



VALIDITY AND WASHBACK OF ENGLISH TESTS IN THE NATIONAL EXAMINATION

Kurniawan Aprianto ✉

Postgraduate Program of Semarang State University, Indonesia

Info Artikel

Sejarah Artikel:
Diterima April 2013
Disetujui Mei 2013
Dipublikasikan Juni 2013

Keywords:
Validity;
Washback;
English Tests;
National Examination

Abstrak

Penelitian ini meneliti validitas butir soal Bahasa Inggris dalam Ujian Nasional Sekolah Menengah Atas dan dampaknya terhadap kegiatan dan proses pembelajaran yang dilakukan guru di kelas. Dalam penelitian ini digunakan gabungan metode dokumentasi dan survei. Data yang diteliti terdiri atas butir soal Bahasa Inggris Ujian Nasional tahun 2010 dan 2011, dan tanggapan guru terhadap pertanyaan-pertanyaan dalam angket tentang butir soal Ujian Nasional dan pelaksanaannya. Data dianalisis dengan menggunakan statistik deskriptif dan analisis isi. Hasil penelitian menunjukkan bahwa ditemukan ketidaksesuaian antara butir soal dengan isi kurikulum Bahasa Inggris. Butir soal itu kurang autentik. Guru menggunakan metode drill, latihan, dan kiat-kiat khusus untuk membantu siswa menyelesaikan soal-soal latihan ujian. Selain itu, guru menyelenggarakan pembelajaran reguler pada dua tahun pertama, sedangkan tahun terakhir difokuskan untuk berlatih mengerjakan soal-soal yang diprediksi akan keluar dalam ujian nasional

Abstract

The study investigates the validity of the English tests in the National Examination of Senior High Schools and its washback to the teaching and learning process in the classroom. This is a combination of a documentary and a survey study. The data consisted of test items of English tests in the National Examination in 2010 and 2011 and teachers' responses to the questionnaires about English tests. The data were analyzed using a simple descriptive statistics and a content analysis. The results show that there were significant incompatibilities between the English test items in the National Examination and the school-based curriculum. The test items were less authentic. Teachers used drills, exercises and tricks in assisting the students to answer questions. In addition, the teachers conducted regular teaching and learning activities during the first two years but focus on practicing on test items that might be found in the National Examination.

© 2013 Universitas Negeri Semarang

✉ Alamat korespondensi:
Kampus Unnes Benda Ngisor, Semarang 50233
E-mail: jurnalpps@unnes.ac.id

ISSN 2087-0108

INTRODUCTION

In Indonesia, English is taught as a foreign language (EFL) in a classroom environment. In keeping the quality of the teaching-learning process, a curriculum is designed and developed from time to time. The newest curriculum is School-based Curriculum (KTSP) which has a close relation to the Competence Based Curriculum (CBC). To measure the students' achievement after the 3-year-study both in junior and senior high school, a final test, called National Examination (NE), is conducted as part of the requirements to graduate from each school level.

At the beginning, the score of NE was used as the basis to decide whether or not a student passed the exam. However, after some years, the score as the passing grade was revised into a combination between NE score and School Examination score. Some scholars claimed that NE mostly gives a relatively bad effect on teaching and learning process. Test-driven on teaching-learning and some cases of cheating were of some negative effects, whether they were done by students or even by teachers. Cheating is a big issue every year and it is one big challenge for the government (Tribun News, 5 April 2012).

NE has also influenced the choice of teaching materials. Teachers tend to use textbooks which match the kinds of questions of NE. It is usually far from activities which lead to communicative goals of EFL. Razmjoo (2007) mentions that textbooks provide a central role in conducting language classroom activities in any educational institution all over the world. He also states that in certain situations, textbooks are the basis of most language input and language practices the students have in their classroom activities. As we know, students mostly have time to interact with the textbook they have, and it means that what is inside the textbook will have a great deal of effect on the students. For teachers, these also provide some aspects of language learning such as contents of lessons, all skills taught and tasks for students. Teachers then can actively change or modify those for the purpose of accomplishing their teaching objectives.

