

Chapter 1

Introduction

Background of the Study

A test is “a method of measuring a person’s ability, knowledge or performance in a given domain” (Brown, 2004). A good test serves a purpose, either formative, to help learners improve their performance or knowledge, or summative, to certify that learners indeed have learned what they should have learned (Office of Medical Education Research and Department (OMERAD), 2011). In terms of language testing, it serves as a means of measuring students’ proficiency in particular skills of the language. Most language tests measure one’s ability to perform a language, such as speaking, writing, reading, or listening to a passage. Thus, a test can be of any scale to gauge some qualities of students after participating in learning for some period. It can be a precious tool for obtaining information that is relevant to several concerns in language teaching, such as providing evidence of the results of learning and instruction which in turn serve as feedback on the effectiveness of the teaching program itself, providing

information that is relevant to making decisions about individuals, i.e. determining what specific kinds of learning materials and activities should be given to students, etc.

(Bachman & Palmer, 1996, p. 8)

In language teaching, there are several types of tests based on some criteria. For instance, formal and informal test, formative vs. summative test, discrete-point vs. integrative test, and classroom-based test vs. standardized test. This study focuses centrally on national standardized test. Brown defines a good standardized test as the typical norm-reference test which aims to place test-takers “on the continuum across a range of scores” and to classify test-takers by their rank (2004). It results from a set of thorough processes of empirical research and development and dictates standard procedures for administration and scoring.

Standardized test is employed to measure the students’ mastery on basic parts of the curriculum in general and the result functions as a portrait of our education quality. An example of a large-scaled standardized test administered in Indonesia is the National Exam (abbreviated into NE) held annually throughout the country to measure students’ achievement at the end of a learning period in each level. This

year, the national exam is administered for secondary level only, leaving elementary school out. The NE is constructed based on the national curriculum, which is a set of subjects and standards used by primary and secondary schools so children learn the same things. It covers what subjects are taught and what standards children should reach in each subject.

The administration of the NE is maintained by the government because it supports the principles of curriculum development and administration. Based on Glatthorn et al's framework (Glatthorn, Boschee, & Whitehead, 2006), NE has the main purpose, that is, to ensure that the students understand and are able to implement the basic skills of curriculum. This part is included in the process of structured basic/written curriculum. It needs to be supported, taught and later tested.

The National Examination (NE) is the latest form of a school leaving examination in Indonesia starting from 2005 until now. It can be defined as a test to measure and evaluate the students' competence nationally by the central government after the process of teaching and learning (The Regulation of the Minister of Education 2005, p.1). The NE is implemented as a way of improving national education quality, which is supported by the Government Regulation Number 19 year

2005 on the National Standards of Education (which later has been altered by Government Regulation of the Republic of Indonesia No. 32 in 2013 about Education National Standards).

As stated in Education National Standard Organization Regulation No. 0022/P/BSNP/XI/2013, the table of specifications used for English National Exam (ENE) 2013/2014 academic year refers to that listed in Education National Standard Organization Regulation No. 0019/P/BSNP/XI/2012 which shows that the examination only covers few listening materials, several reading materials and a small portion of writing materials in a multiple-choice test format. As the aims of teaching and learning process in the English curriculum are to develop not only listening, reading, and writing skills but also speaking skill, ENE that omits the assessment of those skills is claimed to lack authenticity. This is criticized by many that the test does not reflect the students' real language proficiency since it neglects the productive skills.

In the ENE, the skills tested consist of listening, reading comprehension, and writing in a multiple-choice format. These tests may train learners to practice all levels of cognitive skills and are the base of the whole learning process. Evaluating reading and listening comprehension texts and tests is of a

great importance because this task helps learners gain and develop their knowledge. Failing to read and listen as well as comprehend what is read and listened in early learning stages is potential to create educational problems in the learners' life. On the other hand, writing skill is not less important than the two skills previously mentioned. Harmer (2004) points out that writing gives learners opportunities for language processing and helps learners practice and work with language they have been learning. Writing is hard because it is a struggle of thought, feeling, and imagination to find expression clear enough for the task at hand. Therefore, writing has to be learned in school very much the same way that it is practiced out of school.

