

Does the National Examination of English Subject in Indonesia Test What Should Be Tested?

Dwi Rukmini

(wiwikdwirukmini@yahoo.com)

Jumaroh

(kirana.embun@gmail.com)

Abstract. *This article examines the content validity of the test items of English National Examination held by the Indonesian government. The analysis is done by comparing the objectives of English teaching and learning at junior and senior high schools, which are mentioned by the implemented curriculum, with the test items given in English National Examination. The result shows that the English National Examination does not have content-evidence of validity. We can say so since the test-items do not cover all sections required by the curriculum which are listening, speaking, reading, and writing.*

Key words: *content validity, English National Examination*

Introduction

Education is one of the most important things in any nation building process. The quality of education system in a country will greatly influence the development of the country. Therefore, the developmental progress of a country can be measured from the quality of its education because education is a crucial key in improving human resources. Every country has a different way of standardizing their education system. Most countries adopt something which is called 'national examination.' Every level of study (e.g. elementary school, junior high school, senior high school, etc.) is given this kind of standardized test to pass. In Indonesia, we call this *Ujian Nasional* (UN) or the National Examination. Regulation of the Minister of National Education No. 75 of 2009 about national high school examination states that the National Examination is an activity of measuring and assessing of learners' competence at the level of primary and secondary education. The National Examination, previously known as *Ujian Negara* (government test), the Ebtanas, and UAN or *Ujian Akhir Nasional*, is one of the summative evaluations to be done for the basic summary of the students' learning process.

As English is a compulsory subject in all junior high schools, senior high schools, and vocational schools, it becomes a subject which is always tested in the National Examination in Indonesia. Through this annual agenda, it is expected that the development of English teaching in Indonesia can be surveyed regularly. Through the result of this examination, it is expected that the government can regularly evaluate the curriculum and its implementation at schools, and generate new resolution for the betterment of the English teaching and learning process which ends in better students' competency in English. However, there are various negative comments towards the exams. Some people state that the national examination is an unfair way to assess the students' achievement (Davies, 2008); others say that it cannot assess the whole competencies the students' have (Suparno, 2009). There is also an opinion that the National Examination gives a lot of pressure to students and teachers (Widjojo, 2010). Moreover, there is also an opinion that the National Examination, especially in the English subject, is not valid. It does not measure what it should measure. It does not test what should be tested from the students. This opinion will be the basic assumption underlying this article.

What is a Valid Test?

Validity has been identified as the most important quality of test. It concerns the extent to which meaningful inferences can be drawn from test scores (Bachman, 1990). 'A valid test of reading ability actually measures reading ability – not 20/20 vision, nor previous knowledge in a subject, nor some other variable of questionable relevance' (Brown, 2004). In other words, a valid test is test measuring what should be measured, assessing what it should assesses.

To measure speaking ability, a teacher may ask his students to mention any English word as much as they can within fifteen minutes, and then simply counts the words they have mentioned for the final score. Such a test would be easy to be administered. It will be a practical test. It would also be reliable since the scoring would be very dependable. However, such a test cannot be claimed as a valid test of speaking skill since it excludes some consideration such as context or rhetorical discourse elements, comprehension, the organization of the ideas, and so on (Brown, 2004).

Based on Brown (2004), there are five types of evidence that a test is valid. Those types of evidence are content-related evidence, criterion-related evidence, construct-related evidence, sequential validity, and face validity.

If a test contains samples the subject matter about which conclusions are to be drawn, and if it requires the test-takers to perform the behavior that is being measured, we can claim that the test possesses the content-related evidence of validity, or we can say that it has content validity (Mousavi and Huges in Brown, 2004).

Criterion-related evidence of validity or criterion-related validity is the extent to which the criterion of the test has actually been reached. It usually falls into one of two categories: *concurrent* and *predictive* validity. A test has concurrent validity if the result of the test is supported by other concurrent result of other kind of assessment beyond the test. The predictive validity of a test is important since it aims to predict a test-taker's likelihood of future success (Brown, 2004).

