UNNES
UNIVERSITAS NEGERI SEMARANG

INVESTIGATING THE AUTHENTICITY OF ENGLISH TRY-OUT READING TEST ITEMS

A Case of Ninth Graders of SMP Negeri 29 Semarang in the Academic Year of 2018/2019

Final project

submitted in partial fulfillment of the requirements for the degree

of *Sarjana Pendidikan* in English Language Education

by

Putri Primadani

2201415104

ENGLISH DEPARTMENT

FACULTY OF LANGUAGES AND ARTS

UNIVERSITAS NEGERI SEMARANG

2019

i

APPROVAL

This final project entitled *Investigating the Authenticity of English Try-out Reading Test Items (A Case of Ninth Graders of SMP Negeri 29 Semarang in the Academic Year of 2018/2019)* written by Putri Primadani has been approved by a team of examiners of the English Department of the Languages and Arts Faculty of Universitas Negeri Semarang (UNNES) on September 10th, 2019.

Board of Examiners:

1. **Chairman**

   Dr. Hendi Pratama, S. Pd., M. A
   NIP. 198505282010121006

2. **Secretary**

   Arif Suryo P, S. Pd., M. Pd
   NIP. 198306102010121002

3. **First Examiner**

   Dr. Dwi Anggani LB, M. Pd
   NIP. 195901141989012001

4. **Second Examiner**

   Pasca Kalisa, S. Pd., M. A., M. Pd
   NIP. 198909062013032106

5. **Advisor as Third Examiner**

   Novia Trisanti, S.Pd., M.Pd
   NIP. 197611062005012002

Approved by:

the Dean of Languages and Arts Faculty

Dr. Sri Rejeki Urip, M. Hum.
NIP. 196202211989012001

ii

DECLARATION OF ORIGINALITY

I am Putri Primadani declare that this final project entitled *Investigating the Authenticity of English Try-out Reading Test Items (A Case of Ninth Graders of SMP Negeri 29 Semarang in the Academic Year of 2018/2019)* 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 Information derived from the published and unpublished works of others has been acknowledged in the text and a list of references is attached in the references.

Semarang, August 2019

Putri Primadani
2201415104

MOTTO AND DEDICATION

Don't ever let anyone

make you feel like

you deserve less

than what you want

because you are capable

of everything

you want to achieve

and so much more.

This final project is dedicated to:

Allah SWT who always bless me

My honorable parents,

Tumadi and Agatha Endang Mujiyanti

My beloved brother and sister,

Bagus Ferdiyanto and Tiya Ardhiyati

My lovely man

And all my best friends

ACKNOWLEDGEMENT

First of all, all praises are presented to Allah SWT 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 *Investigating the Authenticity of English Try-out Reading Test Items (A Case of Ninth Graders of SMP Negeri 29 Semarang in the Academic Year of 2018/2019)*. This final project is submitted as the final requirement in accomplishing undergraduate program at English Education Study Program, English Department, Faculty of Literatures and Arts, Universitas Negeri Semarang.

In this occasion, I would like to express my gratitude and appreciation to Novia Trisanti, S.Pd., M.Pd as my advisor who has spent her precious time to guide and give excellent suggestions and corrections to me in preparing and completing this final project. Not to mention, my gratitude to the examiners, for the revising and advising to improve my final project better. My sincere thanks to the principal and teachers of SMP Negeri 29 Semarang, especially the English teacher MMVS Pangestuningrum, S.Pd.,MSi who have allowed me to carry out the investigation in their institution.

I would also like to give my deep appreciation to all the lecturers of the English Department who have taught and guided me during my entire study. My gratitude goes as well to my honorable parents, my sibling and his wife who have prayed ceaselessly, given their endless love, support, motivation and also reminded me to finish this final project.

I would also like to express my special gratitude to a man who always accompanies me since I was at fourth semester and gives all of his patience to support and cheer me up in writing this final project.

My sincere gratitude goes to all my best friends, Imma, Dian, Lintang, Rendra, Cituk, Hengki, Setiawan, Puput, Rifa, Muna, Irma, Anggun, Ratih, and Anggita who have never been tired to support and motivate me to finish this study. To all my best friends in English Education Program, who have spent memorable time for recent years. Thanks for keeping me happy and giving me motivation. Last, but not least, my very sincere gratitude for those who could not be mentioned one by one. Finally, I hope this final project may be useful for all the readers either for English teachers or English students, and for people who concern in educational field or anyone who need the information within it.

Semarang, August 2019

Putri Primadani

ABSTRACT

Primadani, Putri. *Investigating the Authenticity of English Try-out Reading Test Items (A Case of Ninth Graders of SMP Negeri 29 Semarang in the Academic Year of 2018/2019).* Final Project. English Department. Faculty of Languages and Arts. Universitas Negeri Semarang. Advisor: Novia Trisanti, S. Pd, M .Pd

Keywords: English try-out test, authenticity of the test tasks, authenticity of the test texts

This study is conducted to investigate the authenticity achieved in the English try-out reading test for ninth-grade students of SMP Negeri 29 Semarang in the Academic Year of 2018/2019. The objectives of this research are to find out the authenticity of the test tasks and the test texts in the English try-out reading test for ninth-grade students of SMP Negeri 29 Semarang in the Academic Year of 2018/2019. A qualitative research designed was used by the writer in this study. Descriptive method was chosen to analyze the data, and the result analysis was presented in descriptive form. The English try-out test consists of 40 items of multiple-choice and 27 reading passages as the researcher analyzed in reading section only. The multiple-choice items were analyzed to find out the authenticity of the test tasks. While the reading passages in the multiple-choice were analyzed to find out the authenticity of the test texts. The result of the study showed that the authenticity of the test tasks is 90.5% and the authenticity of the test texts is 83%. So, the result of the authenticity of English try-out test for ninth-grade students of SMP Negeri 29 Semarang in the Academic Year of 2018/2019 is 86.75%, it is categorized into relatively authentic. The representation of the real-world task indicator had become problematic aspect to reach the higher standard of the authenticity of the test tasks. And the provenance indicator also had become problematic aspect to reach the higher standard of the authenticity of the test texts. Based on the result, the writer suggested to the test-makers to more consider about the authenticity in designing the test. So, the test will be as authentic as possible.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

Chapter I consists of the following subchapters; background of the study, reasons for choosing the topic, research question, objective of the study, significance of the study, limitation of the study, and outline of the study.

1.1 Background of the Study

Language is a very important thing in our life. People use language as a means of communication. By using language, they can communicate with each other, express their feeling, show their purposes, thoughts, and minds to other people. Language served as a tool, a means of increasing the reach; power and effectiveness of individuals or social groups (Hamilton, 1990, p.20).

One of the many very important languages is English. English is the most important and influencing language, i.e. international languages that used in the way to communicate with each other around the world in many sectors such as politics, economics, work, business, and also education. Certainly, learning English is very important for most people in this world. Nowadays, this language takes an important role in the Indonesia Educational System, English subject has been put in the same line with Bahasa Indonesia, Mathematics, and Science for National Examination in Junior High School level. In Indonesia, National Examination plays an essential role, it is one of the requirements that students must pass. By passing the National Examination, students can graduate from a certain education level and continue their study to the next level. National

Examination also has a function to measure students' achievement in particular subjects which are tested. It means that the government can control the quality level of education using National Examination. This exam will give information that encourages and improves the quality of education continually. As stated in *Permbendikbud Nomor 4 Tahun 2018*, National Examination then called as UN, is an activity to measure graduate competency achievement at certain national subjects which refers to the Graduate Competency Standards. To face National Examination, students must prepare themselves, both material and psychological readiness. Nevertheless, there are many problems related to National Examination realization. One of them is students' readiness before taking National Examination. Most of the students are afraid to do the test because they think that the test is too difficult and it really determines their graduation. Especially in the English National Examination, students think that they are not familiar with the test and the use of the language in the test is different from their native language. Here, students' readiness to learn English becomes one thing that should be considered. It is because if students have learned, they will ready to face the exam and get a better result. However, in learning activities in the classroom, students still faced many obstacles due to the lack of readiness in learning English. According to Baden and Major (2004, p.27) "Piaget believed that the activities learners could complete matched views their cognitive aspects and their readiness". It means that the activities of learners related to cognitive aspects and their readiness. So their readiness will certainly affect the cognitive aspect, such as National Examination scores.

In order to prepare students' readiness of taking National Examination, many schools give their students additional courses, such as enrichment, try-out test, etc to add and measure their students' capability. Try-out test itself is a test which has the same level with National Examination and expected with. So, the test should pay attention in developing and constructing items and topics referring to particular specifications and lesson objectives to test students' competencies; the test should be well-prepared.

