An Analysis of EFL Teacher-Made Tests:
Content Knowledge, Cognitive, and Authentic Evidences

Rice Pesiwarissa[1]*,
Pattimura University, Indonesia
e-mail: ricepesiwarissa2014@gmail.com

Karolis Anaktototy[2],
Pattimura University, Indonesia
e-mail: wakwyoya@gmail.com

Hendrik Jacob Maruanaya[3]
Pattimura University, Indonesia
e-mail: hjmaruanaya@gmail.com

Abstract
Teacher-made tests became a common practice in school to assess the students' mastery of the content knowledge and skills. In the development of the test, multiple-choice test form is frequently a preferred option among the teacher. This study examines the content knowledge, cognitive and authentic evidence of teacher-made multiple-choice tests in EFL learning context in Junior High school. The English midterm-test document for grade 7 as the source of data. The test consists of 25 multiple choice items. The analysis showed that test items functioned primarily at the content knowledge of linguistic competence (52%), discourse competence (24%), and interactional competence (24%). The cognitive functioning level is C1 (4%), C2 (32%), C3 (32%) and C4 (32%), while the authentic functioning level is 16%. However, there are 84% items that are considered less authentic with the result of the analysis showing the mean score of the raters is 2.69. Regarding the findings of this study, the teacher needs to develop the authentic tasks in the test.

Keywords: teacher-made test, EFL learning, content knowledge, cognitive, authentic evidence

INTRODUCTION

The teacher plays an important role in teaching students to master English well. Therefore, teachers must have good competence in designing, implementing, and evaluating learning. At the end of the learning process, the teacher must measure how far students absorb the material that has been taught by giving a test.

The test is designed and used to examine or assess the learner's acquired knowledge and skills. According to Brown (2003), a test is a method of measuring a person's ability or knowledge in a given area. In other words, we can say that a test is a method of measurement from the materials that have been given. The teacher or some other institution may set and grade the tests. There are two types of tests based on the test-maker; they are standardized tests and teacher-made tests (Arikunto, 2005). A teacher-made test is a test design by the teacher to measure the students' acquired knowledge and skill. The teacher-made tests may also be employed as a tool for formative evaluation. The teacher creates the test to determine the student's achievement and competency in a certain area.

For that reason, teachers need to make a good quality test, so that the test can measure the students' achievements accurately.

BACKGROUND

To develop a good quality test, teachers must have a good competence in designing the test items. According to Brown (2003), there are five criteria to test the quality of the test namely: practicality, reliability, validity, authenticity, and washback.

However, it is still to be questioned whether the test made by the teacher has good quality or not because the teacher rarely tries out and analyzes the test first before giving it to the students. Mardapi in Widoyoko (2012) stated that there are six steps to develop a test: 1) Create a table of specifications, 2) Create stem item of the test, 3) Analyze the test item, 4) Do the tryout, 5) Analyze the item, 6) Revise. Knowing this fact, the teacher should do the tryout and analyze the test so that the teacher will know the quality of the test. By analyzing the test, the teacher will know which item can be used or revised. Furthermore, by analyzing the test, the teacher will obtain the information to determine the students' progress. Therefore, well-constructed tests can give students the opportunity of assessing their knowledge, and with immediate and constructive feedback; the learners can improve their performance.

In analyzing the test, validity and reliability are the two major criteria which strongly determine the quality of the test. However, authenticity is also an important criterion that has to be considered in analyzing the test. Citing Bachman and Palmer in Brown (2003, p.23) authenticity is "the degree of correspondence of the characteristics of a given language test task to the features of a target language task". Many test items fail to replicate the real-world task, according to Brown (2003), because they are contrived or unnatural in their attempt to target a grammatical form or lexical item.. Furthermore, when discussing authenticity, it is important to define the construct validity and content validity because the concept of authenticity as a well-operated construct is important to achieve sufficient content validity, which in turn helps to ensure that the language tests are accurate in the assessment of the communicative language skills and the level of cognitive required based on the curriculum. These criteria must be considered in analyzing the test so that good quality tests can be obtained.

Therefore, this study aims to examine the content knowledge, cognitive and authentic evidence of teacher-made multiple-choice tests in EFL learning context in Junior High school.

LITERATURE REVIEW

Content Knowledge

The content or language teaching area in EFL learning context can be reflecting into communicative competence. According to Celce-Murcia (2008), there are six communicative competences: sociocultural competence, discourse competence, linguistic competence, formulaic competence, interactional competence, and strategic competence.