Like many tests of this nature, the NE has been subject to a variety of criticism. The administration of NE causes some impacts on its participants. The impacts are called 'washback' effects. Washback is seen as the influence of testing on teaching and learning (Brown, 2004). The major points taken into consideration are that the NE as high-stake test has negative 'washback' effects. In most cases, high stakes test leads to the phenomena of teaching the test and teaching to test. The former includes having some idea of what test questions

will be and teaching the answers to the students. Consequently, this method does not promote learning and critical thinking, and it is absolutely dishonest and could be considered as cheating. The latter involves understanding which concept will be assessed and teaching those concepts to students.

Furthermore, NE is argued to “cause injustice and anxiety in learning among the students” (Sukyadi & Mardiani, 2011). The administration of NE places children under constant stress for their whole academic lives, that teachers and principals are forced to give extra attempts to ensure their students can pass the tests, and that the result of national curriculum testing is for school league tables and branding for recruiting new students. Moreover, it leads to a shift in teachers’ concern that their professional abilities are under-used and raises the phenomena such as teaching to the test, narrowing the curriculum and focusing unequal proportion of resources on students. The form and the content of the test affect the classroom teaching where teachers tend to spend the classroom time on the skills like listening and reading that are on the test and very little on practicing speaking because it is not tested (Nation & Newton, 2009, p. 169). Furthermore, at

the extreme level, national exam betrays the primary purpose of education (Yero, 2001-2002).

It is worth noting that evaluating test items is necessary since it gives a clear portrait of the coherence between what is written in the curriculum and what is tested. Brown and Rodgers (2002, p. 289) define evaluation as “the process of seeking to establish the value of something for some purpose”. It can also bring some recommendations on what should be taught and tested in the next period since what was taught several years ago may not be the same as what the students need now. There is a raising awareness that the key factor to improve the condition of a nation is to improve the quality of human resources within the country through education. It is a key in building and improving individual attitudes and skills to compete in the global market. This needs to be pursued seriously and focused on by the government because the national performance is determined by how the education is undertaken by the community. In consequences, the nations throughout the world are exerting their best to improve their educational programs.

To achieve this, evaluative processes on different fields of curriculum ranging from learning, teaching and assessing

should be carried out to find out the strengths and weaknesses as well. Specifically related to the national exam, Genesee (2001, p. 144) stated that “evaluation in TESOL settings is a process of collecting, analyzing, and interpreting information about teaching and learning”. This process enables us to make informed decisions through which student achievement will increase and educational programs will be more successful.

Researchers have shown that comprehension skills and success in learning L1 and L2 as well as other subjects are closely related. Thus, the comprehension skills should be taught to train students’ cognitive skills ranging from literal comprehension to appreciation comprehension. When these skills are practiced, students can develop not only their lower order thinking skills (LOTS) but also their higher order thinking skills (HOTS) and can effectively respond to testing items which assess the latter skills. LOTS is the foundation of skills required to move into higher order thinking. These are basic skills that are taught very well in school systems and include activities in reading and writing (Wilson, 2000). In lower order thinking, information does not need to be applied to any real life examples, it only needs to be recalled and slightly understood.

Due to this fact, it can be argued that HOTS are fundamental skills that can empower individuals' ability to analyze, to synthesize (to combine knowledge of different sources), to discuss, to judge, and to evaluate. McDavitt (1993, p. 20) says that "Higher Order Thinking Skills include analysis, synthesis, and evaluation and require mastery of previous levels, such as applying routine rules to familiar or novel problems". These skills will help the nation to develop its citizens' capability of critical thinking, problem solving, planning, analyzing, and decision makings.