According to Brown (2004, p.19) a good test can be tested by measuring validity, reliability, practicality, authenticity, and wash back. In this research, the writer only focuses on authenticity analysis. Authenticity is used to investigate the extent to which score interpretations generalize beyond performance on the test to language use. Bachman and Palmer (1996, p.23) argued that authenticity as the degree of correspondence of the characteristic of a given language test task to the feature of a language task. The presence of authenticity in a test will affect test takers' performance. As stated by Bachman and Palmer (1996, p.24) "authenticity has a potential effect on test takers' performance". It means this effect is between those features of authenticity which should be demonstrated if we want to change the speculation about the nature of authenticity to a comprehensive characterization of the notion. Lynch (1982, p.11) also states that authenticity is an important quality for the test development. In this context, authenticity should be considered as an important thing in measuring students' achievement. Authenticity has two important parts; there are test task characteristics and test text characteristics. The test tasks refer to the test

instructions and the optional answers of the items while the test texts refer to the passages used in the test. Spence-Brown (2001, p.463) also stated that authentic tasks have become one of the most important components of tests that aim to measure language learners' ability to apply what they have learned from class to real world situations. In this regard, authenticity is more focused on the correlation between test takers and the test, and it also depends on how they interpret the test texts and the test tasks.

A good test also measures students' skills, abilities, and knowledge by presenting the appropriate test score and fulfill the intended language learning. It means that the test measures what is intended to be measured. Yet, there is often a large gap between students' ability to perform in tests and their ability to use the language in everyday real-life situations, it is because some of the tests lack authenticity. So, the writer decided to analyze the authenticity of the English try-out reading test items of SMP Negeri 29 Semarang in the Academic Year of 2018/2019. The English try-out test consists of 50 items in the form of multiple-choice (reading and writing section). Nevertheless, the researcher only wants to analyze the multiple-choice items and the reading passages on reading section. Finally, by analyzing the authenticity of the test, the test-makers can improve the quality of the test.

1.2 Reason for choosing the topic

In this study, the writer would like to focus on analyzing the authenticity of English try-out reading test items of SMP Negeri 29 Semarang in the Academic

Year of 2018/2019. Every test items in the try-out test should fulfill the criteria of good test items, particularly in the term of authenticity to select qualified test which must have the same quality with the National Examination.

One of the problems when constructing a try-out test is authenticity. The test-makers usually only make a test without considering about the authenticity of the test because their lack of knowledge of authenticity itself or probably because of they think that authenticity is not important to be considered in designing a test. The gap between the language tests and real-life language experience is also rarely bridged by the test-makers whereas a test should demand the students to perform the application of their knowledge, skill, and ability in a real-life situation. Here, the test-makers need to pay attention to this problem because if a test does not reflect the authenticity, students will be confused to use language in context. Besides, it is important that the materials in a test are culturally relevant to students' majors to ease students in understanding the test as well.

1.3 Research Question

Based on the background of the study explained above, the research question is as follow:

1. How is the authenticity of the English reading try-out test tasks of SMP Negeri 29 Semarang in the Academic Year of 2018/2019 in terms of:

    a. The naturalness of language used in the test

    b. The relevance of the test topics for the learners

    c. The representation of the real-world tasks

2. How is the authenticity of the English reading try-out test texts of SMP Negeri 29 Semarang in the Academic Year of 2018/2019 in terms of:

    a. Provenance

    b. Inherent textual qualities

1.4 Objectives of the Study

The objectives of this research are:

1. To find out the authenticity of the English reading try-out test tasks of SMP Negeri 29 Semarang in the Academic Year of 2018/2019 in terms of:

    a. The naturalness of language used in the test

    b. The relevance of the test topics for the learners

    c. The representation of the real-world tasks

2. To find out the authenticity of the English reading try-out test texts of SMP Negeri 29 Semarang in the Academic Year of 2018/2019 in terms of:

    a. Provenance

    b. Inherent textual qualities

1.5 Significance of the Study

Through this study, the writer hopes that the result of this study will give some advantages as follows:

1. Theoretically

This study provides the test-makers about the authenticity of the English reading try-out test items. The researcher hopes that the information could be a reference for the test-makers in designing a test.

So, they will pay more attention to designing a test, especially in authenticity.

2. Pedagogically

The researcher hopes that this information can help the test-makers in designing the good test items by concerning in some criteria of a good test, especially for the authenticity so that the test makers can prevent of making mistakes in constructing the test.

3. Practically

By this study, the researcher hopes to give practical information to the readers about the authenticity of the English reading try-out test items so the test-makers can make a good test.

1.6 Limitation of the Study

The researcher would like to put forward the following limitation to her study:

1. In this study, the researcher analyzed only the English reading try-out test items of SMP Negeri 29 Semarang in the Academic Year of 2018/2019.

2. The English reading try-out test consists of 40 items in the form of multiple-choice items. The researcher analyzed the test tasks and the test texts in terms of its authenticity.

3. The researcher used the theory by Brown (2004) to analyse the authenticity of the test tasks. Authenticity is determined by five indicators namely the naturalness of language used in the test, the contextualization

of the items, the relevance of the test topics for the learners, the presence of thematic test items organization, and the representation of the real-world task or sources. These five indicators build a test which is relevant for students' skills and represent an original situation to the test-takers. In this study, there are three indicators used, namely the naturalness of language used in the test, the relevance of the test topics for the learners, and the representation of the real-world task. Besides that, the researcher also used the theory by Will (2018) to analyze the authenticity of the test texts. Provenance, inherent textual qualities, aspects of post-production, use of the text (appropriate response). These four indicators are needed in order to classify text genres as authentic or non-authentic. But, the researcher only used two indicators, namely provenance, and inherent textual qualities.

1.7 Outline of the study

This study consists of five chapters. Chapter I is an introduction which consists of the background of the study, reasons for choosing the topic, research question, objective of the study, significance of the study, limitation of the study, and outline of the study.

Chapter II presents the review of related literature about this research such as review of previous study, review of theoretical studies, and theoretical framework.

Chapter III presents the method of collecting and analyzing data. It consists of research design, object of the study, role of the researcher, types of

data, unit of analysis, instrument for collecting data, procedure for collecting data, procedure for analyzing data, interpretation of data analyzing result, and triangulation.

Chapter IV presents the result of the study that has been conducted by the researcher. This chapter is divided into three sections. The first section is about the results of the analysis. This section is divided into two sub-sections. The first sub-section discusses the result of the authenticity of the test tasks analysis. The second sub-section consists the result of the authenticity of the test texts analysis. The second section is the interpretation of data analysis result. And the last section is the experts' judgment.

Chapter V presents the conclusions and suggestions based on the result of the study.

CHAPTER II

REVIEW OF RELATED LITERATURE

Chapter II is generally separated into three main sections. First part is the review of the previous studies. This part contains the preview studies which are related to the study and as the reference for conducting the research. The second is review of theoretical study that provides the theories supporting the research. The last section is theoretical framework of analysis which is the framework of how the analysis will be conducted.

2.1 Review of the Previous Studies

Regarding the topic, it is not a new one in writing a final project. In this part, the writer presents the review of the previous studies to describe some researches which have been conducted earlier related to the topic to be used as background knowledge and reference in this study. Related to this study, the writer chooses some literatures about previous researches which are relevant to the test analysis. From those studies, the writer also shows the different between the writer's research and those studies.

The first study wrote by Janet (2018) entitled *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)*. The objectives of this research were to find out the relevance of the English summative test items to the syllabus of curriculum 2013 and the authenticity of the test tasks and the test texts in the English summative test for tenth-grade students of a State Senior High

School in Kendal in the academic year 2015/2016. The result of the study showed that the content of English summative test for tenth-grade students of a State Senior High School in Kendal in the academic year 2015/2016 was 96% valid since most of the test items were relevant to the syllabus of curriculum 2013. Moreover, the result of the authenticity of the test tasks was 84% and for the authenticity of the test texts was 33%. So, the result of the authenticity of English summative test for tenth-grade students of a State Senior High School in Kendal in the academic year 2015/2016 was 59%.