Sociocultural competence reflects the perception of the speaker about how to behave and express messages appropriately within the overall social and cultural sense of communication, in line with the pragmatic factors relevant to language variation. Discourse competence involves the selection, sequence and arrangement of words, structures, sentences and expressions for the achievement of a unified spoken or spoken word. Content areas that distribute to discourse competence are: cohesion, deixis, coherence, and generic structure. Linguistic competence involves the essential elements of communication and include four types of knowledge: the morphological include parts of speech, grammatical inflection, productive derivational processes, the lexical knowledge of both content words (noun, verbs, adjective) and function words (pronouns, determiners, prepositions, and verbal auxiliaries), as well as th   ological for pronunciation and the syntactic include constituent/phrase structure, word or   sic sentence type, modification, coordination, subordinating and embedding. Formulaic competence involves the fixed and prefabricated parts of language that speakers use often in everyday interactions which include the fixed phrases and formulaic chunks (of course, how do you do, etc), collocations (verb object such as : spend money, adverb-adjective such as mutually intelligible, and adjective noun such as tall building), idioms, and lexical frames. Interactional competence involves at least three sub-components: actional competence: knowledge of how to perform common speech acts and speech act sets in the target language; conversational competence includes knowing how to initiate and end conversations, establish and change topics, and gain, hold, and yield the floor; and nonverbal/paralinguistic competence, which involves body language, non-verbal turn-taking signals, gestures, eye contact, batch channel behaviors, haptic behavior, proxemics, and non-linguistic utterances. Strategic competence refers to the mastery of communication strategies which include the communication strategies.

Cognitive Level

According to the updated version of Bloom's Taxonomy, there are six levels of cognitive learning (Anderson, 2001). Each level has an own conceptual framework; Remembering (C1) retrieving, recalling, or recognizing important information from long-term memory; Understanding (C2) convey understanding by one or more types of explanation; Applying (C3) apply knowledge or a skill in a new situation; Analyzing (C4) divide a substance into its basic pieces and identify how the parts relate to one another and/or to an overall structure or purpose; Evaluating (C5) make decisions on the basis of criteria and standards and Creating (C6) combine parts to create a new logical or effective entity or reorganize components to create a new pattern or structure.

Authenticity

Bachman and Palmer in Brown (2003) define authenticity as the degree of correspondence of the characteristics of a given language test task to the features of a target language task, and then

suggest an agenda for identifying those target language tasks and for transforming them into valid test items.

According to Brown (2003), in a test, authenticity may be present in the following ways: 1) the language in the test is as natural as possible; 2) items are contextualized rather than isolated; 3) topics are meaningful (relevant, interesting) for the learner; 4) some thematic organization to items is provided, such as through a story line or episode; and 5) task represent, or closely approximate, real-world task.

METHODOLOGY

The researcher employed descriptive quantitative research design. Ary et al (2010) states that quantitative research uses objective measurement to gather numeric data that used to answer questions or predetermined hypotheses.

This study was conducted in SMP Negeri 9 Ambon which was located on Jalan Wolter Monginsidi, Lateri-Ambon. The sample of this study was the test document since this study needed to analyze an English midterm test for grade VII, made by a teacher. The test document was English midterm test in the second semester of the academic year 2020/2021 for grade 7.

The procedure of data collection involved instrument development and validation before the research took place. There were two major techniques of data collection used in this study namely test administration and document analysis. The test document was analyzed to obtain the data of authenticity of the test and EFL content knowledge and cognitive level in the multiple choice test made by the teacher. In collecting the data, the researcher used documents and a checklist table as research instruments.

To analyze the content of the test, the researcher constructed an indicator /checklist by referring to the content and standard competency of EFL teaching and learning for SMP grade VII based on curriculum 2013 and the concept of content knowledge which categorized into communicative competence by Celce Murcia (2008). In terms of cognitive level, the analysis was based on Bloom's revised taxonomy (Anderson,2001). To analyze the authenticity of the test, the researcher constructed the authenticity checklist table which developed based on criteria set by Brown (2003). There were three raters to assess the authenticity of the test items. The first rater is an English Study Program lecturer. He is a senior lecturer who teaches English Language Testing. The second rater is an English supervisor of Dinas Pendidikan Kota Ambon and the third rater is the researcher.

Data about content knowledge and cognitive level of the test that have been collected were analyzed quantitatively. The researcher calculated the frequency and found out the percentage of the test item by using the percentage formula.