In consequence, raising the awareness among teachers and educators as well as the society that curriculum and educational processes are responsible for building learner's critical thinking is deemed very crucial. If the ENE is designed to test students' HOTS, most teachers' and students' activities in the classroom will be oriented toward improving these skills due to the teacher's tendency to neglect items which are not tested. In turn, this practice will be beneficial for students for their whole academic lives. Otherwise, if the tests are dominated with questions' asking the students' LOTS, they will be low achievers who are merely capable of focusing on lower order thinking skills (LOTS). This is in line with Jacob

in Sukyadi & Mardiani (2011) who states that high school national graduation exams increased the rate of drop outs and hinder the development of higher order thinking skills.

Good test items are those items that can assess the performance of learners effectively. Since language testing has such a powerful influence on classroom instruction, it is important for educators to be informed about the question types in examination, especially a high-stake exam such as the National Exam. With this knowledge, educators can evaluate the level of comprehension and the students' competence to process high order thinking skills. Students' interactions with questions directly influence their future learning outcomes (Armbruster & Ostertag, 1993). The implication is that higher order questions would promote higher order processing of the text.

In brief, analyzing the ENE items is a process that sheds some light on the strengths and weaknesses of listening and reading comprehension texts and tests as well as writing performance test and their classifications of LOTS and HOTS.

The Research Questions

The secondary stage and particularly 12th grade is considered a critical stage in the academic life of the students who are about to attend university. Professors at Indonesian universities usually complain of secondary school graduates attending the universities to pursue their higher education lacking of their higher thinking skills. The researcher has noticed that university professors feel that Indonesian college students do not demonstrate effective ability to answer questions of the higher thinking skills types. Students commonly find difficulties in answering reading and listening comprehension questions at the levels of higher order thinking skills.

This issue has triggered research interest to investigate the nature of questions used in the ENE for Senior High School students and its efficacy to develop the 12th grade students linguistically, mentally and intellectually. This study thus is aimed to answer these questions:

To what extent are the lower order thinking skills (LOTS) and higher order thinking skills (HOTS) available in the questions in the ENE of 2013/2014 academic year in the light of suggested criteria?

- a. To what extent do the questions in the ENE 2013/2014 academic year include literal comprehension?
- b. To what extent do the questions in the ENE of 2013/2014 academic year include reorganization comprehension?
- c. To what extent do the questions in the ENE of 2013/2014 academic year include inferential comprehension?
- d. To what extent do the questions in the ENE of 2013/2014 academic year include evaluation comprehension?
- e. To what extent do the questions in the ENE of 2013/2014 academic year include appreciation comprehension?
- f. Are the proportions of items assessing students' LOTs and HOTS consistent among the twenty test packages in the ENE of 2013/2014 academic year?

The Purposes of the Study

According to the research questions that are defined above, the aim of this research is to formulate a checklist of criteria for evaluating LOTS and HOTS in the ENE of 2013/2014 academic year, to identify the proportions of LOTS and HOTS in the listening and reading comprehension questions as well as writing performance item in the ENE of 2013/2014 academic year, and to compare the consistency of the number of items assessing students' LOTS and HOTS among the twenty test packages in the ENE of 2013/2014 academic year. As regulated in Education National Standard Organization Regulation No. 0020/P/BSNP/I/2013, the 20 packets of the test items are professionally designed by the test designers to reflect the same table of specifications listed in Education National Standard Organization Regulation No. 0019/P/BSNP/XI/2012 that share the same level of difficulty, quality, and validity.

The purpose of this research is not to criticize the test designers but to provide feedback about the quality of the test items under HOTS and LOTS perspectives. Additionally, it is aimed to provide current information in the

field of language testing that will lead to the improvement of ENE for Senior High School in Indonesia, and thus to the improvement of English teaching and learning in Indonesia.