One more important function of a whole process of learning is on how teachers are conducting the assessment. Fulcher (2010:1) states that "tests are mostly used to place learners into classes, to discover how much they have achieved, or to diagnose difficulties that individual learners may have". Moreover, tests could motivate students to study. Commonly, students do not want bad scores.

Another thing is that the differences between the presentation of language testing and the textbook give more or less the method used by the teachers in teaching EFL. Students in the third year of senior high school are sometimes treated differently from those who are in the first and second year. They tend to be much more focused on kinds of testing. For the same reason, the teacher gives their students such test drills. They use books such as workbooks which are designed to be as close as possible to the final exam (NE). The students prepare for the test by becoming familiar with the kinds of test, not by increasing their language competence as a whole (Fulcher, 2010). However, it seems that both students and teachers avoid having clear communicative objectives.

Syllabus is one aspect of curriculum but not exactly the same. Syllabus is details of the contents of a course of instruction and list what will be taught and tested (Richards, 2001). By that definition then syllabus of reading will specify kinds of reading skills that will be taught and practiced during the course, the functions, topics, or other aspects in reading that will be taught, and the order in which they will appear in the course. A curriculum specifies the needs of a group of learners, aims and objectives to manage the needs and a certain program to conduct in order to achieve the aims and objectives. A curriculum also contains appropriate syllabus, course structure, teaching method, and material as well as the evaluation of a language program which should be done as a measurement of the result along the implementation of the curriculum.

Richards (2001) then summarizes the notion curriculum as consisting of aims and objectives, the content, organization, and evaluation. The curriculum development is of planning and implementation processes and of the developing and renewing curriculum due to the emergence of newer need analysis, situational analysis, planning learning outcomes, course organization, selecting and preparing new teaching materials, providing for effective teaching and evaluation. Curriculum is designed by the government to maintain the similarity of learning goal of the country. This is very important because if there are significant differences in different areas then it is difficult to measure students' learning achievement together and the standard is also difficult to gain.

Testing, including all language testings, is one form of measurement. Brown (2004:3) mentions that "a test is a method of measuring a person's ability, knowledge, or performance in a given domain", so that there should be limitation

on what students will be tested. There is always a potential for error when we measure something. Bachman and Palmer (1996) believe that there is no such thing as a good test or bad test, or even such thing as the best test even for a certain situation. Testing still has a very important role as a part of learning process. As one type of assessments, a test becomes a favorite model for teachers to know the level of understanding the students have achieved.

To make sure that a test will really measure what has been taught, the validity of the test is very important. Fulcher (2010) gives more explanation about validity which has at least five aspects, i.e.: a) the substantive aspect, if the test can justify what can be shown that the test we draw about the knowledge, skills, and abilities of the test taker; b) the structural aspect, whether the test provides information on a number of different skills, it should be structured according to the skills of interest; c) the content of the test should be representative to the content of a course of study (or of a particular domain) in which we are interested; d) generalizability, whether it is predictive of ability in contexts beyond these modeled in the test; e) the external aspect, the relationship of the scores of the test to other measures of the same or different skills and abilities. Furthermore, Brown (2004) says that testing is an administrative procedure to measure students' muster at identified time stated in the syllabus. Developing a reliable and valid test is then a need for teachers or the government to apply what they have decided and stated on the curriculum and syllabus.

The problem for most test developers is that the dilemma between validity and reliability of a test, for its quality. This is a problem in Indonesia since English is not a second language in our society or it is very hard to find the exposure of English in our real life. However, the test usefulness can be described in terms of the six test qualities: Reliability, construct validity, authenticity, interactiveness, impact, and practicality.

The National Exam is a standardized test intended to check the students' achievement on their competences which is conducted nationally by the government. During the first couple of years in conducting the national examination, the test was determined fully for students' graduation. The passing grade was decided by the government and every student had to reach the score in order to graduate. But starting 2011, the combination between NE and School exam (developed and conducted by the school itself) was designed by MONE to determine students' grade

for graduation.

