

QUALITY OF LISTENING TEST VALIDITY OF THE NATIONAL LEAVING EXAM FOR SMA

Wuwuh Yunhadi

University of Kutai Kartanegara

Email: wuwuh_dr@yahoo.co.id

Abstract

This study reports quality of content validity of items of listening test of National Leaving Exam for SMA in Samarinda is constructed. The quality was identified from task coverage, item difficulty, objective of item, and test format. This study used content analysis with qualitative design. The study revealed that content validity revealed three major findings. Task coverage is dominated by the first and the second year of SMA (8 items are from first year, 8 items are from second year, and 4 items are from third year of SMA). Of 20 items, 9 (45%) are easy, 7 (35%) moderate, 4 (20%) hard. The questions in the items are specified into 9 (45%) general, 5 (25%) comprehension, 4 (20%) application, and 2 (10%) analysis. The analysis on the Table of Specification reconstructed from the English listening test showed that the test has low content validity.

Keywords: National Leaving Exam, listening, content validity.

Received: 15 January, 2016; Accepted: 15 Marech, 2016

1. INTRODUCTION

Reconstruction of tests for National Leaving Exam (NLE) or Ujian Akhir Nasional (UAN) and tests generally in adherence to content coverage is doubtful for two reasons. First, the construction of tests depends on test-makers' logic. When it uses a test-maker's logic, the quality of content validity of a test should be analyzed to prove its content validity. This implies that the appropriateness between the test items and the instructional objectives that have been taught is not completely appropriate. Indirectly, this case will cause the second problem, indicating inappropriateness between the Table of Specification of the test and its content, from which will influence the quality of content validity of the test.

In determining the content validity, the curriculum is useful for the test-makers and the teachers. The test-makers use the curriculum as their reference to construct the test items and the teachers use it to elaborate items of the test coverage. Although the test-makers and teachers use the same curriculum for reference, there is no guarantee that the instructional objectives in the curriculum have been properly described.

This study focuses on the analysis of English Listening test for NLE for SMA in Samarinda. The third year students of SMA are divided into two study programs: (1) science, and (2) social sciences. The tests of NLE for science classes are Mathematics test, Indonesian language test, and English tests; whereas tests for social science classes include Economics test, Indonesian language test, and English test. This implies that the English test for NLE for

SMA exists in each study program. The test consists of two sections: (1) listening test section, and (2) reading test section. Listening section is a new model of NLE because in several previous periods, the listening test was not used in English test for SMA.

As a new test, listening test arises problem particularly from the views of Table of Specification. The Table of Specification is a matrix that shows task coverage, objective of items, test format, and item difficulty. Particularly, this study examines the appropriateness between the English listening test and instructional objectives in English curriculum for SMA. The items should be analyzed in order to prove quality in content validity.

Content validity as one of logic validities demands a content appropriateness between measured skill and test used to measure it. The appropriateness is shown in the kind of skills demanded to do the test compared with the kind of skills, which becomes the objective of measurement (Djiwandono, 1996:92-93).

Furthermore, what is the importance of content validity? Hughes (1989:22-23) says that the first importance of content validity is the greater a test's content validity, the more likely it is to be an accurate measure of what it is supposed to measure. A test in which major areas identified in the specification are under represented at all is unlikely to be accurate. The second one, such a test is likely to have a harmful backwash effect. Areas, which are not tested, are likely to become areas ignored in teaching and learning. Too often the content of tests is determined by what is easy to test rather than what is important to test.

The above description will cause a question: how to determine the content validity of the test? For this question, Saukah (1997:171) says that content validity is determined by using table of specification (blueprint). It contains of task coverage, objective of items, difficulty index (p), and the format of the test. Actually, the Table of Specification is a plan used as a reference to develop a test that focuses on task coverage, proportion of items, difficulty index, and the test formats used in test being developed.