The second study was conducted by Natalia (2013) entitled *Analysis of Authenticity of Teachers' Made Assessment and Its Contribution to Students' English Achievement in Elementary Schools in Buleleng Reegency.* The authenticity of teachers' made assessment was seen in terms of teachers' assessment planning, assessment implementation as perceived by teachers, teachers' assessment implementation as perceived by students. This study was intended in order to know the contribution of those aspects to students' English achievement. The research design for this study was Exploratory Mix-Method design, involving 33 English teachers. The result of data analysis showed that teachers' assessment planning, assessment implementation as perceived by teachers and students had sufficient level of authenticity. However, it was also found that the three aspects were integratedly gave significant contribution to students' English achievement.

The next study was carried out by Adnyayanti (2013) entitled *Analysis of Authenticity of Teacher's Made Assessment and Its Contribution to Students'*

English Achievement (A Study in Junior High Schools in Buleleng Regency). 35 Junior High School English teachers in Buleleng Regency were used as the subjects in this research. The findings demonstrated that teachers' made assessments have sufficient authenticity level of planning, have high authenticity level of assessment implementation as perceived by teachers, have sufficient authenticity level of assessment implementation as perceived by students, and is considered as having high level on the students' English achievement after being assessed with teachers' made assessment. The findings also showed that only assessment implementation as perceived by students and simultaneous predictors gave significant contribution to students' English achievement. Assessment planning and assessment implementation as perceived by teachers did not show significant contribution because there was found inconsistent finding between what teachers planned with what teachers have implemented. This study has positive contribution to the importance of considering the authenticity of assessment in teaching English as a foreign language.

The fourth study is from Purwandani (2017) entitled *The Analysis of The Authenticity of Authentic Reading Materials in Students' Text Book.* The objective of this research was to examine the compatibility of the current authentic reading materials in students' textbook with the criteria of authenticity. The subjects were twelve English teachers. The sources of the data were documents of three authentic reading passages which were taken from Pathway to English 2 by Erlangga Publisher. The instrument was the combination of criteria authenticity checklist proposed by McGrath (2002) and the criteria of good reading materials

proposed by Arias (2007). The result showed that the authentic reading materials in the selected textbook were compatible with the criteria of authenticity. This indicates that the contents of the texts are appropriate to use as teaching materials.

Then, the study conducted by Widyaningrum (2012) entitled *Content Validity and Authenticity of the 2012 English Test in Senior High School National Examination.* The study aimed to find out how valid the content of the English test items of National Examination year 2012 for senior high schools is and how authentic the English test items of National Examination year 2012 for senior high schools is. The data were obtained from the document and analysed by using checklists. Besides, to maintain the validity of the analysis, a triangulation was done by distributing a questionnaire to four experts in language assessment. There were two findings resulted from the analysis. First, the content of the National Examination year 2012 was 98.8% valid since almost all of the contents were relevant to the test specifications. However, there were three reading test versions which failed to represent kinds of texts, namely explanation text. Second, the National Examination year 2012 met the criteria of authenticity with percentage 79.5% since some listening and reading test items failed to conform to authenticity criteria. Natural language use, the relevance of the test topics, and real-world representativeness became problematic aspects to meet the higher standard of authenticity.

The sixth study is from Geribka (2016) entitled *A Study on The Content Validity and Authenticity of English National Examination 2015 for Junior High School*. The study aimed to find the content validity and authenticity of English

National Examination 2015 items. The research objects were the documents of the English National Examination. The instruments of this research were checklists. The method of the data analysing was matching method. To obtain results ralted to the research question, the researcher analysed the data by using checklist in form of checklist tables. This research demonstrated two findings, the content validity of the examination items reached 94% and the authenticity of the topics of the National Examination reached 89%.

The seventh study was done by Hapsari (2017) entitled *The Authenticity of Assessment Used in Reading Class at The Third Semester of English Education Program in Universitas Pekalongan Academic Year 2016/2017*. The objectives of the study were to describe the types of assessment used in raeding at the third semester in Universitas Pekalongan, to describe the most frequently used assessment in reading class, to know the authenticity of the assessment used by the lecturer in reading class, and to describe the problems faced by lecturer in applying authentic assessment in reading class. The result of the study showed that the lecturer used two types of assessment, namely performance in the form of oral presentation and teacher observation in the form of anecdotal records. Performance and teacher observation had high level of authenticity.

The next study was carried out by Nuryanto (2018) entitled *The Validity Analysis of Reading Test Items on National Standard School Final Examination for 12th Grade of Man 1 Semarang*. The objectives of this study were, to examine whether the test items of NSSFE are match the materials, to examine whether the test items of NSSFE are match the expected learning objectives, to examine

whether the test items of NSSFE meet the teachers' perception, to examine whether the test items of NSSFE meet the students' perception. The subjects of the study were three English teachers and 40 students of Natural science major from 12th grade of MAN 1 Semarang. The data sources used in this study were: documents, and informants. The techniques of collecting data used were document analysis, interview and questionnaire. The findings of this study showed that the test has fulfilled the criteria of having content validity since the findings of the current study were, the test items of NSSFE are match the materials, the test items of NSSFE are match the learning objectives, the test items of NSSFE meet the teachers' perception, and the test items of NSSFE meet the students' perception.

The ninth study was conducted by Apriliani (2013) entiled *Item Analysis of English Tryout Test in Detik-Detik Ujian Nasional Bahasa Inggris 3 SMP/MTs Published By Intan Pariwara.* The general objective of the research was obtaining an objective description of the structure of a good test item. The writer conducted two activities. The first was library activities, the writer selected some books which gave information, or supporting data for reference. Then the second was field activity, it was used to collect the data. The result showed there are validities of this test are good in face validity, good in content validity and good in construct validity. The reliability of the test is 0.87, so this test is perfect reliable. In the term of difficulty level this item categorized as easy item because the mean is 0.73. There are 2 difficult items, 19 desirable items, and 29 easy items. From the point of view of discrimination power, it can be concluded as good because the

mean of the discrimination power is 0.59. There are 10 excellent items, 38 good items and 2 satisfactory items.

The tenth study wrote by Cindrakasih (2017) entitled *Analysis of English Try-out Test of The Ninth Grade Students at MTSN Probolinggo.* The objectives of the study were to analyze the content validity, the reliability , the item distractor, the item difficulty and the discriminating power of the English try out test of the ninth grade students at MTSN Probolinggo. The subject of this study was 3 classes of the third grade students of MTSN Probolinggo with 91 students. The findings of the research showed that the content validity based on the basic competence are 96% appropriate and 4% unappropriate with the syllabus. The reliability in Packet 1 is moderate and the rest packets have high criteria. Packet 1, 2, 3, 4 and 5 has 26, 35, 31, 24 and 28 options of excellent distractor. The item difficulty in Packet 1, 2, 3, 4 and 5 is there were 14, 18, 10, 11 and 16 difficult items. In the result of SPSS 23.0 of discriminating power, Packet 1, 2, 3, 4 and 5 has 13, 19, 17, 26 and 15 very good items. The conclusion is the content validity based on syllabus is high however in SPSS, the content validity is low. The reliability is moderate and high. The item difficulty is low. The item distractor is high and the discriminating power is low.

There was also another research about authenticity that had been written by Lewkowicz (2000) entitled *Authenticity in language testing: some outstanding questions.* This study took a look at the concept of authenticity and the way this notion has evolved in language testing and also at the importance of authenticity for test takers. A group of 72 first-year students from the University of Hong

Kong were identified for this study, they were given two language tests within a period of three weeks. The test administered first to all the students was a 90-item multiple-choice test based on a TOEFL practice test and the second test administered was an EAP (English for academic purposes) test. The results of the study showed that valuable insights can be obtained about authenticity from test-taking informants and that their response to test tasks may be much subtler and more pragmatic than testers might prefer to believe and the results also raised the issue of the importance accorded to authenticity in the literature. They showed that test takers' perceptions of authenticity vary.

The next research paper was written by Qichun (2005) entitled *The Analysis of Authenticity in Language Test.* This study focused on analysis of one of major test qualities – authenticity. It has been placed upon the theoretical analysis and operational practices that differences in test purposes have influenced the authenticity, it also included the best time to take stock of the authenticity factor and how to keep balance among the test qualities affecting the usefulness of the test. The result of the study was the overall test quality or test usefulness was determined by the appropriate balance among all the six test qualities. The authenticity of test depends to a great extent on the test constructors understanding and knowledge of the characteristics of test takers, so any test which goes beyond the scope of test takers topical language knowledge and abilities fails to be an authentic test.

The next study was written by Joy (2011) entitled *The duality of authenticity in ELT.* This study aimed to find out the dual-perspectives on

authenticity in English language teaching. The problem of the study was that the concept of authenticity has generated the quite a ripple in ESL/EFL classrooms. The method used in this study was a survey of literature has been done to trace the genesis of authenticity and its various facets. And the result found that a balanced outlook on authenticity is required by the teachers to promote better teaching-learning experiences.