The material of NE is based on the indicators designed by BSNP (Board of National Education Standards). This indicators reflect the current curriculum and based on competence standard and basic competence mandated in the Government Regulation (PP) No. 22 year 2006 (*Peraturan BSNP No 13/P/BSNP/XII/2011*). The national level committee makes the questions for National Examination based on Basic Competence and content standard as stated in Ministry Rule No. 22 year 2006. (MONE Regulation no. 59 year 2011 article 23(2)). It refers to the extent to which a given test score as an indicator on how well or the ability(ies) we want to measure. In other words, a test has construct validity if it accurately measures a theoretical, non-observable construct or trait. The construct validity of a test is worked out over a period of time on the basis of an accumulation of evidence (Bachman and Palmer, 1996; Sackett, 2012).

The notion of "washback" is prevalent in language teaching and testing literature, but it is seldom found in dictionaries. Some writers used the term "washback" while others preferred "backwash" to describe the effects or influences brought by tests or examinations. The impact of a test is related to individual (at micro level) who is taking the test and to the educational system or society (at macro level) (Bachman and Palmer, 1996). On the other hand, washback or backwash is "the effect of testing on teaching and learning, it can be harmful or beneficial" (Hughes, 1989). There has been a perception that washback influences teaching content but it does not affect the teaching methods.

This paper addresses the validity of the English tests in the National Examination of Senior High Schools and its washback to the teaching and learning process in the classroom.

METHODS

This is a descriptive study and a simple descriptive statistic and content analysis was adopted. This study utilized research models from previous studies. For example, the main basis for the present study was Bharati and Suwandi (2006) study, which also focused on NE and its relevance to Competency Based Curriculum. However, the present study differed from Bharati et al. (2006) in some aspects: the sample of the present study was more specific, i.e. NE text from 2010-2011, and the curriculum was also more specific and more detail (School-based curriculum is such a competency based curriculum

concerning the school condition); the Regulation from Ministry of National Education was also considered. Moreover, more data about teachers' perspectives on viewing English test in National Exam was also taken into consideration.

There are two kinds of objects in this study. First, written object which consists of questions of National Exams, competence standard and basic competence in the syllabus of senior high school. The questions of National Exams in year 2010 and year 2011 are then taken as the first object. The second object is information from teachers of senior high schools in Mataram about the process of the teaching-learning of English in the classroom. That information was gained through questionnaires. Teachers from various senior high schools were the participants of the study. In conducting this survey all respondents were asked the questions that were appropriate to them, and so that, when those questions are asked, they are always asked in exactly the same way (Brace, 2004). There are 23 senior high schools which consist of 10 public schools and 13 private schools in Mataram. The combination between probability and non-probability sampling techniques were used. The randomized quota sampling was conducted

to determine the participants of the study (Cohen et al., 2000). Thus, each school was represented by one English teacher as the object of the study.

The written data, the last two consecutive NE questions (2010-2011), were collected including the listening materials. While about school-based curriculum, competence standard and basic competence of English subject enclosed in Ministry Rule no 26 year 2006 was considered as the data.

A survey using a questionnaire as the instrument was employed to collect the second data. The instrument was partly developed using three-point Likert-scale questionnaire covering their views and opinions dealing with the role of NE to improve students' competences in English as mentioned in school-based curriculum, a choice depending on NE or teacher-made evaluation, academic advantages, material development regarding to NE, components of English skills to be examined, the function of scores from NE, and their readiness to conduct the evaluation. Each question was completed with the open ended question for further explanation of the choice considered. The table below shows some sample questions in the questionnaire.