In evaluation of language teaching, task coverage refers to curriculum and objective of items that refers to Bloom's taxonomy. They are knowledge, comprehension, application, analysis, synthesis, and evaluation. Bloom in Gronlund (1985:514) defines each aspect in Bloom's taxonomy as follows:

- 1) **Knowledge.** Knowledge is defined as the remembering of previously learned material. This may involve the recall of a wide range of material, from specific facts to complete theories, but all that is required is the bringing to mind of the appropriate information. Knowledge represents the lowest level of learning outcomes in the cognitive domain.
- 2) **Comprehension.** Comprehension is defined as the ability to grasp the meaning of material. This may be shown by translating material from one form to another (words of numbers), by interpreting material (explaining or summarizing), and by estimating future trends (predicting consequences or effects). These learning outcomes go one step beyond the simple remembering of material, and represent the lowest level of understanding.
- 3) **Application.** Application refers to the ability to use learned material in new and concrete situations. This may include the application of such things as rules, methods, concepts, principles, laws, and theories. Learning outcomes in this area require a higher level of understanding than under comprehension.

- 4) **Analysis.** Analysis refers to the ability to break down material into its component parts so that its organizational structure may be understood. This may include the identification of the parts, analysis on the relationship between parts, and recognition of the organizational principles involved. Learning outcomes here represent a higher intellectual level than comprehension and application because they require an understanding of both the content and the structural form of the material.
- 5) **Synthesis.** Synthesis refers to the ability to put parts together to form a new whole. This may involve the production of a unique communication (theme or speech), a plan of operations, (research proposal), or a set of abstract relations (scheme for classifying information). Learning outcomes in this area stress creative behaviors, with major emphasis on the formulation of new patterns or structures.
- 6) **Evaluation.** Evaluation is concerned with the ability to judge the value of material (statement, novel, poem, and research report) for a given purpose. The judgments are to be based on the definite criteria. These may be internal criteria (organization) or external criteria (relevance to the purpose) and the student may determine the criteria or be given them. Learning outcomes in this area are the highest in the cognitive hierarchy because they contain elements of all of the other categories, plus value judgments based on clearly defined criteria.

To make a good test, use these action verbs, teachers will be helped to make item appropriate with measured skills. Davis in Nugroho (2001:4) gives prescription for action verbs as in Table 1.

Table 1 Action Verbs for Achieving Cognitive Aspects

Aspects	Action Verbs			
Knowledge	define know reproduce	Write choose call	Underline Enlist label	mention propose measure
Comprehension	identify present point out	draw give a reason formulate	explain consider oppose	choose propose classify
Application	predict calculate	choose evaluate	arrange point out	find demonstrate
Analysis	analysis separate	identify solve	give a reason compare	criticize spell out
Synthesis	combine arrange	conclude connect	generalize argue	enclose shorten
Evaluation	evaluate	determine	consider	maintain

The third component in table of specification is difficulty index. Difficulty index or item difficulty is differentiated into 3 levels: easy, moderate and difficult. The last component is the test format; it can be multiple choice, matching, completion, essay, etc. In example below, A, B and C are code for the format of the test (multiple choice, completion or essay).

To get a clear explanation, Nugroho (2001:3) gives an example of the Table of Specification and the table is below. It is a matrix of the Table of Specification that contains of task coverage, objective of item, the test format, and the item difficulty (p).

Table 2 Example for Table of Specification

Subject :
 Class :
 Task Coverage :
 Time :

Task coverage	Item Difficulty (p)	Knowledge			Comprehension			Application			Analysis			Synthesis			Evaluation			Total
		A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	
	Easy																			
	Moderate																			
	Difficult																			
	Easy																			
	Moderate																			
	Difficult																			
	Easy																			
	Moderate																			
	Difficult																			
	Easy																			
	Moderate																			
	Difficult																			
	Easy																			
	Moderate																			
	Difficult																			
Total Based on (p)																				
Total based on objective of items																				

The English Listening test for SMA is said to have quality of content validity if the test materials represent the materials that have been taught based on the curriculum. The curriculum being used is an English curriculum for SMA and it consists of standardized competence, basic competence, the indicators, and basic materials for each class in SMA.

To obtain an English listening test for SMA that has quality in content validity, English teachers and test makers should use the English curriculum for their activities. Table below shows the outline of English curriculum that has been limited on the action competence in listening comprehension. The table consists of the class, the indicators, and the core materials. The class in the table is meant as the year of SMA whether the first year, second year, or the third year of SMA. Then, the indicators are used to indicate the materials for each class. The last component is the core materials. It is expected to give examples for each indicator in each class.

