

Content Validity and Authenticity of the 2012 English Test in the Senior High School National Examination

Frisca Ayu Desi Widyaningrum

Carla Sih Prabandari

English Language Education Study Program

Sanata Dharma University

ABSTRACT

This paper discusses content validity and authenticity of the English test items in National Examination (UN) year 2012. It is worth discussion because UN, which was administered nationally, was the most important standardized test to assess Indonesian students' competence. 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 writers employed a qualitative research with document analysis to conduct the analysis of both content validity and authenticity of the English test items. The data were obtained from the document and analyzed 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.

Keywords: content validity, authenticity, English test items, National Examination, document analysis

1. INTRODUCTION

National Examination in Indonesia is the highest standardized test employed to assess and measure Indonesian students' competence (Education Ministry Regulation No. 59/2011). By passing National Examination, Indonesian students are able to graduate from a certain education level and to continue their study to the further education level. Therefore, the administration of National Examination is regulated orderly by Education Ministry as well as the test items should be well-prepared

and referring to particular test specifications and lesson objectives. Mentioned in Education Ministry Regulation (No. 22/2006), National Examination materials are generally based on Competence Standard and Basic Competence of each level of educational units which are concluded in Content Standards (*Standard Isi*). Furthermore, Competence Standard and Basic Competence of each level of educational units become reference to create Graduate Competence Standard (*Standard Kompetensi Lulusan*) which consists of test specifications.

Due to the reasons above, the test-makers need to pay attention at least to the test's content validity and authenticity in order to make good test items, particularly in National Examination. Content validity helps the test reflect the measured skills which should be performed by students. American Psychological Association (1985) advances validity of a test is to reveal the relevant test scores (as cited in Rudner and Shafer, 2002, p.12). Seif (2004) explains that if a test does not have content validity, the test-examiners may not be able to determine that the students achieve the set of learning objectives in a particular level of education (as cited in Jandhagi and Shaterian, 2008, p.2).

The test-makers also need to pay attention to the authenticity of the test. Authenticity is important since it builds figures of the target language used in the real situation for students (Brown, 2004). Students will be confused to use language in context unless National Examination reflects authenticity. Moreover, it is important that the materials used in the test are relevant to students' majors in order to ease students in comprehending the content.

The researcher analyzes content validity and authenticity of English test items on National Examination in order to obtain more information about the quality of English items of National Examination year 2012 for senior high schools. The analysis was conducted using document analysis. This is supported by Fraenkel and Wallen (2008) who state that document analysis is useful to prevail information in dealing with educational matters (p.497). In this research, the primary document which is analyzed is the listening recording and the five reading test versions of English test of National Examination year 2012 for senior high schools. The study is based on following two questions:

1. How valid is the content of English test items of National Examination year 2012 for senior high schools related to the lesson objectives and the test specifications?
2. How authentic is English test items of National Examination year 2012 for senior high schools related to the criteria of authenticity set by Brown?

2. LITERATURE REVIEW

A language test is a systematic method to measure someone's capability, knowledge, or performance in a certain domain in its relation with the language use. In order to meet usefulness of a language test, the test should meet a good test's criteria, for instance: reliability, validity, practicality, and authenticity (Brown, 2004). Therefore, the language test should have high quality since it is a measurement of students' capability. In terms of methods, National Examination is a kind of paper-and-pencil language tests or written test and it belongs to receptive tests because it tests somebody's receptive skills such as listening and reading skills. Besides, National Examination is categorized into achievement tests in terms of test purposes (McNamara, 2000).

As an achievement test, National Examination corresponds to the classroom lessons, units, or curriculum (Brown, 2004). The bases of composing National Examination are the Competence Standard, Basic Competence, and Graduate Competence Standard. In order to meet the function as an assessing tool, language test such as National Examination should meet at least two of the principles of language assessment namely content validity and authenticity.

A valid listening test is a test where the content is composed based on the blueprints. If the topics are relevant with the test specifications, the listening test is valid (Brown, 2004). A valid reading test is

a test where the content is composed based on the blueprints. If a language test does not meet content validity it probably affects the students' capability to perform the intended skill and the students are probably not capable to answer the test questions (Seif, 2004). In order to check the validity of language test, the test-designers or teachers are able to check it by matching the test items with the relevant test specifications and lesson objectives.