Table 1. Sample questions in the questionnaire

Category	Sample questions / statements
The role of NE to improve students' competences	I believe that English in National Examinations (UN) improves students' competences in English
Government-developed or teacher-made evaluation	I believe that all questions in UN should be developed by the government and not by the teacher(s) in each school.
Academic advantages for teachers	I have got a positive thing from the presence of National Exam to my teaching-learning process.
Material development regarding to NE	Do you spare your time to discuss the syllabus and the material with your colleagues?

The written objects (questions of the latest English test in National Exam and the latest English curriculum) were analyzed using qualitative descriptive analysis as content analysis did. They are categorized based on the similarities and differences and generate some inferences based on those similarities and differences. Three analyses were done: 1) comparing between the Curriculum (School-based curriculum) and the operational indicators of English test in National Exam from two consecutive years (2010 and 2011); and 2) comparing between the questions of English test in National Exam from two consecutive years (2010 and 2011) and the operational indicators of the test from the same years; 3) analyzing the compatibility of the questions of the English test

in NE (2010 and 2011) and the School-based curriculum. This analysis was conducted to find out the validity of the test concerning what had been mandated in the curriculum. This is very crucial because English test in NE should evaluate what students had learned.

The questions of the questionnaires employing Likert-scale and closed-ended questions were summarized using descriptive statistics. Open-ended questions were analyzed according to the tradition of content analysis and were categorized based on the similarities. In other words, these were analyzed qualitatively to reveal the patterns of relation among verbal responses made by the respondents (Sulistyo, 2009). Respondents answers in the questionnaire were coded as 1 (for

“Yes” answer), 2 (for “In between” answer), and 3 (for “No” answer). The data then were tabulated to find out the frequency of the answer. To get deeper and detailed answer, the respondents were to propose the reason to elaborate their answers. After the coding was complete, the data were further analyzed. Firstly, simple descriptive statistics (% of each answer) to find out the trend was applied and then content analysis to establish the inference by categorization was conducted. Secondly, the data gathered from these analyses were then described and used to answer the present research questions.

RESULTS AND DISCUSSION

As most experts consider that language ability consists of four language skills: listening, reading, speaking, and writing, the language proficiency test should contain all language skills. One biggest problem is how to conduct productive skills assessment. These two skills, speaking and writing, required different instruments compared to the other two. They could not be done by machines because assessing those skills involved human cognitive activity which could not be replaced by machinery. Doing such assessment to measure a very large number of test takers would face some problem especially when conducted at relatively the same time. The following discussion focused on two receptive language skills, i.e. listening and reading, found in English Test in National Examination.

Based on the data, the representativeness of the test is discussed in a number of aspects:

a. Test Coverage

In listening section, all questions were covered all indicators of Graduate competency Standard. But in Reading section (NE 2011), some points in the indicators were not covered in the test items.

b. Test Relevance

Communicative competences required a test which accommodate all skills integratedly while in the English test of NE measure skills in isolation one another. The English tests in National Examination were far from being relatively relevant. Or we could say that they were less relevant to achieve the goal of Competency Based Curriculum.

c. Program Coverage

If the English test in National Examination were supposed to show the mapping of educa-

tional progress in Indonesia, it seemed to succeed for some extent. The test would reveal how well students answer the questions. However, it was only part of students' proficiency since the tests only portrayed students' receptive skills.

Authenticity is a quality about to what extent a test task related to the target language use task. In other word, it provides an investigation to which score interpretations generalize based on the performance of the test to language use in the target language use domain. We realize that it is not easy to capture target language use task since students in Indonesia commonly do not have real life use of English in their daily lives. Otherwise, what we need to do is making language instructional TLU domain, that is, situations in which language is used for the purpose of teaching and learning the language. I adopt some characteristics of authenticity by Mueller (2012). Related to the characteristics proposed by Mueller, English Test in National Examination had relatively low in authenticity.

Interactiveness of the test shows the test taker's individual characteristics are involved in accomplishing a test task. This includes language ability (language knowledge and strategic competence), topical knowledge and affective schemata. In understanding whether a test task brings a relatively high Interactiveness or not, all the components must be regarded. But sometimes a test task does not need to have a high level of interactiveness, the minimum set of acceptable level would be enough. Based on the indicators of listening and reading section of the English test in NE, the degree of interactiveness of the test task was relatively high.