Table 3 Outline of English Listening Curriculum for SMA

No	Class	Indicators	Core Materials
1.	1	Respond correctly conversation in transactional/interpersonal discourse such as: <ul style="list-style-type: none"> • Receiving things/service • Introduction • Meeting • Agreement • Accepting promises • Replying thanks expression • Replying congratulations • Showing attention, surprise, sympathy, happiness, unbelievable 	Examples for conversational expression: <ul style="list-style-type: none"> • Thank you so much. • Pleased to meet you too. • Hello! • See you. • Take care! • Thanks for the invitation. • That's OK. • Let's meet at four. • That's all right.

The above description implies that the English Listening test items for NLE 2006/2007 for SMA in Samarinda should be evaluated to measure students' listening comprehension. It is viewed from the construction of the test and the Table of Specification of test. The construction of the test depends on the test-makers' logic and it will cause the uncertainty of content validity. In other side, it is estimated that the Table of Specification of the test and its items are not appropriate because it has not been checked. The hesitancy of content validity of English listening test for SMA in Samarinda encourages the researcher to evaluate the quality of content validity of the test.

2. METHOD

This study employed evaluation research as the design of analysis. Carol in Darma (2005:34) cites that evaluation research is aimed to increase the rationality of policy making. Using objective information of the outcomes of programs, wise decisions can be made on the budget allocation and program training. Program that yield good result will be expanded; those that make poor showings will be abandoned or drastically modified. In addition, Carol in Darma (2005:34) argues that the purpose of evaluation research is to measure the effect of a program against the goal for contribution in the decision-making. Furthermore, Budiharso in Darma (2005:34) states that the contribution can be useful to evaluate the methodology of the program, the outcomes of the program, the criteria used in the program and the social purpose of the program.

The current study was aimed at seeing the process of producing English listening test for NLE of SMA students. The main purpose of the study was to clarify the methods how the English listening test of NLE is constructed. A set of NLE used as a test for SMA in Samarinda was selected its listening items and the items were evaluated their criteria. The evaluation was based on the criteria of procedure of developing a standardized test.

Based on the above rationale, the variable of this study was quality of content validity of English listening test of NLE for SMA in Samarinda. The quality was determined using the Table of Specification reconstructed from the test and consists of task coverage, item difficulty, objective of item, and test format. The task coverage is based on the English curriculum for SMA and the objective of item refers to Bloom's taxonomy, which consists of knowledge, comprehension, application, analysis, synthesis, and evaluation. In this regard, the focus of this study was to find out whether the English listening test of NLE for SMA in Samarinda was appropriately constructed using basic standard.

3. FINDINGS

3.1 Quality of Item Contents of the English Listening Test of NLE

The quality of content validity of the English listening test or the quality of item contents was showed by table 4 below. The table consists of the source and the task coverage of the English listening test items refers to the English curriculum for SMA.

(1) The source of the English listening test items

The English listening test items were dominated by the items of first and second year of SMA (8 items were from the first year of SMA, 8 items were from the second year of SMA, and 4 items were from the third year of SMA).

(2) Task coverage of the English listening test items

From the task coverage, it shows that the English listening test's coverage was covered by 5 items of conversational expressions, 6 items of implied meaning, 8 items of specific content questions, and 1 items of writing comprehension. This task coverage was determined by coverage of each item supported by the instructional objectives in the English curriculum for SMA. The items of conversational expressions measure the students' comprehension in expression used in English

conversation, the items of implied meaning measure the students thinking toward the text meaning, the items of specific content questions measure the students' ability in comprehension about a text or a conversation, and the item of writing comprehension measures the students' comprehension in writing, specifically to test the students' ability in structure.