Authenticity is one of the important language assessment facets since it resemble how the language test show the real-world tasks and true language use (Richards, 2001). It performs the true language in context and they help students by providing appropriate information about the target language (Richards, 2001). In addition, authenticity is a matter of appropriateness of the content and construction of both test tasks and test texts as well as it is not used to teach grammar or language discourse. Instead, it shows genuine and reliable language (Richards, 2001).

In order to determine the authentic assessment, the test-designers should consider two important parts of authenticity namely test task characteristics and test text characteristics (Bachman and Palmer, 1996). Task characteristics include five aspects namely the naturalness of test language, the contextualized items represented in the test, the relevance of the test topics and the learners, the existence of some thematic organization items, and the representativeness of the world tasks (Brown, 2004). The naturalness of test language in reading test items consists of linguistic aspects namely typography, lexis, morphology, syntax, and semantics.

The naturalness of test language shows the appropriateness of the test language to the target language.

The target language use of the English test on National Examination is American English and British English. It is because American English and British English becomes international language as means of communication spoken by most of people throughout the world. The naturalness of listening test refers to the existence of hesitations, white noise, and interruptions in listening tests (Brown, 2004). The contextualization of the test items refers to the test items organizations which are related to the existence of some thematic organization items. Another indicator is relevance of the test topics and the learners which has meaning that the materials should be appropriate to learners. The last indicator is that the tasks should represent the real-world tasks which mean that authentic materials are taken from real-world sources. Besides the test tasks, the test text characteristics become important in order to achieve authenticity and the text characteristics adapt the five indicators of test authenticity. There are three indicators used to check authenticity of reading texts namely the naturalness of test language, the relevance of the test topics and the learners, and the representativeness of the world tasks.

3. DISCUSSION

The results of the analysis on both content validity and authenticity of the test items are presented in the following table:

Table 1 The Percentages of Validity and Authenticity of the Test Items

No	Research Findings	Percentages
1.	The validity of the English test items according to Competence Standard-Basic Competence and Graduate Competence Standard	98.8%
2.	The authenticity of the English test items	79.5%

The table shows that the percentages of validity and authenticity of English test items on National Examination year 2012 were not able reach the highest percentages namely 100% due to some causal factors which were found in data analysis. The reasons are described as follows:

3.1 Content Validity of the Listening Test Items According to Competence Standard and Basic Competence

According to the analysis carried out by the researcher, there was none of the listening test items represents samples of *responding to short spoken functional texts* as stated in Competence Standard and Basic Competence for senior high schools. It is that listening learning topic was not written in Graduate Competence Standard as one of the test materials. Instead, all listening test materials on National Examination year 2012 make reference to the learning topics stated in Competence Standard and Basic Competence for senior high schools. That is supported by Brown (2004), he argues that test specifications include the general outline of the test and the test tasks (p.50). The test specifications in Graduate Competence Standard referred to a certain curriculum and it consisted of only the general outline of whole materials and skills due to test practicality. Therefore, it was not a matter as long as all materials in the listening test make reference to Competence Standard and Basic Competence.

3.2 Content Validity of the Listening Test Items According to Graduate Competence Standard

The results which were obtained show that each listening test items one up to 15 reflected the content of the test specifications on Graduate Competence Standard. It correlates to APA (1954) that content validity refers to the scale of the correlation between the content of the assessment items and domain of interest.

The listening test items includes the listening skill which are going to be measured in the listening test section. It does not include measurements or test items which measure other skills like speaking, reading or, writing skills.

3.3 Content Validity of the Reading Items According to Competence Standard and Basic Competence

Test items on versions A57, B69, C71, D32, and E45 were considered 100% valid in case of the content since the items on each test version of National Examination represented reading topics of learning which was written in Competence Standard and Basic Competence, not other skills. This is in line with Seif (2004) who claims if test does not meet validity in its content, there will possibly be two negative outcomes(as cited in Jandhagi and Shaterian, 2008, p.2). In this case, the two negative outcomes would not happen since all test items in the five test versions of National Examination year 2012, based on Competence Standard and Basic Competence.