$\% = \frac{f}{N}x100$

Where: % = percent
f = frequency

Data of authenticity of the test that have been collected were analyzed by comparing the score from the three raters. To know the extent of the authenticity of the test and level of agreement of the raters, the researcher used Pearson Product Moment with the assistant of the

SPSS application to correlate the data. The size of the relationship among the raters' scores are expressed in numbers called the correlation coefficient (R) that is between -1 to +1 where if approaching +1 then the relationship is stronger and positive. Meanwhile, if close to -1 then there is a stronger relationship but the direction is negative. If the correlation coefficient is zero then it means there is no relationship at all among the raters' scores. Napitulu et al (2018) provide the interpretation of correlation coefficient in the table below:

Table 1. Interpretation of Correlation Coefficient

Coefficient Interval	Correlation
0.00 – 0.199	Very Weak
0.20 – 0.399	Weak
0.40 – 0.599	Medium
0.60 – 0.799	Strong
0.80 – 1.000	Very Strong

FINDINGS

The content of the test developed by the teacher of SMP Negeri 9 Ambon was analyzed based on the conceptual framework of communicative competence (Murcia,2008). There are six elements to represent language communicative competence. They are sociocultural competence, discourse competence, linguistic competence, formulaic competence, interactional competence, and strategic competence.

From the content of the test analysis, it was found that the test represents the discourse, linguistic and interactional competence, while sociocultural, formulaic and strategic competence were not included in the test items. The description of each competence tested in the midterm test is described below.

● *Discourse Competence*

Discourse competence relates to 'the selection, sequencing, and arrangement of words, structure, sentences, and utterances to achieve a unified spoken or written text' (Murcia, 2008). From the analysis of the test content, it was found that 7 questions were constructed to test this knowledge on students.

Table 2. The Result of Discourse Competence Analysis

Midterm Test	Element of Discourse Competence			
	Cohesion	**Deixis**	**Coherence**	**Generic structure**
Item no	-	-	8,17,18,19	9,16
Total =6	-	-	4 items	2 items

The data show that from the 25 multiple choice test items, six test items dealt with coherence and generic structure as a part of discourse competence. Knowledge on coherence accounts for 4 items while generic structure accounts only for 2 items.

To measure the percentage of the test items, the percentage formula is used. Therefore the percentage for the coherence is 16% and generic structure is 8%.

● *Linguistic Competence*

Linguistic competence relates to syntactic (sentence patterns), morphology, lexical, and phonology. From the analysis of the test content, it was found that 13 questions were constructed to test this knowledge on students.

Table 3. The Result of Linguistic Competence Analysis

Midterm Test	Element of Linguistic Competence			
	Syntactic	**Morphological**	**Lexical**	**Phonological**
Item no	1,2,3,4,10 11,12,13,14,15	-	22,23,24	-
Total = 13	10 items	-	3 items	-

The data show that 13 test items dealt with linguistic competence include syntactic knowledge which consist of 10 items (40%) and lexical knowledge account for 3 items (12%), while morphological, and phonological gain no place in the test.

● *Interactional Competence*

Interactional competence deals with the actional competence and conversational competence. The actional competence relates to knowledge of language function which covers interpersonal exchange, information, opinions, feelings, suasion, problems, and also future scenarios, while the conversational competence relates to knowledge of extending the conversation. Based on the content of the test, it was found that 6 questions were developed to assess students' knowledge of interactional competence.

Table 4. The Result of Interactional Competence Analysis

Midterm Test	Element of Interactional Competence	
	Actional Competence	Conversational Competence
Item no	5,6,7,20,21,25	-
Total = 6	6 item	-

The data show that 6 items relate to actional competence which cover asking and giving factual information and interpersonal exchange while conversational competence gains no place in the test. The percentage of the actional competence in the test is 24%.

The following table shows the whole competence elements tested in the midterm test constructed by the teacher.

Table 5. Content of The Test In Term of Communicative Competence Framework

No.	Content Knowledge	Total test Item	Percentage
1.	Sociocultural	-	-
1.	Discourse	6	24%

2.	Linguistic	13	52 %
3	Interactional	6	24 %
4.	Formulaic	-	-
5.	Strategic	-	-

Linguistic competence accounts for 13 items (52%) and it represents the majority of content knowledge followed by discourse competence 6 items (24%) and interactional competence 6 items (24%). Sociocultural competence, formulaic competence and strategic competence were not found in the content of the test.

