semarang: Unnes.
- Johnson, K. (2001). *An Introduction to Foreign Language Learning and Teaching*. Harlow: Pearson Education Limited.
- Kahsay, M. Y. (2014). Assessment of Content Validity in Ethiopian General Secondary Education Certificate of English Examinations (2009-2012). *Global Scholars Journal of Linguistics and Communication Studies*, 1(1), 1-13. Retrieved from <https://pdfs.semanticscholar.org/17af/dfd3964c0cfe66f5>
- Kristiana. (2014). *An Analysis on the Content Validity of Summative Test for the Second Grade Students of Junior High School*. Faculty of Tarbiyah and Teachers. Jakarta: State Islamic University Syarif Hidayatulloh. Retrieved from [repository.uinjkt.ac.id/dspace/bitstream/123456789/25380/1/kristiana-fitk.pdf](http://repository.uinjkt.ac.id/dspace/bitstream/123456789/25380/1/kristiana-fitk.pdf).
- Lei, M. I. (2015). Validity of Teacher-Made Assessment. *Asian Social Science*, 11(5), 193-200. Retrieved from <http://www.ccsenet.org/journal/index.php>
- Linn, R. L. & Gronlund, N. E. (1995). *Measurement and Assessment in Teaching*. London: Prentice- Hall Inc.
- Madsen, H. S. (1983). *Techniques in Testing*. New York. Oxford University Press.
- Mahardika, S. H. (2018). *Investigating the Content Validity and Discrimination Power of English Summative Test (The Case of Tenth Graders of a Private Senior High School in Semarang in the Academic Year of 2016/2017)*. Faculty of Languages and Arts. Semarang: Unnes.
- Masuwai, A. (2016). *Evaluating the Face and Content Validity of a Teaching and Learning Guiding Principles Instrument (TLGPI): A Perspective Study of Malaysian Teacher Educators*. Faculty of Education and Human Development. Sultan Idris Education University. Retrieved from <http://journalarticle.ukm.my/9891/1/2x.geografia-si-mac16-azwani-edam>
- Miles, M. B. & Huberman, A. M. (1984). *Qualitative Data Analysis: A Source book of New Methods*. Beverly Hills, California: Sage Publications.

- Moritoshi, T. P. (2002). *Validation of the Test of English Conversation Proficiency*. Department of English. Edgbaston: University of Birmingham. Retrieved from <http://www.birmingham.ac.uk/Documents/collegeartslaw/cels/essays/mateflt/esldissertations/MoritoshiDiss.pdf>.
- Nawati, I. D. (2009). *Content Validity of the English National Writing Test for Junior High School Based on the Graduate Competence Standard of KTSP Curriculum*. Faculty of Languages and Arts. Semarang: Unnes. Retrieved from <https://lib.unnes.ac.id/4701/>
- Nawawi, H. (1993). *Metode Penelitian Bidang Sosial*. Yogyakarta: Gajah Mada University Press.
- Nofiyanti. (2011). *Analysis on the Content Validity of the Summative Test for the First Year Students of Junior High School*. Faculty of Tarbiyah and Teachers. Jakarta: State Islamic University Syarif Hidayatulloh. Retrieved from <repository.uinjkt.ac.id/.../1/103071-nofiyanti-fitk.pdf>.
- Nugroho, C. A. (2015). *A Comparative Analysis of Teacher-Made English School Final Exam of Ninth Graders of SMPN 17 Semarang and SMP Kesatrian 2 Semarang Academic Year 2013-2014*. Faculty of Languages and Arts. Semarang: Unnes. Retrieved from <http://lib.unnes.ac.id/20369/>
- Ozer, I., Fitzgerald, S., & Sulbaran, M. E. (2013). *Reliability and content validity of an English as a Foreign Language (EFL) grade-level test for Turkish primary grade students*. Kent State University. Kent: Elsevier Ltd. Retrieved from [www.sciencedirect.com](http://www.sciencedirect.com).
- Palacio, M., Gaviria, S., & Brown, J. D. (2016). Aligning English Language Testing with Curriculum. *Issues in Teachers' Professional Development*, 18(2), 63-77. Retrieved from <http://dx.doi.org/10.15446/profile.v18n2.53302>.
- Parsi, G. (2017). Beliefs on Learning and Teaching Language Components: The Case of Iranian EAP and EFL Learners. *Advances in Language and Literary Studies*, 8(3), 5-6. AIAC: Australia. Retrieved from <http://dx.doi.org/10.7575/aiac.all.v.8n.3p.5>
- Prasetyo, A. B. (2015). *Analysis of Content Validity in the English Final Examination in the MA Al Ma'arif*. English Educational Program. Tulungagung: State Islamic Institute (IAIN). Retrieved from [www.repo.iain-tulungagung.ac.id/2777/2/.com](http://www.repo.iain-tulungagung.ac.id/2777/2/.com)
- Putri, Y. F. (2014). *Analysis of Teacher-Made English Final Second Semester Test for the Year Eleven Students of SMAN 1 Ambarawa in the Academic Year of 2008/2009 Based on the Representativeness of Content Standard*. Faculty of

- Languages and Arts. Semarang: Unnes. Retrieved from <https://lib.unnes.ac.id/6520/>
- Rakhmah, F. (2009). *A Content Validity Analysis of English Summative Teacher-Made Test for the Eleventh Grade of SMA in Brebes*. Faculty of Languages and Arts. Semarang: Semarang State University. Faculty of Languages and Arts. Semarang: Unnes.
- Salwa, A. (2012). *The Validity, Reliability, Level of Difficulty and Appropriateness of Curriculum of the English Test (A Comparative Study of The Quality of English Final Test of The First Semester Students Grade V Made By English KKG of Ministry of Education and Culture and Ministry of Religion Semarang)*. Postgraduate Program of Linguistics. Semarang: Diponegoro University.
- Sireci, S. G. (2014). Validity Evidence Based on Test Content. *Psicothema*, 26(1), 100-107. Retrieved from <http://www.psicothema.com/pdf/4167.pdf>
- Siregar, S. H. (2014). *Content Validity of English Summative Test for Eleventh Grade Students Academic Year 2013/2014*. Jakarta: Unesa. Retrieved from [https://www.academia.edu/8745174/content\\_validity\\_of\\_english\\_summative\\_test\\_for\\_eleventh\\_grade\\_students\\_academic\\_year\\_2013\\_2014\\_by\\_sampai\\_haji\\_siregar\\_13783510?auto=download](https://www.academia.edu/8745174/content_validity_of_english_summative_test_for_eleventh_grade_students_academic_year_2013_2014_by_sampai_haji_siregar_13783510?auto=download).
- Sugiyono. (2010). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Tinambunan, W. (1988). *Evaluation of Student Achievement*. Jakarta: Depdikbud.
- Trisanti, N. (2017). Authentic Assessment Dilemma: Study on the English Teacher's Need Analysis of Assessment in the Classroom Context. *2<sup>nd</sup> Unnes-TEFLIN National Seminar*.
- Ur, P. (1996). *A Course in Language Teaching: Practice and Theory*. Cambridge: Cambridge University Press.
- Widdowson, H.G. (1983). *Learning Purpose and Language Use*. Oxford: Oxford University Press.
- Widyaningrum, F. A. D. (2014). *Content Validity and Authenticity of the 2012 English Test in the Senior High School National Examination*. Faculty of Teachers Training and Education. Yogyakarta: Sanata Dharma University. Retrieved from <http://e-journal.usd.ac.id/index.php/LLT/article/view/279>
- Wiersma, W. & Jurs, S. G. (1990). *Educational Measurement and Testing*. Boston: Allyn and Bacon.