Table 4 The Task Coverage of the English Listening Test

No	The Source of the Items	Task Coverage of the Items	Item numbers
1.	First year of SMA	<ul style="list-style-type: none">• Conversational expressions• Specific content questions• Implied meaning (not directly stated)	1, 4, 5 8 11, 12, 19, 20
2.	Second year of SMA	<ul style="list-style-type: none">• Writing comprehension• Conversational expressions• Specific content questions	3 7 9, 10, 13, 14, 15, 16
3.	Third year of SMA	<ul style="list-style-type: none">• Conversational expressions• Implied meaning (not directly stated)• Specific content questions	2 6 17, 18
Total			20 items

3.2 Table of Specification Reconstructed of English Listening Test

The table of specification of the NLE test consists of the items of the English listening test with its answer, the task coverage, the item difficulty, the objective of item, and the source of each item. The items were taken from the English listening test of NLE for SMA in Samarinda; item difficulty consists of three levels: (1) easy, (2) moderate, and (3) difficult; the objective of item depends on the actions verbs used in each item; and the source of item is based on the English curriculum for SMA. The explanation of each component is described below:

1) *Item difficulty of the test*

The item difficulty of the English listening test was dominated by easy items (9 items are easy items, 7 items are moderate items, and 4 items are difficult items). This proportion happens because of the items almost measure the basic students' comprehension in English

listening. Table 5 presented to summarize the item difficulty of the test. The table consists of (1) level of difficulty; easy, moderate, and difficult, and (2) item numbers for each level.

Table 5 Item Difficulty of the Test

No	Item Difficulty	Item Numbers	Total
1.	Easy	1, 2, 4, 8, 9, 10, 16, 18, 20	9 items
2.	Moderate	5, 6, 7, 11, 13, 14, 19	7 items
3.	Difficult	3, 12, 15, 17	4 items
Total			20 items

2) *Task coverage of the test items*

The task coverage of the test items was dominated by specific content questions that usually appear in the English listening test (8 items are specific content questions, 6 items are implied meaning, 5 items are conversational expressions, and 1 item is writing comprehension). The table below, table 6 will make it clear. The table consists of task coverage of items and the item numbers for the task coverage.

Table 6. Task Coverage of the Test

No	Task Coverage	Item Numbers	Total
1.	Conversational expressions	1, 2, 4, 5, 7	5 items
2.	Implied meaning	6, 11, 12, 18, 19, 20	6 items
3.	Specific content questions	8, 9, 10, 13, 14, 15, 16,17	8 items
4.	Writing comprehension	3	1 item
Total			20 items

3) *The objective of the test items*

The objective of item was dominated by items for knowledge (9 items are for knowledge, 5 items are for comprehension, 4 items are for application, and 2 items are for analysis). As we know, knowledge is the lowest level in Bloom's taxonomy. The table 7 below showed the analysis on the objective of items. It consists of objective of item and item numbers for each objective.

Table 7. The Objective of the Test Items

No	Objective of Item	Item Numbers	Total
1.	Knowledge	3, 6, 9, 10, 11, 12, 13, 14, 16	9 items
2.	Comprehension	1, 7, 18, 19, 20	5 items
3.	Application	4, 5, 8, 17	4 items
4.	Synthesis	-	-
5.	Analysis	2, 15	2 items
6.	Evaluation	-	-

Total		20 items

4) *The source of the test items*

The source of the test item was dominated from the first and second year of SMA (8 items were from first year, 8 items were from second year, and 4 items were from third year of SMA). Table 8 below summarized the result of the analysis on the source of item. The table consists of the source of item and the item number for each source.

Table 8. The Source of the Test Items

No	The source of item	Item Numbers	Total
1.	First year	1, 4, 5, 8, 11, 12, 19, 20	8 items
2.	Second year	3, 7, 9, 13, 14, 15, 16	8 items
3.	Third year	2, 6, 17, 18	4 items
Total			20 items

The conclusion of analysis showed in the tables above figured out that the Table of Specification was in unbalanced proportion. It should in balance proportion. It would be better if the items were from first, second, and third year in balanced proportion.

4. DISCUSSION

Based on the findings of the study, a discussion is focused to give comment of the result of analysis on the content validity of the English listening test of NLE for SMA in Samarinda. The result of analysis in this study to prove the content validity is based on the Table of Specification reconstructed from the English listening test. The analysis involved the task coverage of the test, item difficulty, the objective of item, the source of item, and the test format.