Another research that discussed the analysis of authenticity was a research by Shomoossi and Ketabi (2007) entitled *A Critical Look at the Concept of Authenticity.* This article is intended to briefly overview the concept of authenticity which cannot be defined similarly for all contexts. In other words, authenticity is a relative concept, materialized within each context with the interaction of its participants. Authentic materials require at least two features: communicative potential and relevance. Therefore, literature on authenticity suggests that authenticity is subject to pragmatic variation which cannot be defined in vacuum and that its defining characteristics lie in the context. This argument has implications for teacher education and material development because of the changing situation of ELT practice around the world. The final conclusion is that the notion of authenticity within the global context must be considered in the light of the pragmatic appropriateness of the materials used and the interaction tasks set in relation to learners' needs and interests. The knowledgeable teacher is the nexus for empowering materials and tasks that are 'authentic'

The next study was undertaken by Wang (2008) entitled *Authenticity in Language Testing*. This paper begins with a brief overview of authenticity in language testing. This is followed by a discussion of several characteristics of authenticity in particular types of tests. The final conclusion of this study was authenticity is one of the crucial components that test designers should include if the goal of testing is to reflect test takers' actual language competence and abilities. Test designers should understand test takers' goals of learning a language, clarify the objectives of their tests, investigate the real situations on which their tests are based, and experiment with the authenticity and practicality of the tests in order to avoid the negative backwash caused by assessment tasks.

The next research paper was written by MacDonald, Malcolm, Badger, Richard and Dasli, Maria. (2006) entitled *Authenticity, Culture and Language Learning*. This paper reviews four key conceptualisations of authenticity which have emerged in the field of applied linguistics: text authenticity, authenticity of language competence, learner authenticity and classroom authenticity. Text authenticity can lead to a poverty of language; authenticity of competence can lead to a poverty of performance; learner authenticity can lead to a poverty of interpretation; classroom authenticity can lead to a poverty of communication. This paper proposes that a pedagogy of intercultural communication be informed by a more hybrid view of authenticity as a process of subjectification, derived from the Heideggerian concept of self-concern.

Another research paper was written by Shomoossi and Tavakoli (2010) entitled *Authenticity and Authentication in Language Testing: an Operational*

Perspective. This paper reviews the concept of authenticity in language testing and intends to find a balance between theoretical and operational levels through the concept of authentication. The conclusion of this paper are the authenticity as a test feature may be side-lined by a more important notion in test validity, i.e. the authentication process, the notion of authenticity within the global context must be considered in the light of the pragmatic appropriateness of the materials used and interaction tasks set in relation to learners' needs and interests. The knowledgeable teacher is the nexus for empowering materials and tasks that are authentic for specific groups of learners.

The next research paper was written by Alkubaidi (2009) entitled *Authenticity in Language Test Design.* This paper suggests a possible definition of authenticity wherein a test may be seen to be authentic in terms of language testing. It is possible to design an authentic test by the definition proposed. This paper argues and defends its definition by various readings, and support the notion of the impossibility to create a truly authentic test. The conclusion of this paper is authenticity in test design cannot be fully realized; however, we may attempt to authenticate a test through a progressive continuum of a series of testing so that a single test is not the only source of assessment. In each test, specific constructs should be designed to measure the 'can dos'. Additionally, to come close to real-life situations as well as to simulate natural responses from learners, primary resource material should be intertwined with the learning process to produce the required register for assessment.

The next study was written by Doye, Peter (1991) entitled *Authenticty in Foreign Language Testing*. This paper is about the congruence between the test situation and the real-life situation the learner is expected to master. Some language educators make authenticity a major criterion of test quality. However, complete congruence of test and real-life situation is impossible. The solution is to find a reasonable balance between authenticity abstraction in tests. Pragmatics, with its analyses of speech acts and their characteristics, can be helpful in finding the right degree of abstraction for testing.

The last study was written by Gilmore (2007) entitled *Authentic Materials and Authenticity in Foreign Language Testing*. This article reviews some of the wide-ranging issues and research surrounding authentic materials and authenticity in foreign language learning. The paper discusses four important areas of concern: i) the gap between authentic and textbook discourse; ii) the English-as-a-world-language debate; iii) authenticity and motivation; iv) text difficulty and task design and their effects on language acquisition.

All of the research studies above are about authenticity analysis. Those studies have some similarities in the theory which is used. But, the study presented in this final project is different from the previous studies. In this study, the writer conducts her research only on analysis of authenticity in the English reading try-out test items for the ninth grade students.

2.2 Review of the Theoretical Studies

In this subchapter, the writer explains about the theoretical background of the study, those are:

2.2.1 Testing, Assessment and Teaching

Testing, assessment, and teaching are three basic interrelated concepts. The relation among testing, assessment, and teaching are found in a diagram below:
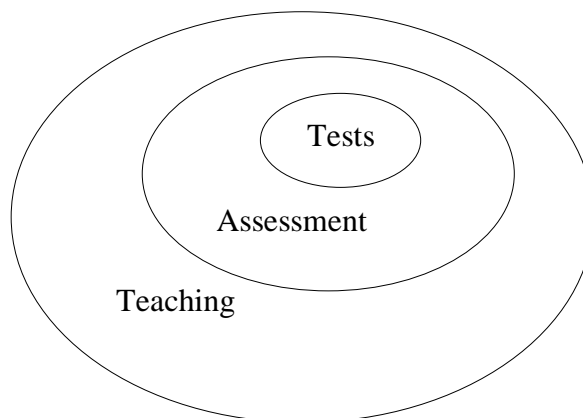
Figure 2.1 Tests, assessment, and teaching

Based on the diagram above, Brown (2004, p.5) shows that teaching overlaps assessment, measurement, and tests in order to evaluate learners' competence. Teaching sets up the practice games of language learning: the opportunities for learners to listen, think, take risks, set goals, and process feedback from the "coach" and then recycle through the skills that they are trying to master.

As stated in Brown (2004, p.4) assessment, on the other hand, 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 students' performance. Bachman (2004, p.7) also stated assessment can be thought of broadly as the process of collecting information about a given object of interest according to procedures that are systematic and substantively grounded. A product, or outcome of this process, such as a test score or a verbal description, is also referred to as an assessment.

2.2.2 *Test*

As reported by Brown (2004, p.3) a test in simple terms, is a method of measuring a person's ability, knowledge, or performance in a given domain. A test is first a method. It is an instrument – a set of techniques, procedures, or items – that requires performance on the part of the test-taker. To qualify as a test, the method must be explicit and structured. Second, a test must measure. Some tests measure general ability, while others focus on very specific competencies or objectives. Next, a test measures an individual's ability, knowledge, or performance. Testers need to understand who the test-takers are, their previous experience and background, that they can determine appropriate kind of test to their ability and how to interpret their scores. Fourth, a test measures performance, but the results imply the test-taker's ability, or, to use a concept common in the field of linguistics, competence. Most

language tests measure one's ability to perform language, that is to speak, write, read, or listen to a subset of language but the others can also infers general competence. Finally, a test measures a given domain. In the case of a proficiency test, even though the actual performance on the test involves only a sampling of skills, that domain is overall proficiency in a language-general competence in all skills of a language. A well-constructed test is an instrument that provides an accurate measure of the test-taker's ability within a particular domain.

Bachman (1995, p.20) also claimed that a test is "a measurement instrument designed, to elicit a specific sample of an individual's behavior". As one type of measurement, a test necessarily quantifies characteristics of individuals according to explicit procedures. What distinguishes a test from other types of measurement is that it is designed to obtain a specific sample of behavior.

Based on the definitions of test above, it can be concluded that a test is a standardized device to measure students' achievement after getting materials from the previous meetings given by teacher.