The third type of evidence concerns with construct-related validity or construct validity. The term 'construct' itself refers to any theory, hypothesis, or model that attempts to explain observed phenomena in our universe of perception (Brown, 2004). Construct-related validity asks 'Does this test actually tap into the theoretical construct as it has been defined?' Tests are operational definitions of construct in that they operationalize the entity that is being measured (Davidson, Hudson, and Lynch in Brown, 2004).

Consequential validity encompasses all the consequences of a test. It includes the consideration of the test accuracy in measuring the intended criteria, the impact of the test on the preparation of the test-takers, the effect that the test has on the learners, and the social consequences of a test's interpretation and use. In Brown (2004), Gronlund encourages teachers to 'consider the effect of tests on students' motivation, subsequent performances in a course, independent learning, study habits, and attitude toward school work'.

The last kind of evidence of validity is the face validity. It refers to 'the degree to which a test *looks* right, and *appears* to measure the knowledge or abilities it claims to measure, based on the subjective judgment of the examinees who take it, the administrative personnel who decide on its use, and other psychometrically unsophisticated observers (Mousavi on Brown, 2004).

The five types of evidence above are used to examine the validity of a test. However other writers propose another way to do so. Based on their opinion, examining validity of a test requires a validation process by which a test user presents evidence to support the inferences or decisions made on the basis of test scores (Crocker & Algina, 1986). Validation studies of language performance assessment are mainly concerned with three types of validity: construct validity, predictive validity, and content validity.

Construct validity is associated with two distinctive approaches to performance assessment development; they are the construct-centered approach and the task-centered approach (Bachman, 2002). The task-centered approach has been favored over the construct-centered approach by some proponents of performance assessment. For instance, a group of researchers at the University of Hawaii at Manoa argue that performance on a task-based test itself is the construct of interest, indicating that predictions to be made are about the test-takers' abilities to accomplish certain tasks (Brown, Hudson, Norris, & Bonk, in press, as cited in Bachman, 2002). Building on this definition, task-based language performance assessment is regarded as one type of performance assessment where the construct of interest is task performance itself. One potential problem with this approach, however, is that inferences may not be made beyond a specific testing context, which thus severely weakens the interpretation and generalization of test results (Bachman, 1990). Bachman (2002) therefore argued that both task-centered and construct-centered approaches should be adopted in the performance-based test design.

The other pivotal validity considerations in second or foreign language testing are predictive validity and content relevance and coverage. Since the major purpose of performance tests is to provide predictive information about how well the test-takers will use the second language under specific target conditions, predictive validity has been one of the primary concerns in performance assessment (Wesche, 1985). How accurately a prediction can be made relies on the degree of content validity.

Content validity involves two crucial concepts: content relevance and content coverage (Bachman, 1990). Content relevance refers to the extent to which the aspects of ability to be assessed are actually tested by the task, indicating the requirement to specify the ability domain and the test method facets (Bachman, 1990). Content coverage concerns the extent to which the test tasks adequately demonstrate the performance in the target context, which may be achieved by randomly selecting representative samples (Bachman, 1990).

The second aspect of content validity is similar to that of content representativeness, which also concerns the extent to which the test accurately samples the behavioral domain of interest (Bachman, 2002). Some problems in investigating content validity have been identified by language testers (Bachman, 2002). First, difficulties may arise in defining the target language universal domain in a situation where examinees come from diverse backgrounds and have widely ranging needs in language use. Furthermore, even when the target language universal domain can be well defined, selecting representative samples from that domain may be problematic (Bachman, 2002). As pointed out by Hughes (1981), it is quite difficult to sample representative language skills as a result of inadequate needs analyses and the lack of comprehensive and complete descriptions of language use. This sampling problem may complicate and lengthen the test design (Jones, 1979).

Some attempts have so far been made to identify representative samples. For instance, Branden, Depauw, and Gysen (2002) highlighted the value of needs analysis for sampling tasks in the instructional and learning contexts. However, needs analysis has been challenged in cases where test-takers come from various backgrounds. These challenges may pose a serious extrapolation problem beyond a specific testing context.