3.4 Content Validity of the Reading Test Items According to Graduate Competence Standard

The data analysis shows that there were three test versions did not represent explanation text. In this case, those three test versions of National Examination year 2012 for senior high schools lacked content validity. If the reading test of National Examination completely refers to Graduate Competence Standard, there will be 13 kinds of reading materials (message, letter/e-mail, advertisement, narrative text, news item, recount text, announcement, report, descriptive text, explanation, exposition text, discussion, and review) on the reading test. Instead, test versions A57, C71, and D32 only demonstrated 12 kinds of reading materials while B69 and E45 could demonstrate all 13 kinds of reading materials which were written on Graduate Competence Standard.

It showed the difference in quality of content validity in all five test version. This is in line with ElindDriana's statement, an observer in educational field, *Tempo* (Tuesday, 11 September 2012). ElindDriana (2012) states composing several versions of National Examination should consider that all those various test versions should have the same quality of difficulties. Therefore, it is important that each test version should have the same kinds of test instructions and test topics in order to meet validity of the result.

3.5 Authenticity of the Listening Test Items

. There was one significant problem related to the naturalness of language use in listening test items. There was no significant problem related to other factor namely contextualization of the test items, thematic item organization, relevance of the test topics to the learners, and real-world representativeness. The language used in the conversations was similar to the real-world conversations and there were also some word reduction in order to make the conversations natural. In the listening test question number 2, for instance, the woman reduced the word *did* and *not* into *didn't*. However, there was no hesitations and white noise found in the conversations. Therefore, the conversations sounded like designed recordings. According to Brown (2004), there are two of three features which can be used to express natural language use in listening comprehension section; they are hesitations and white noise (p.28).

Afterwards, all listening test items on National Examination year 2012 are considered as contextualized items because the test items are developed from two learning topics integrated in the blueprints namely transactional/interpersonal expressions and monologue texts. Besides, all learning topics of the fifteen test items on the

listening test are relevant for the learners. The learners in this context are senior high school students and the learning topics used in the conversations are about *asking for and giving direction, expressing pleasure, thanking, complaining, asking for and giving information, and offering help*. All topics in the listening test take place in daily-life situation. In the listening test on National Examination year 2012, the researcher found out that four test items are organized in a form of story lines. Lastly, the real-world representativeness could be exhibited in all listening test items. The conversations and the spoken monologue texts often take place in daily-life situation.

1.6 Authenticity of the Test Tasks

The preliminary data shows that the total different test items from A57, B69, C71, D32, and E45 were 123 test items and there were 50 different passages employed in it. The researcher also recognized that most of the test tasks had problem to fulfill the naturalness of language used in the test instructions and the optional answers as well as the relevance of a particular test topic for the learners. Although the language test was not intended to test some grammatical or lexical items, the test-designers should avoid linguistic mistakes in order to represent highly authentic reading test.

According to Richards (2001), the visible characteristic of authentic materials was that it provides true language (pp.252-253). It means that there should no linguistic mistakes such as typographical mistakes, lexis, morphemes, word orders and grammar (syntactic matters), diction, and meaning (semantic matters) in the test tasks in order to avoid test takers' confusion in understanding the test instructions. From 123 test tasks or test instructions there are only 105 test items which meet the natural language use criterion. Consequently, the test takers were possibly confused in

understanding the meaning. It was related to Widdowson (1976) who emphasizes that authenticity is not only about the quality of a text at all but authenticity is reached when the readers understand the writer's intention (p.264). The other mistake belongs to morphosyntactic mistake which is related to singular and plural forms. It is related to the use of determiner as well.

The researcher also considered that all the reading test items on National Examination year 2012 are contextualized. All test tasks were developed from certain learning topics namely functional texts and essays. In relation to the thematic items organization, the researcher identified there were 118 test tasks constructed thematically while there were five test items constructed independently. Besides, the test tasks on the reading test do not attempt to ask for English grammatical forms but it indicated asking for information or the meaning of some vocabulary. Lastly, the relevance of a particular test topic to senior high school students becomes a problem in the reading test tasks of National Examination year 2012 since there are two test tasks found in A57 test version were considered not relevant to senior high school students.