2.2.3 Type of Tests

(1) Based on the Objective of the Test

According to Brown (2004, p.43), there are five types of tests:

a. Language Aptitude Tests

Brown (2004, p.43) stated a language aptitude test is designed to measure capacity or general ability to learn a foreign language and ultimate

success in that undertaking. Language aptitude tests are ostensibly designed to apply to the classroom learning of any language. Two standardized aptitude tests have been used in the United Stated are the Modern Language Aptitude Test (MLAT) and the Pimsleur Language Aptitude Battery (PLAB).

b. **Proficiency Tests**

Brown (2004, p.44) stated a proficiency test is not limited to any one course, curriculum, or single skill in the language; rather, it tests overall ability. Hughes (2003, p.11) noted proficiency tests are designed to measure people's ability in a language, regardless of any training they may have had in language. Proficiency tests have traditionally consisted of standardized multiple-choice items on grammar, vocabulary, reading comprehension, and aural comprehension. Proficiency tests are almost always summative and non-referenced. They provide results in the form of a single score, which is a sufficient result for the gate-keeping role they play of accepting or denying someone passage into the next stage of a journey. A typical example of a standardized proficiency test is the Test of English as a Foreign Language (TOEFL). It is widely used as a proficiency test, which is used to measure the English language proficiency of foreign students who wish to study in the United States.

c. **Placement Tests**

Placement tests are used to place a student into a particular level or section of a language curriculum or school. A placement test usually

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. Placement test come in many varieties: assessing comprehension and production, responding through written and oral performance, open-ended and limited responses, selection (e.g. multiple choice) and gap-filling formats, depending on the nature of a program and its needs. The ultimate objective of a placement test is to correctly place a student into a course or level and the secondary benefits to consider include face validity, diagnostic information on students' performance, and authenticity. (Brown, 2004, p.46)

d. Diagnostic Tests

A diagnostic test is designed to diagnose a specified aspects of a language. Hughes (2003, p.15) told that diagnostic tests are used to identify learners' strengths and weaknesses. They are intended primarily to ascertain what learning still needs to take place. For example, a test in pronunciation might diagnose the phonological features of English that are difficult for learners and should therefore become part of a curriculum. Writing would elicit a writing sample from students that would allow the teacher to identify those rhetorical and linguistic features on which the course needed to focus special attention. Diagnostic tests should elicit information on what students need to work on in the future. (Brown, 2004, p.46-47).

e. **Achievement Tests**

According to Brown (2004, p.47-48) achievement test is related directly to classroom lessons, units, or even a total curriculum. Achievement tests are limited to particular material addressed in a curriculum within a particular time frame and are offered after a course has focused on the objectives in question. Hughes (2003, p.13) stated that achievement tests are directly related to language courses, their purpose being to establish how successful individual students, groups of students, or the courses themselves have been in achieving objectives. The primary role of an achievement test is to determine whether course objectives have been met and appropriate knowledge and skills acquired by the end of a period or instruction. Achievement tests are often summative because they are administered at the end of a unit or term of study. They also play an important formative role. An effective achievement test will offer washback about the quality of a learner's performance in subsets of the unit or course. This washback contributes to the formative nature of such tests.

(2) Based on the Function the Test Gives

Bloom in Harrys and Vallete (2003, p.10-11) divides the type of the test as follows:

a. **Formative Test**

This test is given during the course of instruction; its purpose is to show which aspects of the chapter the student has mastered and where remedial

work is necessary. The formative test is usually graded on a pass-fail basis, and students who fail are given the opportunity to study and then take the test again.

b. Summative Test

The summative test is usually given at the end of a marking period and measures the "sum" total of the material covered. On this type of a test, students are usually ranked and graded.

Brown (2005, p.1-2) has different idea in dividing the test based on its function. They are:

a. Norm-Referenced Test (NRT)

The test is designed to measure global test ability. In norm-referenced tests, each test-taker's score is interpreted in relation to a mean (average score), median (middle score), standard deviation (extent of variance in scores), and/or percentile rank.

b. Criterion-Referenced Test (CRT)

The test is usually produced to measure well-defined and fairly specific instructional objective. Criterion-referenced test are designed to give test-takers feedback, usually in the form of grades, on specific course or lesson objectives.

(3) Based on the Way the Test is Scored

a. Objective Test

Objective test is one for which there is a specific correct response; therefore, whether the item is scored by one teacher or another, whether it

is scored today or last week, it is always scored the same way. (e.g multiple-choice, fill-in-the-blank and short-answer questions).

b. Subjective Test

Subjective test is one that does not have a single right answer. A short composition or an impromptu interview may be scored in different ways by different teachers, and even by the same teacher scoring the answer twice under different circumstances.

(4) Based on the Constructor of the Test

According to Harrys and Valette (2003, p.1) there are two types of test:

a. Standardized Test

It is formal, large-scale, standardized instruments which are prepared by professional testing services to assist institutions in the selection, placement, and evaluation of students.

b. Teacher-Made Test

It is generally prepared, administered, and scored by one teacher. In this situation; test objectives can be based directly on course objectives, and test content derived from specific course content.

2.2.4 Type of Test Items

There are some types of test items that stated by Brown (2004, p.194) in the following:

1. Multiple-Choice

By far the most popular method of testing a reading knowledge of vocabulary and grammar is the multiple-choice format, mainly for reasons of practicality: it is easy to administer and can be scored quickly. The most straightforward multiple-choice items may have little context, but might serve as a vocabulary or grammar check.

Multiple choice items are made up of an item stem, or the main part of the items at the top, a correct answer, which is obviously one choice, and distractor which are those choice that will be counted as incorrect (Brown, 2004, p.48). Multiple-choice or selection items types were developed to overcome a number of the weakness of the composition test that we noted earlier. The multiple-choice item consists of (1) a *stem* or *lead*, which is either a direct question or an incomplete statement, and (2) two or more choices or responses, of which one is the answer and the others are *distracters* - that is the incorrect responses (Harrys and Valette, 2003, p.7).

According to Brown (2004, p.56):

1. Multiple-choice items are all respective, or selective, response items in that the test-takers chooses from a set of responses rather than creating a response.

2. Every multiple-choice item has a stem, which presents a stimulus, and several options or alternatives to choose from

3. One of those options, the key, is the correct response, while the others serve as distracters.

2. Matching

The test-taker's task is simply to respond correctly, which makes matching an appropriate format. Matching tasks have the advantage of offering an alternative to traditional multiple-choice items or fill-in-the-blank formats and are sometimes easier to construct than multiple-choice items, as long as the test designer has chosen the matches carefully. Some disadvantages do come with this framework, however. They can become more of a puzzle-solving process than a genuine test of comprehension as test-takers struggle with the search of the match, possibly among 10 or 20 different items.

3. Gap-filling

An extension of simple gap-filling tasks is to create sentence completion items where test-takers read part of a sentence and then complete it by writing a phrase. The obvious disadvantage of this type of task is its questionable assessment of reading ability. The task requires both reading and writing performance, thereby rendering it of low validity in isolating reading as the sole criterion.

4. Cloze

It is, the ability to fill in gaps in an incomplete image (visual, auditory, or cognitive) and supply (from background schemata) omitted details. Cloze tests were developed for native language readers and defended as an appropriate gauge of reading ability. Cloze tests are usually a minimum of two paragraphs in length in order to account for

discourse expectancies. They can be constructed relatively easily as long as the specification for choosing deletion and for scoring are clearly defined. Typically every seventh word (plus or minus two) is deleted (as known as fixed-ratio deletion), but may cloze test designers instead use a rational deletion procedure of choosing deletions according to the grammatical or discourse functions of the words. Rational deletion also allows the designer to avoid deleting words that would be difficult to predict from the context.

2.2.5 Characteristics of a Good Test

According to Brown (2004, p.19) all good tests possess five qualities: practicality, reliability, validity, authenticity, and washback.

a. Practicality

A test should be practical across time, cost, and energy. Dealing with time and energy, tests should be efficient in terms of making, doing, and evaluating. Then, the tests must be affordable. It is quite useless if a valid and reliable test cannot be done in remote areas because it requires an inexpensive computer to do it (Heaton, 1975:158-159; Weir, 1990: 34-35; Brown, 2004: 19-20). Brown in his book *Language Assessment Principles and Classroom Practice clearly* also mentioned that an effective test is practical. This mean that it: is not excessively expensive; stays within appropriate time constraints; is relatively easy to administer; and has a scoring or evaluation procedure. A test is impractical if it is prohibitively expensive, consumes more time and money than necessary to

accomplish its objective, and takes a few minute for a student to take, while several hours for an examiner to evaluate.

b. *Reliability*

A reliable test is consistent and dependable (Brown, 2004, p.20). It means that the test-result of the test takers or students should yield similar with the result that they get in another occasion. Factors affecting reliability are (Brown, 2004: 21-22):

1. **Student-related reliability**: students personal factors such as motivation, illness, anxiety can hinder from their 'real' performance.

2. **Rater reliability**: either intra-rater or inter-rater leads to subjectivity, error, bias during scoring tests.

3. **Test administration reliability**: when the same test administered indifferent occasion, it can result differently.