Most teachers (47.37%) believed that English test in National Exam has improved students' language ability, but a relatively high percentage of teachers (36.84 %) did not really believe that it supported students' improvement for some reasons. They said that ET in NE did not assess all skills so that it was far from the whole description of students' improvement. Some others, 15.79% of teachers, actually said the same thing as those who did not really believe that ET in NE improved of students' competences. It only assessed perceptive skills (reading and listening) which were not communicative as not all skills are included, whereas actually they are inseparable.

In seeing the curriculum, almost 100% (94.74%) of teachers had the same opinion that the current curriculum was already ideal for the time being, i.e. having communicative competences as the teaching-learning goal. Only one respondent gave 'in between' opinion and the rea-

son was that in her opinion the evaluation of the learning process should be a collaboration of two stakeholders of the national education, i.e. the government and the teachers. She actually agreed that the current curriculum provided students with more communicative goal. But the final test (NE) did not really assess students' achievement. There were still gaps between the school-based curriculum and the final test (NE).

Final assessment, as one component of the most influential and critical factors to decide whether a student has finally completed the study at senior high level, is importantly conducted by the right party. 12 teachers (63.16%) agreed that it was the government's job to conduct the assessment. In contrast, 4 teachers (2.05%) answered 'No' with some different perspectives. Some of them stated that only the teachers knew their students best, so that in developing the final assessment, the teachers should be in charge. Moreover, the current curriculum was actually built up by the teachers, and that was why called School-based Curriculum. The rest of the respondents (3 teachers or 15.79% of all) answered 'in between'.

Regarding to the regulation that final test held by the school was also a part of the whole score, teachers played in this role. They tried to make the score as high as possible to mark up the final score. It could be conducted by having a relatively easier questions or having remedial test for those who got relatively low scores.

Ten teachers (52.63%) felt that they had a positive gain from it, directly or indirectly. Some other teachers (5 in number or 26.32%) felt that they did not get anything from the NE so far. They stated that the pressure on the teachers that the students must succeed in National Exam made the teaching-learning process dull and tiring. No fun at all. So they came to the conclusion that it has a negative effect on the teaching-learning process as it should have been a communicative learning process (elaborating all skills) but it turned out to be non-communicative activities during teaching-learning process. While four teachers (21.05%) seemed to have indirect benefit from the presence of National Exam. They felt alright as long as the UN was conducted well. They indirectly learned about the variation in the test task to measure students' achievement.

Certain passing grade was to be an exit requirement. But this leads to some disadvantages for students and teachers. In teachers' opinion (the respondents), students had already got additional burden as they should surpass the passing grade in order to graduate. This made them focus more on the score, not on the actual compe-

tences they had to achieve. This tended to make students commit dishonesty. The students lacked motivation to learn the language and to practice it. On the teachers' side, the impact was more or less the same. As teachers were part of the educational system, they would feel ashamed if their students managed bad scores in NE. This could make them do everything in order to make their students succeeded in their final examination. According to this phenomenon, what the respondents (15 respondents out of 19) did was to prepare everything related to National Examination, practicing kinds of test which were like the model of National Examination. Giving question drills to the students during the last semester was only an option they did besides giving students more time in the afternoon to have more practices.

The findings about the teachers' point of view regarding the carrying out of the final assessment of English in National examination showed that only 26.32% of the respondents believed that English test in National Exam was still good for their students as it gave a significant effect on students' motivation. This test was a kind of a qualified standard test because the government did it by involving highly qualified teachers / researchers in developing the test. But most respondents (63.16% who answer 'No' and 10.53% who answer 'in between') said that there were still many things to consider about the test such as the area (skills) assessed, scoring system, its passing grade, its reliability, and some other external aspects (e.g. socio-cultural aspect).