As claimed by Bachman (2002), 'ill-defined or indeterminate relationships between assessment tasks and target language universal tasks affect extrapolation' (pp. 458-459). Some empirical attempts have been made to investigate the extrapolation issue in relation to generalizations across test tasks. For example, the findings in Brindley and Slatyer's (2002) study demonstrated generalizability problems in performance assessment and indicated an urgent need for a detailed exploration of sources of variation that may affect test-takers' performance. The comparability of various tasks has often been questioned

and thus the generalization of performance on a certain test task to the broader universe of test tasks has been called into question (e.g., Bachman, 2002).

Based on the long description about validity and the topic of this essay, the types of evidence that will be used to analyze the validity of the English national examination for junior and senior high schools in Indonesia is the content-related evidence or the content validity.

Evaluating the Content Validity of a Test

On his book, Brown (2004) explains two steps to evaluate the content validity of a classroom test. The first is 'Are the classroom objectives identified and appropriately framed?', and the second is 'Are lesson objectives represented in the form that lends itself to assessment?'

The first step deals with how any objective of the teaching and learning activities is well framed. It must be clearly stated (no ambiguous constituent), neither too easy nor too difficult, the standard is clearly defined, can be assessed, and neither too general nor too specific. The second suggests that the content validity of an existing test should be apparent in how the objectives of material being tested are represented in the form of the content items.

However, both steps are used to evaluate the classroom tests. Since the observed object of this writing is the national examination, the first step is expected to be fulfilled well because all the objectives are well-prepared and well-presented in the existing curriculum. The second step is the concern of this essay and will be discussed in the next subsections.

Students' Competencies Expected by the Curriculum

To analyze the content validity of the English national examination, we have to take a look on the objectives of the whole set of teaching and learning process. Those objectives will be compared to the items of the examination afterwards. To make the explanation simple, I will only present the learning objectives of senior high school students required by the curriculum. The explanation of the English examination in junior high schools will only be presented briefly. Bellow are the learning objectives of senior high school students based on *KTSP* or school-based curriculum. I do not analyze the 2013 curriculum since the national examination based on this curriculum has not been conducted yet.

Grade X, Semester 1

Listening

1. Understanding transactional and interpersonal conversation in everyday context
2. Understanding short functional texts and monologues in the form of recount, narrative and procedure texts in everyday context

Speaking

3. Expressing meaning through transactional and interpersonal conversation in everyday context
4. Expressing meaning within short functional texts and monologues in the form of recount, narrative and procedure in everyday context

Reading

5. Understanding the meaning of short functional texts and simple essay in the form of recount, narrative and procedure texts in everyday context and to access science

Writing

6. Expressing meaning in written short functional texts and simple essay in the form of recount, narrative, and procedure in everyday context

Grade X, Semester 2

Listening

7. Understanding transactional and interpersonal conversation in everyday context
8. Understanding short functional texts and monologues in the form of recount, narrative, descriptive, and news item texts in everyday context

Speaking

9. Expressing meaning through transactional and interpersonal conversation in everyday context
10. Expressing meaning within short functional texts and monologues in the form of recount, descriptive and news item texts in everyday context

Reading

11. Understanding the meaning of short functional texts and simple essay in the form of narrative, descriptive and news item texts in everyday context and to access science

Writing

12. Expressing meaning in written short functional texts and simple essay in the form of narrative, descriptive, and news item in everyday context

Grade XI, Semester 1

Listening

1. Understanding formal and sustained transactional and interpersonal conversation in everyday context
2. Understanding meaning in short functional texts and monologues in the form of report, narrative and analytical exposition texts in everyday context

Speaking

3. Expressing meaning through formal and sustained transactional and interpersonal conversation in everyday context
4. Expressing meaning within short functional texts and monologues in the form of report, narrative and analytical exposition in everyday context

Reading

5. Understanding the meaning of short functional texts and simple essay in the form of report, narrative and analytical exposition texts in everyday context and to access science

Writing

6. Expressing meaning through essay in the form of report, narrative, and analytical exposition in everyday context

Grade XI, Semester 2

Listening

7. Understanding formal and sustained transactional and interpersonal conversation in everyday context
8. Understanding meaning in short functional texts and monologues in the form of narrative, spoof, and hortatory exposition texts in everyday context