4. **Test reliability**: dealing with duration of the test and test instruction. If a test takes a long time to do, it may affect the test takers performance.

c. *Validity*

When teachers want to assess their students, they will deal with a question "how to measure students' ability". According to Harrys and Valette (2003, p.19) in the selection of any test, two questions must always considered: (1) What precisely does the test measure? And (2) How well does the test measure?. It implies that teachers should be able to design a measurement to bring up students' potentials as they wish. It is called

validity and a good test should be valid. Brown (2004, p.22) defines validity as the extent to which inferences made from assessment results are appropriate, meaningful, and useful in terms of the purpose of the assessment. Gronlund and Waugh (2009, p.46) also state that validity is concerned with the interpretation and use of assessment results. From these definitions, it can be inferred that when a test is valid, it can elicit students' certain abilities as it is intended to. The valid test can also measure what it is supposed to measure..

d. *Authenticity*

Authenticity is the fourth major principle in language testing. It is a concept that is a little slippery concept to define in evaluating and designing tests. According to Bachman and Palmer (1996, p.23), authenticity as the degree of correspondence of the characteristics of a given language test task to the features of a target language task, and then suggest an agenda for identifying those target language tasks and for transforming them into valid test items. Authenticity in a test task is likely to be enacted in the "real world" because many test item types fail to simulate real world tasks.

e. *Washback*

In large scale assessment, washback generally refers to the effects the tests have on instruction in terms of how students prepare for the test. Washback also occurs in the information that washes back to students in the form of useful diagnoses of strengths and weaknesses. According to

Brown (2004, p.29) the challenge to teachers is to create classroom tests that serve as learning devices through which washback is achieved. Students' incorrect responses can become windows of insight into further work. Their correct responses need to be praised. Teachers can suggest strategies for success as part of their "coaching" role. Washback enhances a number of basic principles of language acquisition: intrinsic, motivation, autonomy, self-confidence, language, ego, interlanguage, and strategic investment, among others.

2.2.6   *Authenticity*

There are many definitions of authenticity stated by the experts. According to Bachman and Palmer (1996, p.23) authenticity as the degree of correspondence of the characteristics of a given language test task to the features of a target language task. Brown (2004, p.28) also argued that authenticity in a task should stimulate real word tasks. It means that a test should cover materials which are related to the real-word. Lynch (1982) said authenticity is an important quality for the test development. Authenticity is more focused on the connection between the test takers and the test itself, and it depends on how the individual interprets the test tasks, which will affect both the validity and reliability of test scores (Bachman, p.10). Widdowson (1990) defined authenticity was a quality of the outcome present if the audience could realize the author's intentions which would only be possible where the audience was aware of the conventions employed by the writer or speaker. Bachman and Palmer (1996) also

stated that authenticity is an important quality for language tests for two reasons. The first is because it provides a link between test performance and the Target Language Use (TLU) tasks and domain to which we want to generalize. The target language use of the English try-out test is American English and British English. The second is because the way test takers perceive the relative authenticity of test tasks can potentially facilitate their performance. In this case, original materials or topics are absolutely needed. Based on the definitions of authenticity above, it can be seen that authenticity has become more important in a test. Authenticity is an important quality of language test development that shows the real-world sources and true language use.

In this research, the criteria of authenticity set by Brown (2004) and Will (2018). Both of these theories are used to examine the authenticity of English try-out reading test items of SMP Negeri 29 Semarang in the Academic Year of 2018/2019, such as the language in the test is as natural as possible, items are contextualized rather than isolated, topics are meaningful (relevant, interesting) for the learner, some thematic organization to items is provided, tasks represent or closely approximate real-world tasks. These are used to analyze the test tasks. While provenance, inherent textual qualities, aspects of post-production, use of the text (appropriate response) are used to analyze the test texts.

The naturalness of language used in reading test items includes linguistic facets, such as typography, lexis, morphology, word orders and

grammar (syntactic matters), diction, and semantics. It shows the appropriateness of the test language to the target language in order to avoid test makers' confusion in understanding the test instructions.

The contextualization of the test items means that the test items are related to the existence of some thematic organization items. Contextualization is putting language items into a meaningful and real context rather than being treated as isolated items of language for language manipulation practice only. Contextualizing language tries to give real communicative value to the language that learners meet.

The relevance of the test topics and the learners means that the materials should be appropriate to learners, fulfil the core competence, the basic competence, and relevant to the objective of the material. Many of authentic passages have difficult level of language which make difficult for learners who have low level of language ability.

Some thematic organization to items is provided means that the test-makers should organize the test items orderly.

The last indicator to analyse the test tasks is the test tasks should represent the real-world task, it means that authentic materials are taken from real-world sources. The test tasks should be designed to obtain certain information related to the text rather than asking grammatical form or lexical items.

Provenance can lead to a broad definition that is based on the producer's communicative intention, which means the intention to convey

a real message. There are three elements to determine the provenance of a text: real source, real-life text, and native-like or natural language.

Inherent textual qualities means that a text produced for language learning that looks authentic by virtue of idiomatic features or a certain range of vocabulary. All of the texts fulfill inherent textual qualities if they were grammatically correct, and also had good cohesive and coherence.

Aspects of post-production means the aspect after the production which must be happen, that is simplification of text. It is used to make texts suitable to the learners' level of language proficiency without changing the original text.

The last indicator to analyse the test texts is use of the text (appropriate response). It means texts should reflect the original communicative purpose on which they are based, so it can activate learners' existing knowledge of the target language and culture.

To be authentic, the passage used in the test which covers topics should meet the criteria of authenticity. In this study, the writer only uses three indicators of authenticity to analyze the English try-out reading test tasks. Namely, the naturalness of language used in the test, the relevance of the test topics, and the representation of the real-world tasks. And also two authenticity indicators to analyze the English try-out reading test texts. Namely, provenance and inherent textual qualities.

2.2.6.1 Task Authenticity

Nunan (1989, p.5) has defined task is a communicative work in which students are involved in comprehending, manipulating, producing, or interacting in the target language while their attention is principally focused on meaning rather than form. Newman and Wehlage (1993) have said that tasks prepare students on activities they encounter in the office, at home, or any other social contexts. A task is authentic when it related to real-world relevance and provides opportunities for language use in such a way that students waive inhibitions while using the language. The opportunities for language use include taking right decisions, presenting persuasive opinions, filling in the missing data and resolving problems or it is also known as reasoning-gap, opinion-gap, information-gap, and problem-solving tasks.

2.2.6.2 Text Authenticity

Authentic material gives opportunities for students to see, experience, and perform the functions of the text which are realistic in nature. Nunan (1989, p.99) said that authentic material is produced for purposes other than to teach language. Crossley, Louwerse, McCarthy, and McNamara (2007) have defined all text-based articles intended for the target audience as authentic texts. Here, texts are sorted from many and different sources, countries, and it adapts to suit the needs and appropriateness for the test takers. According to Widdowson (1994, p.386) every text is produced in a specific context and the very act of extracting a