3.7 The Authenticity of the Test Texts

The result of the analysis shows that most test texts face problem to fulfill the naturalness of language used in the test passages and the real-world representativeness as well as the relevance of a particular test topic to the learners. According to the data, there was only 36% of the test texts which met the indicator of naturalness of the language used in the test texts. The failure of the test passages to meet the indicator was caused by the existence of linguistic facets like: typographical mistakes, lexis, morphemes, word orders and grammar (syntactic matters), diction, and meaning.

Afterwards, there were only 98% of the test text topics which were relevant to senior high school students. The topic was not relevant to senior high school students because the passage used specific terms related to electrical installation.

Almost all of the passages used in the reading test failed to represent the real-world context even though the topics of the passages were rational and based on real-world context. Unlike what Brown (2004) states that authentic reading passages are taken from real-world sources (p.28), meanwhile the test-designers of the English test items did not mention the sources where the passages were taken from. Another reason was that the samples of the formal letters, announcements, and the advertisements look unnatural viewed from the format and design.

3.8. Other Findings

The main goal of the Education Ministry by applying different kinds of test version in National Examination year 2012 was to clamp down on students' fraudulence in the implementation of National Examination. From the pre-analysis, the researcher found out interesting results. It was that there were several similar passages and test questions used in all five test versions. The other interesting findings were that most of the passages and the test tasks in test version C71 were similar to the passages in test version D32. The difference between both test versions was only found in the test item numbers 39, 40, 41 of both test version since the passages used in each test version, related to those three questions, were different. It implied a mismatch between the Education Ministry's objectives to apply several test versions in National Examination and the facts founded in reading test items of National Examination year 2012.

4. CONCLUSIONS

There were several findings in this research concerning to answer two research questions as follows. First, it was found that the content of the test items (including listening and reading test items) are 98.8% valid. Second, the results of analyzing the authenticity of National Examination listening test items year 2012 show that the listening and reading test items are authentic. It reaches 79.5% as the percentages of the authenticity. According to the findings, it implies that the English test items of National Examination year 2012 need evaluation and improvement since both content validity and authenticity of the English test items on National Examination year 2012 are not able to reach the highest percentages.

REFERENCES

- American Psychological Association. (1954). *Technical recommendations for psychological tests and diagnostic techniques: Preliminary proposal*. American Psychologist, 7, 461-476.
- American Psychological Association. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. USA: Pearson Education, Inc.
- Fraenkel, J. R., & Wallen, N. E. (2008). *How to design and evaluate research in education* (7thed.). Boston: McGraw-Hill Higher Education.
- Gronlund, N. (1998). *Assessment of student achievement* (6thed.). Needham Heights, MA: Allyn & Bacon.
- Jandaghi, G., & Shateria, F. (2008). Rate of validity, reliability and difficulty indices for teacher-designed exam questions in first year high school. *International Journal of Human Sciences* [Online]. 5:2. Retrieved September 19, 2013 from <http://www.insanbilimliri.com>
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Oxford Advanced Learner's Dictionary (7thed). (2005). Oxford: Oxford University Press.
- Peraturan Menteri Pendidikan Nasional No. 22 tahun 2006 tentang Ujian Nasional.
- Peraturan Menteri No. 59 tahun 2011 tentang Ujian Nasional.
- Richards, J. C. (2001). *Curriculum development in language teaching*. New York: Cambridge University Press.
- Seif, A. A. (2004). *Assessment and Evaluation in Education*. Tehran: Doran Publication. Retrieved September 19, 2013 from <http://www.insanbilimliri.com>
- Tempo Interactive. (June, 2012). National exams warrant re-evaluation. *Tempo*. Retrieved on August 31, 2012, from <http://www.tempo.co/read/news/2012/06/.../055409165>
- Anggrita Desyani. (September, 2012). Rencanavariasi 20 soal ujiannasional dikritik. *Tempo*. Retrieved on September 19, 2013, from <http://www.tempo.co/read/news/2012/09/11/079428853/Rencana-Variasi-20-Soal-Ujian-Nasional-Dikritik>
- Widdowson, H. G. (1976). The authenticity of language data. In John F. Fanselow & Ruth H. Crymes (Eds.), *TESOL '76* (pp. 261-270). Washington, D.C.: TESOL.