Speaking

9. Expressing meaning through formal and sustained transactional and interpersonal conversation in everyday context
10. Expressing meaning within short functional texts and monologues in the form of narrative, spoof and hortatory exposition in everyday context

Reading

11. Understanding the meaning of short functional texts and essay in the form of narrative, spoof, and hortatory exposition texts in everyday context and to access science

Writing

12. Expressing meaning in written short functional texts and simple essay in the form of narrative, spoof, and hortatory exposition in everyday context

Grade XII, Semester 1**Listening**

1. Understanding formal and sustained transactional and interpersonal conversation in everyday context
2. Understanding meaning in short functional texts and monologues in the form of narrative, explanation, and discussion texts in everyday context

Speaking

3. Expressing meaning through formal and sustained transactional and interpersonal conversation in everyday context
4. Expressing meaning within short functional texts and monologues in the form of narrative, explanation, and discussion in everyday context

Reading

5. Understanding the meaning of short functional texts and essay in the form of narrative, explanation and discussion texts in everyday context and to access science

Writing

6. Expressing meaning in written monologues in the form of narrative, explanation, and discussion review accurately, fluently, and understandably in everyday context

Grade XII, Semester 2**Listening**

7. Understanding formal and sustained transactional and interpersonal conversation in everyday context
8. Understanding meaning in short functional texts and monologues in the form of narrative and review in everyday context

Speaking

9. Expressing meaning through interactional text, especially formal and sustained transactional conversation in everyday context
10. Expressing meaning within short functional texts and monologues in the form of narrative and review accurately, fluently, and understandably in everyday context

Reading

11. Understanding written monologue in the form of narrative and review accurately, fluently, and understandably in everyday context and to access science

Writing

12. Expressing meaning in written monologues/essays in the form of narrative and review in everyday context

Based on the long description above, we can see that through the English curriculum, the government expects that the English teaching and learning activities in senior high schools should cover the four language skills: listening, speaking, reading, and writing. Therefore, the English national examination should cover all of the competencies mentioned by the curriculum. The analysis comparing the English national examination and the learners' competencies required by the curriculum (the government) will be presented in the next subsection.

Does the national examination of English subject in Indonesia test what should be tested?

In brief, the expected competence in listening from Grade X to Grade XII is the learners are able to understand transactional and interpersonal conversation in daily

context, and to understand meaning in various short functional texts and monologues in daily context. Then in speaking skill, the learners are able to express meaning through formal and sustained transactional and interpersonal conversation in everyday context, and to express meaning within various short functional texts and monologues in daily context. In reading skill, the learner are expected to be able to understand the meaning of written short functional texts and essay in the form of various text types in everyday context and to access science. In the last language skill, writing, the learners are expected to be able to express meaning through essays in the form of various text types in everyday context.

If we take a look on the composition of the competencies that should be reached by the students, we will find that there are two listening competencies from six competencies required in each semester. Similar proportion is allocated to speaking skill. However, only one competence is required for each of reading and writing skill each semester. It means that from the six competencies required in each semester, the comparison between listening : speaking : reading : writing is 2 : 2 : 1 : 1, or 2 : 1 in the comparison of aural oral skills and the literacy skills. This proportion describes the goal or the expectation of the curriculum is emphasized in the communicative competence. Through this proportion, it is expected that the test items in English national examination for senior high school students will has similar proportion, or at least it includes all of the four language skills: listening, speaking, reading, and writing.

However, the contradiction appears if we take a look on the actual sheets of the English national examination. Each set of test items consists of fifty questions in the form of multiple choices. The comparison between spoken and written test is 15 : 35. Those items only covers listening, reading, and writing test. From fifty items, fifteen items are listening section items, thirty one items are reading section, and the last four are writing section. There is no item testing the students' speaking skill.

The test makers may argue that the speaking items are integrated into the items of listening section. However, how can we know that a student can speak or not simply by ask him to select a possible answer? Just like Brown's statement saying that a valid test is test measuring what should be measured, assessing what it should assesses (2004). We cannot simply ask the students to choose some possible responses for certain situations then claim that the students can speak well. It is in line with the example presented by Brown (2004). In such cases, Brown points out that 'such a test would be easy to be administered. It will be a practical test. It would also be reliable since the scoring would be very dependable.' However, such a test cannot be claimed as a valid test of speaking skill since it excludes some consideration such as context or rhetorical discourse elements, comprehension, the organization of the ideas, and so on.