text from its original source, even if it is left in its entirety, could be said to 'disauthenticate' it since authenticity is 'non-transferable'.
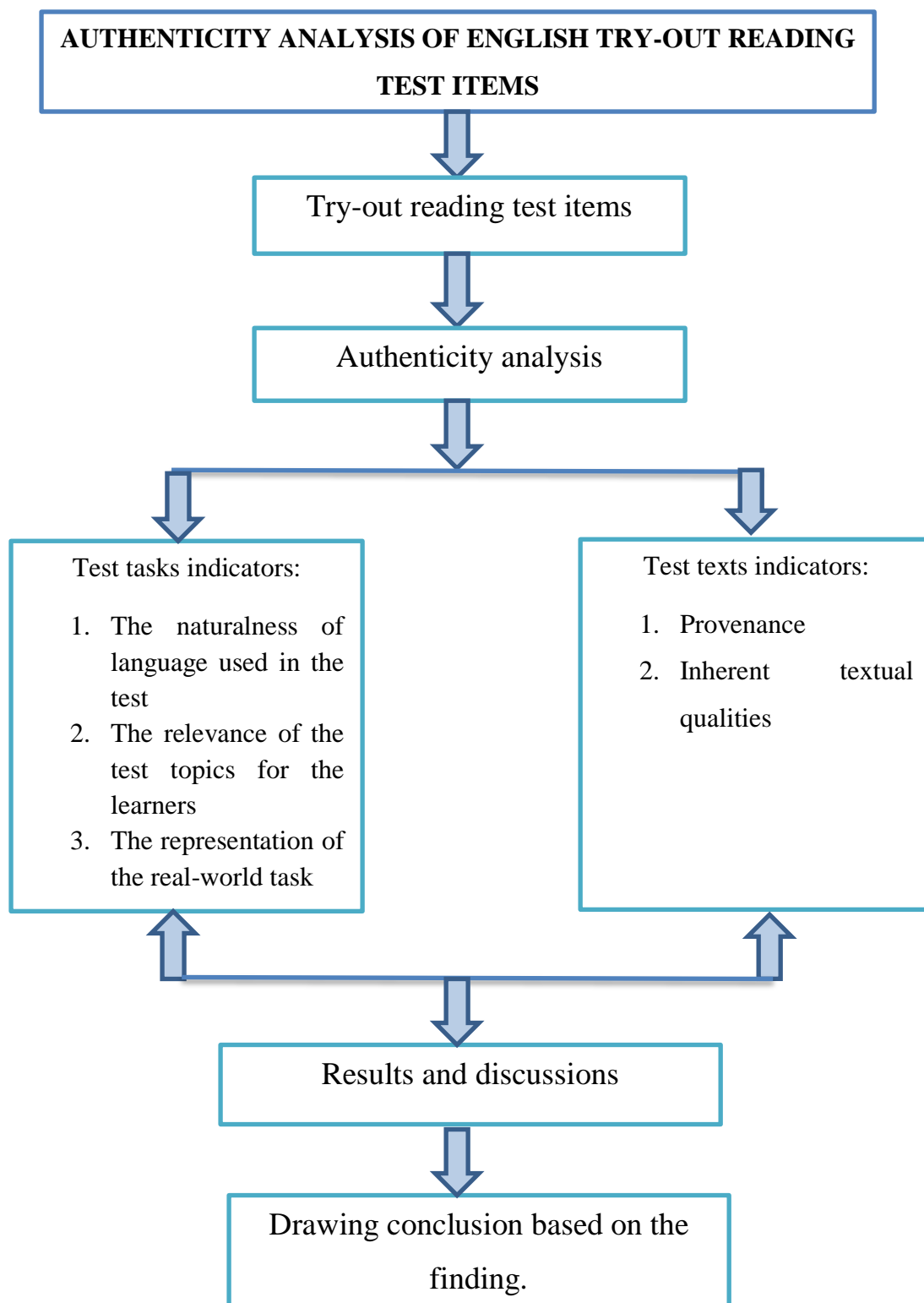
2.3 Theoretical Framework

Brown (2004, p.3) argues that a test is a method of measuring a person's ability, knowledge, or performance in a given domain. Test is used to measure the achievement or understanding of the students after being taught by the teacher in the previous meetings. It can be said that a test is a main instrument used by the teacher to gather information of the students in their achievement. In order to measure and obtain students' achievement, a test should cover five principles of assessment, there are practicality, reliability, validity, authenticity, and washback. These five principles should be applied in developing and constructing tests especially language tests since the language tests, especially English test is employed in National Examination. National Examination is categorized into achievement test because the test is given at the end of the course (Alderson, 2003, p.12).

In Indonesia, students are prepared to do a try-out test before taking National Examination. Try-out test itself is a test which has same level with National Examination and expected with. In other words, the try-out test is a test to measure the students' readiness before they take the National Examination. In measuring students' readiness, the try-out test should have good criteria so that the test can inform students' capabilities.

In this study, the writer purposed to find out the quality of the English try-out reading test items of SMP Negeri 29 Semarang in the Academic Year of

2018/2019. The writer analyzed the quality of the test in terms of the authenticity with the authenticity indicators set by Brown (2004) and Will (2018) to prove the authenticity of the English try-out reading test items.

To simplify the explanation about the analysis framework of this study, the writer created a chart of analysis framework that can be seen in the figure 2.2 below:

Figure 2.2. Analysis framework of study


```
AUTHENTICITY ANALYSIS OF ENGLISH TRY-OUT READING
                  TEST ITEMS
                      ↓
          Try-out reading test items
                      ↓
             Authenticity analysis
                      ↓
```

Test tasks indicators:

1. The naturalness of language used in the test
2. The relevance of the test topics for the learners
3. The representation of the real-world task

Test texts indicators:

1. Provenance
2. Inherent textual qualities

```
             Results and discussions
                      ↓
        Drawing conclusion based on the
                   finding.
```

CHAPTER V

CONCLUSIONS AND SUGGESTIONS

Chapter V presents the conclusions based on the result of the study. Furthermore, some suggestions are also provided in this chapter for those who concerned in the education field.

5.1 Conclusions

There are two objectives of this study. First, to find out the authenticity of the English try-out reading test tasks of SMP Negeri 29 Semarang in the Academic Year of 2018/2019 in terms of the naturalness of language used in the test, the relevance of the test topics for the learners, and the representation of the real-world tasks. Second, to find out the authenticity of the English try-out reading test texts of SMP Negeri 29 Semarang in the Academic Year of 2018/2019 in terms of provenance and inherent textual qualities. After conducting this study, it could be concluded that most of the test items are authentic since the authenticity of the test tasks analysis reached 90.5% and the authenticity of the test texts analysis reached 83%. So, the average of the authenticity of the test tasks and the test texts analysis is 86.75%, it is categorized into relatively authentic. In short, the researcher makes the conclusions of this study.

Based on the results of the authenticity analysis of the English try-out reading test items of SMP Negeri 29 Semarang in the Academic Year of 2018/2019, it could be drawn the conclusions as follows:

5.1.1 *The Authenticity Results of the Test Tasks Analysis*

a. In the naturalness of language used indicator, it was found that there were 32 of 40 tasks that fulfil all of the criteria, 7 of 40 tasks that fulfil two criteria, and one of 40 tasks that fulfil one criterion. In this indicator, it reached 92.5% of percentage.

b. In the relevance of the test topics indicator, it was found that all of the tasks fulfil all of the criteria (100%). It means that the test makers made the English reading try-out test based on the Graduate Standard Competence or *SKL* which is designed based on Curriculum 2013.

c. In the representation of the real-world tasks, it was found that there were 12 of 40 tasks that fulfil all of the criteria, 27 of 40 tasks that fulfil three criteria, and one of 40 tasks that fulfil two criteria. In this indicator, it reached 81.8% of percentage.

5.1.2 *The Authenticity Results of the Test Texts Analysis*

a. In the provenance indicator, it was found that there were 8 of 27 texts that fulfil all of the criteria, and 19 of 27 texts that fulfil two criteria. In this indicator, it reached 77.78%.

b. In the inherent textual qualities, it was found that there were 22 of 27 texts fulfil all of the criteria, and 5 of 27 texts that fulfil only one criterion. in this indicator, it reached 90.7%.

5.2 Suggestions

Concerning from the conclusion above, the researcher would like to offer some suggestions to all of the teacher and the test-makers. First, the test-makers

should more pay attention in typing the test tasks and the test texts because there are still some linguistic mistakes that will make the test-takers feel confused. Such as grammatical mistake, typographical mistake, morphological mistake, etc. Second, the test texts will be more authentic and understandable by the test-takers if the test texts are taken from the country that the test-takers are come from. Third, the test-makers should put the real-source of the test tasks and the test texts.

Lastly, from the conclusions and suggestions that have been explained above, the writer hopes that the next test especially for the English try-out test in SMP Negeri 29 Semarang is better and can fulfill all of the authenticity requirements of a test. Besides, the writes hopes that this result of the analysis could be used as an example and reference in the preparation of the next test and for future researchers who have an idea to analyse the same topic as mine, which is about authenticity. The writer suggested for the future researchers to expand this study with a deeper analysis to find out the authenticity of a test and hopefully the data will be more valid. Then, the future researchers are better to involve some experts for assessing the data and the results.

REFERENCES

Adnyayanti, N. L. P. E., Marhaeni, A. A. I. N., & Artini, L. P. (2013). *Analysis of Authenticity of Teachers' Made Assessment and Its Contribution To Students' English Achievement (A Study in Junior High Schools in Buleleng Regency), 1*. Language Education Department. Singaraja: Universitas Pendidikan Ganesha. Available at http://oldpasca.undiksha.ac.id/e-journal/index.php/jpbi/article/view/804.

Aldersen, J. C., Clapham, C., & Wall, D. (2003). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.

Alkubaidi, M. A. (2009). *Authenticity in Language Test Design*, 1–12.Available at https://www.academia.edu/19239014/Authenticity_in_Language_Test_Design.