The analysis showed that the English listening test under study has low validity. Finding on coverage of contents English listening test that presented the task coverage, item difficulty, and the objective of item showed that the English listening test under study has low content validity. Moreover, the table of Specification reconstructed from the English listening test figured out that the proportion of each component in the Table of Specification was not balanced in proportion. It means that the English listening test of NLE for SMA in Samarinda has low content validity. From this result, it can be known that it is not easy to make a test that has high content validity. This is supported by previous research about the validity of the English EBTANAS test. Riandy (1997) analyzed Item Analysis of English EBTANAS Test 1995/1996 for SMP in Samarinda using descriptive study. His research revealed that the items of English EBTANAS test in academic year 1995/1996 for SMP in Samarinda were not valid.

5. CONCLUSION

Based on the objectives and findings of the study, some conclusions are presented as follows:

- 1) The analysis on content validity revealed three major findings on task coverage, item difficulty, and objective of item. Each finding is summarized below:
 - a) Task coverage of the English listening test is dominated by the first and the second year of SMA (8 items are from first year, 8 items are from second year, and 4 items are from third year of SMA).
 - a) Item difficulty of the English listening test almost consists of easy items (9 items of 20 items are easy, 7 items are moderate, and 4 items are difficult).
 - b) The English listening test items are dominated by known test (9 items of 20 items are for knowledge) and several items are for other objectives (5 items are for comprehension, 4 items are for application, and 2 items are for analysis)
- 2) The analysis on the Table of Specification reconstructed from the English listening test of NLE for SMA in Samarinda showed that the test has low content validity.

6. SUGGESTIONS

In order to improve the result of research in line the scope and limitation of the study as well as the significance of the study, the following suggestions are proposed:

- 1) The content validity of the English listening test of NLE has low content validity. It is suggested test-makers increase their understanding on constructing the standardized tests, specifically on the components of Table of Specification (i.e.: task coverage, item difficulty, objective of item, and the test format).
- 2) The analysis showed that the Table of Specification reconstructed from the English listening test under study have no good balance in objective of item and in the source of item. It is suggested the items for each objective is adjusted, such as example for a test that consists of 15 items: 6 items are for knowledge, 5 are for comprehension, 3 items are for application, and 1 item is for analysis; whereas for the source of item, it is suggested the items are from first, second, and third year of SMA with balanced proportion.

REFERENCES

- Arikunto, Suharsimi. 1996. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: PT. Bina Rupa Aksara.
- Budiharso, T. 2004. *Prinsip dan Strategi Pengajaran Bahasa*, Surabaya: Lutfansyah Mediatama.
- Darma, Erwin. 2004. *Item Quality of English Reading Test of UAN for SLTP*. Thesis. Samarinda: College of English Education of Mulawarman University.
- Diknas. 2003. *Kurikulum 2004 Standar Kompetensi Mata Pelajaran Bahasa Inggris Untuk SMA*. Jakarta: Departemen Pendidikan Nasional.
- Djiwandono, Soenardi. 1996. *Tes Bahasa dalam Pengajaran*. Bandung: Penerbit ITB.
- Hidriyani. 2004. *Developing Reading Test for the Sixth Year Students of Elementary School*. Thesis. Samarinda: College of English Education of Mulawarman University.
- Hughes, Arthur. 1991. *Testing For Language Teachers*. Great Britain: Bell & Brain Ltd.
- Madson, Harold. 1983. *Technique in Testing*. New York: Oxford University Press.

Citation: Yunhadi, Wuwuh. 2016. Quality of Listening Test Validity of
the National Leaving Exam for SMA. *JARES*, (2016), 1(1): 33-44.

- Melda, Kamilatun. 2003. *Listening Comprehension ability of the First Year Students of SLTPN 5 Samarinda in Academic Year 2002/2003*. Thesis. Samarinda: College of English Education of Mulawarman University.
- Nugroho, Dwi. 2001. *Posisi Kisi-kisi Tes dalam Rancangan Pembelajaran*. Samarinda: Sekretariat FKI FKIP UNMUL.
- Riandy, Firman Dwi. 1997. *Item Analysis of English EBTANAS Test 1995/1996 for SMP in Samarinda*. Samarinda: College of English Education of Mulawarman University.
- Saukah, Ali. 1997. *Evaluasi Hasil Pengajaran Bahasa Inggris*. Jakarta: Universitas Terbuka.
- Tuckman, Bruce. W. 1975. *Measuring Educational Outcomes: Fundamentals of Testing*. New York: Harcourt Brace Jovanovich Inc.