All respondents mentioned that there was nothing they could do except preparing their students to be able to answer the test and motivating them to be honest when doing the test. They simply tried to deliver what they called communicative language learning before the last semester because their sixth semester was totally for preparing students to face the English test in NE. Furthermore, they also gave some points of view to reduce disadvantages they already faced. Here are their expectations to NE:

1. There should be an effort to bring back a more communicative language teaching because schools are not cram courses. It is to make students really have communicative competence. A test is only a part of learning process, so they hoped that the government would not use it to be an exit requirement from schools anymore.
2. The government should provide a more comprehensive test, not only assessing receptive skills.
3. Highlighting the function of the English

test. The test is to measure students' level of ability in acquiring English. So that cheating in doing the test can likely to be avoided because the students would not be afraid of the penalty, i.e. failing to finish the study.

4. Removing the passing grade as the exit requirement because it likely makes students not perform their actual performance. Standardized tests such TOEFL and IELTS can portray the level of competences which can be the alternative assessments for students to do.

Let the schools to be the institutions that can manage the graduation themselves though there is still a kind of monitoring systems from the Ministry of National Education. This hopefully leads to school independency and responsibility

CONCLUSION

Due to the relatively low validity, teachers actually wanted a better test system in assessing students' competences. They preferred to conduct more comprehensive test task which includes all language skills. This kind of test would encourage students to focus on language ability. The teachers also propose that there should not be a passing grade for students to reach in order to finish their study. The test system such as IELTS and Cambridge English Exam would give a description about the level of achievement, not judging whether he/she has passed the exam or not.

REFERENCES

- Bachman, L. F. and A. S. Palmer. 1996. *Language Testing in Practice: Designing and Developing Useful Language Test*. Oxford: Oxford University Press.
- Bharati, D. L. and Suwandi. 2006. UAN and its Relevance to the New Curriculum, KTSP. Proceeding of *TEFLIN International Conference* 2006.
- Brace, I. 2004. *Questionnaire Design: How to Plan, Structure and Write Survey Material for Effective Market Research*. London: Kogan Page Ltd.
- Brown, H. D. 2004. *Language Assessment: Principles and Classroom Practices*. New York: Pearson.
- BSNP. 2006. *Panduan Penyusunan Kurikulum Tingkat Satuan Pendidikan Jenjang Pendidikan Dasar Dan Menengah*.
- Cohen, L. and L. Manion and K. Morrison. 2000. *Research Method in Education*. London: Routledge Falmer
- Fulcher, G. 2010. *Practical Language Testing*. London: Hodder Education
- Government Regulation no. 45 year 2010
- Government Regulation. 22 year 2006
- Hughes, A. 1989. *Testing for Language Teachers*. Cambridge: Cambridge University Press
- Kattington, L. E. 2010. *Handbook of Curriculum Development*. New York: Nova Science Publishers, Inc.
- Krippendorff, K. 2004. *Content Analysis: An Introduction to its Methodology*. California: Sage Publications, Inc.
- Lightbrown, P. M. and N. Spada. 2006. *How Languages are Learned*. Oxford: Oxford University Press
- Ministry of National Education Regulation 22 year 2006
- Ministry of National Education Regulation 59 year 2011
- Razmjoo, S. A. 2007. High Schools or Private Institutes Textbooks? Which Fulfill Communicative Language Teaching Principles in the Iranian Context? *Asian EFL Journal*, Vol. 9, No. 4. 2007, pp. 126-140
- Richards, J. C. 2001. *Curriculum Development in Language Teaching*. Cambridge: Cambridge University Press.
- Spratt, M. 2005. Washback and the Classroom: the Implication for Teaching and Learning of Studies of Washback from Exams. *Language Teaching Research*. Vol 9, No. 1, (2005); pp. 5-29.
- Sulistyo, G. H. 2009. English As A Measurement Standard In The National Examination: Some Grassroots Voice. *TEFLIN Journal*, Volume 20, Number 1, February 2009; pp. 1-24.