Another minor section in the national examination is the writing skill. There are four questions in writing section. Three of them are in the form of short-answer and sentence completion tasks. This type of test belongs to intensive (controlled) writing test. However, the argument comes because of the rest the writing section. In this item, the students are asked to rearrange some sentences into a good paragraph. Since even in his book Brown (2004) does not mention this kind of test, we can say that this test item is not a valid way to measure the students' competence in writing. The example of this test item is given below.

32. Arrange these sentences below into a good order.

1. Grate the cheese into the bowl and stir.
2. Then turn omelette with a spatula when it brown.
3. First, crack an egg into a bowl.
4. Heat the oil in a frying fan, and pour the mixture into the frying fan.
5. Next cook both side and it is ready to serve.

- A. 3 - 5 - 1 - 2 - 4 - 6
- B. 3 - 6 - 1 - 4 - 2 - 5
- C. 3 - 2 - 6 - 1 - 4 - 5
- D. 3 - 2 - 4 - 1 - 6 - 5
- E. 3 - 1 - 6 - 4 - 2 - 5

The explanation above is given to answer the question by Brown (2004): 'Are lesson objectives represented in the form of test specifications?' Then, to conclude that the English national examination has content validity or not, I will also consider the two crucial concepts by Bachman (1990): content relevance and content coverage. Talking about the content relevance, the test items of the English national examination are relate to the objective of the English teaching and learning process required by the curriculum. However, when we talk about the content coverage, I think we can conclude that the examination does not fully cover the competence required by the curriculum. In the examination, the speaking skill even has never been tested.

When we turn to the English national examination for junior high school students, we can see that it even has worse content coverage since none of the test items covers the listening and speaking sections. Almost all of the items only test the reading skills of the students. However, this condition leads to a bad effect where in some schools, the English teaching and learning process is test-oriented and minimizes the allocation time for aural-oral skills activities.

Conclusion

From the explanation of the previous subsections above, we can conclude that the English national examination does not have content-evidence of validity. We can say so since the test-items do not contain all sections required by the curriculum which are listening, speaking, reading, and writing.

If we consider the content-coverage of the examination, we can also say that the English national examination does not cover all of the learning objectives required by the implemented curriculum.

References

- Bachman, L. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. 2002. Some reflections on task-based language performance assessment. *Language Testing*, 19, 453-476.

- Branden, K., Depauw, V., & Gysen, S. 2002. A computerized task-based test of second language Dutch for vocational training purposes. *Language testing*, 19, 438-52.
- Brindley, G., & Slatyer, H. 2002. Exploring task difficulty in ESL listening assessment. *Language Testing*, 19, 369-94.
- Brown, H.D. 2004. *Language Assessment: Principles and Classroom Practices*. White Plains: Longman.
- Crocker, L., & Algina, J. 1986. *Introduction to classical and modern test theory*. Belmont: Wadsworth Group/Thomson Learning.
- Davies, R. 2008. *Exams important part of education process*. From <http://www.thejakartapost.com/news/2008/11/09/exams-important-part-education-process.html>.
- Hughes, A. 1981. Conversational cloze as a measure of oral ability. *English Language Teaching Journal*, 35, 161-168.
- Jones, R. 1979. *Performance testing of second language proficiency*. In E. Briere & F. Hinofotis (Eds.), *Concepts in language testing*. Washington, DC: TESOL.
- Suparno, P. 2009. *Quality of Education and the national exams*. From <http://www.thejakartapost.com/news/2009/12/19/quality-education-and-national-exams.html>.
- Wesche, M. 1985. *Introduction*. In P. C. Hauptman, R. LeBlanc, & M. B. Wesche (Eds.), *Second language performance testing*. Ottawa: University of Ottawa Press.
- Widjojo, O. 2010. *Letter: the National Examination*. From <http://www.thejakartapost.com/news/2010/04/07/letter-the-national-examinations.html>.