Apriliani, A., (2013). *Item Analysis of English Tryout Test in Detik-Detik Ujian Nasional Bahasa Inggris 3 SMP/MTs Published By Intan Pariwara*. Faculty of Teacher Training and Education. Kudus: Muria Kudus University. Available at http://eprints.umk.ac.id/2300/1/Hal._Judul.pdf.

Bachman, Lyle F. (1995). *Fundamental Considerations in Language Testing*. Oxford : Oxford Press.

Bachman, L. F., Angeles, L., & Kunnan, A. (2015). *Statistical Analyses for Language Assessment* Workbook and CD-ROM, (January).

Bachman, Lyle F. & Palmer, Adrian S. (1996). *Language Testing in Practice*. New York: Oxford University Press.

Brown, D. H. (2004). *Language Assessment: Principles and Classroom Practices. (4th ed.)*. New York: Longman.

_____. (2005). *Principles of Language Learning and Teaching*. Pearson Education.

Chako, T. 2017. *Triangulation in Qualitative Research! Why the Use of Triangulation necessary in Social Research?*. Available at https://www.academia.edu/36277240/Triangulation_in_qualitative_research_Why_the_use_of_triangulation_necessary_in_social_research.

Cindrakasih, D. B., (2017). *Analysis of English Try Out Test of the Ninth Grade Students at MTSN Probolinggo.* Available at http://jiesjournal.com/index.php/jies/article/download/122/98.

Creswell, John W. (2012). *Educational Research Fourth Edition*. USA: Pearson Education Inc.

Creswell. J. W, & Miller. D. L. (2000). *Determining Validity in Qualitative Inquiry.* Available at https://www.tandfonline.com/doi/abs/10.1207/s15430421tip3903_2.

Crossley, Louwerse, McCarthy & McNamara (2007). *Language Teaching*, *41*(3), 409-429.

Denzin, NK. (1978). *Sociological Methods*. New York: McGraw-Hill.

Doye, P. (1991). *Authenticity in Foreign Language Testing*. Available at https://eric.ed.gov/?id=ED350819.

Geribka, M., (2016). *A Study on the Content Validity and Authenticity of English National Examination 2015 for Junior High School*. Faculty of Teachers Training and Education. Yogyakarta: Sanata Dharma University. Available at https://repository.usd.ac.id/6322/.

Gilmore, A. (2007). *Authentic Materials and Authenticity in Foreign Language Testing*, 97–118. Available at https://www.researchgate.net/publication/231910134_Authentic_materials_and_authenticity_in_foreign_language_learning.

Gronlund, N. E., & Waugh, C. K. (2009). *Assessment of Student Achievement. Ninth Edition*. Upper Saddle River, N.J.: Pearson Education Inc.

Hales, D. (2010). *An Introduction to Triangulation*. Switzerland: Unaids.

Hamilton, D. (1990). *Learning about Education*. Open University Press.

Hapsari, H. S., (2017). *The Authenticity of Assessment Used in Reading Class At The Third Semester Of English Education Program In Universitas Pekalongan Academic Year 2016/2017*. Faculty of Magister. Surakarta: Universitas Muhammadiyah Surakarta. Available at http://eprints.ums.ac.id/53544/.

Harrys & Valette. (2003). *Principles Language Testing I*. New York: McGraw Hill.

Hayes. R. L. (1987). *Research in Humanistic Education.* Available at https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.2164-4683.1987.tb00299.x.

Honorene, J. (2017). *Understanding the Role of Triangulation in Research*. Available at

http://www.srjis.com/pages/pdfFiles/149544238718.%20HONORENO%20J OHNSON.pdf

Hughes, A. (2003). *Testing for Language Teachers*. Cambrigde: Cambridge University Press.

Janet, A. (2018). *Exploring Content Validity and Authenticity of the 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: Universitas Negeri Semarang.

Joy, J. J. L. (2011). *The Duality of Authenticity in ELT*, *7*(2), 7–23. Available at https://www.semanticscholar.org/paper/The-Duality-of-Authenticity-in-ELT-Joy/a08a8316a5d119ecf2589d635c1f8d491444a279.

Lambert. A. V., & Clinton. E. L. (2012). *Qualitative Descriptive Research : An Acceptable Design*. (n.d.), (4), 255–256.

Lewkowicz, J. A., Centre, E., & Kong, H. (2000). *Authenticity in language testing : some outstanding questions*, *17*(1988), 43–64. Available at https://pdfs.semanticscholar.org/477c/57ac781e87c108a27a9b9158211118609e95.pdf.

MacDonald, Malcolm, Badger, Richard and Dasli, Maria. (2006). *Authenticity, culture and language learning. Language and Intercultural Communication*, 6 (3-4). pp. 250-261. Available at http://wrap.warwick.ac.uk/78443.

Miles, Matthew B. and Huberman, A. Michael. (1984). *Qualitative Data Analysis: A Source book of New Methods*. Beverly Hills. CA.: Sage Publications.

Nassaji, H. (2015). *Qualitative and descriptive research : Data type versus data analysis*. https://doi.org/10.1177/1362168815572747.

Natalia, S. A. M. K. D., Marhaeni, A. A. I. N., & Dantes, N. (2013). *Analysis of Authenticity of Teachers' Made Assessment and Its Contribution to Students' English Achievement in Elementary Schools in Buleleng Regency.* Language Education Department. Singaraja: Universitas Pendidikan Ganesha.Available at http://oldpasca.undiksha.ac.id/ejournal/index.php/jpbi/article/view/727.

Nawawi, H. Hadari. (1993). *Metode Penelitian Bidang Sosial*. Yogyakarta: Gajah Mada University Press.

Newmann, F. M., & Wehlage, G. G. (1993). *Five Standards of Authentic Instruction. Educational Leadership*. 50(7), 8-12.

Nunan, D. (1989). *Designing Tasks for the Communicative Classroom*. United Kingdom: Cambridge University Press.

Nuryanto, M., (2018). *The Validity Analysis of Reading Test Items on National Standard School Final Examination for 12th Grade of Man 1 Semarang*. Faculty of Magister. Surakarta: Universitas Muhammadiyah Surakarta. Available at http://eprints.ums.ac.id/68416/2/FRONT%20PAGES.pdf.

Patton, MQ. (1999). *Enhancing the Quality and Credibility of Qualitative Analysis*. HSR: Health Services Research. 34 (5) Part II. pp. 1189-1208.

Peraturan Badan Standar Nasional Pendidikan No. 11 tentang Kisi-kisi Ujian Nasional Sekolah Menengah Pertama/Madrasah Tsanawiyah Tahun Pelajaran 2018/2019.

Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia No. 04 Tahun 2018 tentang Penilaian Hasil Belajar oleh Satuan Pendidikan dan Penilaian Hasil Belajar oleh Pemerintah.

Purwandani, A. D., Raja, P., & Suparman, U. (2007). *The Analysis of the Authenticity of Authentic Reading Materials in Students' Text Book*, (2002), 1–9. Faculty of Teacher Training and Education. Lmapung: Universitas Lampung. Available at https://www.neliti.com/id/publications/192584/the-analysis-of-the-authenticity-of-authentic-reading-materials-in-students-text.

Qichun, Z. (2005). *The Analysis of Authenticity in Language Test*. Available at http://www.celea.org.cn/teic/59/59-122.pdf.

Richards, J. C. (2001). *Curriculum development in language teaching*. New York: Cambridge University Press.

Savin-Baden, M. & Major, C.H. (2004). *Foundations of Problem-based learning*. Buckingham: SRHE and Open University Press.

Shomoossi, N. (2015). *Authenticity and Authentication in Language Testing : an Operational Perspective*, (January 2009).

Shomoossi, N. (2007). *A Critical Look at the Concept of Authenticity*, *4*(1), 149–155. Available at https://www.researchgate.net/publication/268813193_A_Critical_Look_at_the_Concept_of_Authenticity.

Spence-Brown, R. (2001). *The Eye of the Beholder: Authenticity in An Embedded Assessment Task*. Language Testing 18: 463-81.

Wang. (2008). *Authenticity in Language Testing*.

Widdowson, H. G. (2004). *Teaching Language as Communication*. Oxford: Oxford University Press.

Widyaningrum, F. A. D., (2012). *Content Validity and Authenticity of the 2012 English Test in the Senior High School National Examination*, *16*(1), 23–29. English Language Education Study Program. Yogyakarta: Sanata Dharma University. Available at http://ejournal.usd.ac.id/index.php/LLT/article/view/279.

Will, L. (2018). *Authenticity in English Language Teaching*. New York: Waxmann.