Scope and Limitation

This study was restricted to evaluating LOTS and HOTS in 50 listening and reading comprehension questions as well as writing performance tests in the twenty test packages of ENE for Senior High School of 2013/2014 academic year. The roles, the importance, and the issue of authenticity of ENE were not discussed in detail as they are beyond the scope of this research. Due to the constraint of time and finance, it was not possible to investigate the issue of test validity, reliability, the level of difficulty, and the item discriminability, but only to concentrate on specific relevant questions as stated previously.

The whole documents analyzed of the proposed study were all the 20 (twenty) test packages of ENE of 2013/2014 academic year. In the school year 2013/2014, ENE test items consisted of 15 listening questions, 31 reading questions and 4 writing questions. The time allotted to do the exam was 120 minutes. The study took place in the 2013/2014 academic year and was restricted to the used tool in the form of a checklist

based on Barrett's taxonomy as the criterion for the evaluation process.

Theoretical Framework

Relevant literature is reviewed and discussed to provide a conceptual and theoretical basis upon which the field of language testing can be understood. This study is primarily anchored on the Barrett's Taxonomy of Comprehension (1980), which discusses the different levels of Comprehension namely: literal, reorganization, inferential, evaluation and appreciation. The theory assumes that learners move from the literal understanding to another, until the learner fully understands and appreciates the cognitive and aesthetic aspects of the material.

One interesting aspect of the Barrett taxonomy, according to Armbruster & Ostertag (1993), is the subdivision of categories according to specific type of information targeted by the question (e.g. recognizing and recalling main ideas, inferring cause and effect relationships, identification with characters and incidents). It contributes to the usefulness of Barrett's taxonomy as a guide for constructing questions on a

variety of levels as well as for judging questions that have already been created. It can be used to evaluate students' comprehension of text. Therefore, the researcher decided to use Barrett's taxonomy as the theoretical framework for this study.

Definition of Key Terms

In providing a better understanding about the key concepts of the study, the following terms are operationally defined.

National Exam:

a standardized exam for English Subject for Senior High School students held in Indonesia of 2013/2014 academic year in which all test takers answer the same questions under the same conditions, in multiple-choice format.

Evaluation:

analysis of language testing items to find out the availability and classification of LOTS and HOTS based on Barrett's taxonomy.

Language testing:

an instrument that provides an accurate measure of test-takers' ability within language domain.

Lower order thinking skills (LOTS):

the ability to do literal and reorganization comprehension based on cognitive skills of Barrett's taxonomy.

Higher order thinking skills (HOTS):

the ability to do inferential, evaluation, and appreciation comprehension based on cognitive skills of Barrett's taxonomy.

Literal comprehension:

a level of comprehension which requires the student to identify or mention ideas and information which are explicitly stated in the selection.

Reorganization comprehension:

a level of comprehension which requires the student to analyze, synthesize, and/ or organize ideas or information explicitly stated in the selection.

Inferential comprehension:

a level of comprehension demonstrated by the student when he or she uses the ideas and information explicitly stated in the selection, his or her intuition, and his or her personal experience as a basis for conjectures and hypotheses.

Evaluation comprehension:

a level of comprehension which deals with judgment by comparing ideas presented in the selection with external criteria provided by the teacher, other authorities, or other written sources, or with internal criteria provided by the reader's experiences, knowledge, or values and focuses on qualities of accuracy, acceptability, desirability, worth, or probability of occurrence.

Appreciation comprehension:

a level of comprehension which involves all the previously cited cognitive dimensions of reading or listening, for it deals with the psychological and aesthetic impact of the selection on the reader/listener.

The Significance of the Study

The study aims to achieve the following:

1. To inspire language testing designers to modify, organize, and enrich English language testing items which develop students' LOTS and HOTS accordingly.
2. To be used as a reference for future researchers. It will be a useful material for future studies in a sense that it will

impart findings about the LOTS and HOTS of the National Exam.