The analysis of the content of the test items in terms of the cognitive level was based on Bloom's revised taxonomy (Anderson,2001) which cover the lowest order thinking skill to the highest order thinking skill. The result analysis of the cognitive level can be seen in the table below:

Table 6. The Result of Cognitive Level Found In The Test

Cognitive Level	Detail of Basic Competence	Item No	Total	Percentage
C1	Remembering the structure of the descriptive text	9	1	4%
C2	Understanding the structure of the descriptive text by giving information related to the description of people.	5,7	2	32%
	Understanding the language feature of descriptive text by giving information related to animal descriptions	11	1	
	Understanding the social function of descriptive text by giving information related to the description of objects.	16	1	
	Understanding the structure of descriptive text by giving information related to the description of objects	17	1	
	Identifying the language feature of transactional texts that involve the act of giving and asking for information related to the character of objects.	23	1	
	Identifying the language feature of transactional texts that involve the act of giving and asking for information related to the character of people.	24,25	2	
C3	Applying language feature of descriptive text by giving information related to people's descriptions	1,2	2	32 %
	Applying descriptive text structure by giving information related to the description of	3,4,10	3	

	people.			
	Applying the language feature of descriptive text by giving information related to animal descriptions.	13,14	2	
	Applying the structure of the descriptive text by giving information related to animal descriptions.	19	1	
C4	Analyzing the structure of the descriptive text by giving information related to the description of people.	6,8	2	32 %
	Analyzing the structure of transactional texts involving the act of giving and asking for information related to the character of people.	21	1	
	Analyzing the structure of transactional texts involving the act of giving and asking for information related to the character of objects	22	1	
	Comparing the language feature of descriptive texts by giving information related to animal descriptions.	12,15	2	
	Comparing the structure of the descriptive text by giving information related to the description of objects	18	1	
	Comparing the structure of descriptive texts by giving information related to people's descriptions.	20	1	

In terms of the intended cognitive level present in the developed test, it was found that the cognitive level 2 (C2), level 3 (C3) and level 4 (C4) were dominant. Each is equal in number which is 8 items (32%) respectively, while the cognitive level 1 (C1) accounts for only one item (4 %).

The authenticity of the test is measured based on Brown's theory of authenticity in language tests that contain: natural language use, contextualized rather than isolated, relevancy, thematic organization, and real-world like tasks. Likert scale of 3 to 1, with 3 indicating 'agree', 2 indicating 'slightly agree' and 1 indicating 'disagree ', was used to assess the authenticity of the developed test.

There were three raters including the researcher to assess the authenticity of the test. The first rater is a senior lecturer from the English Study Program who teaches English Language Testing. The second rater is an English supervisor of Dinas Pendidikan Kota Ambon and the third rater is the researcher.

The five authenticity criteria in the table are coded: 1) natural language use, 2) contextualization, 3) relevancy, 4) thematic organization and 5) real word task. The result of the analysis of the authenticity can be seen in the table below:

Table 7. The Result of Authenticity Analysis 1

Item	Rater 1						Rater 2						Rater 3					
No	1	2	3	4	5	Average	1	2	3	4	5	Average	1	2	3	4	5	Average
1	3	3	3	1	3	2,6	3	3	3	1	3	2,6	2	1	1	1	1	1,2
2	3	3	3	1	3	2,6	3	3	3	1	3	2,6	2	1	1	1	1	1,2
3	3	3	3	1	3	2,6	3	3	3	1	3	2,6	2	1	1	1	1	1,2
4	3	3	3	1	3	2,6	3	3	3	1	3	2,6	2	1	1	1	1	1,2
5	3	3	3	3	3	3,0	3	3	3	3	3	3,0	2	3	3	3	1	2,4
6	3	3	3	1	3	2,6	3	3	3	1	3	2,6	2	2	3	2	1	2,0
7	3	3	3	1	3	2,6	3	3	3	1	3	2,6	2	3	3	2	1	2,2
8	3	3	3	3	3	3,0	3	3	3	3	3	3,0	2	3	3	2	1	2,2
9	3	3	3	3	3	3,0	3	3	3	3	2	2,8	3	1	3	1	1	1,8
10	3	3	3	3	3	3,0	3	3	3	3	3	3,0	2	3	3	3	3	2,8
11	3	3	3	3	3	3,0	3	3	3	3	3	3,0	2	3	3	3	1	2,4
12	3	3	3	3	3	3,0	3	3	3	3	3	3,0	2	3	3	3	1	2,4
13	3	3	3	3	3	3,0	3	3	3	3	3	3,0	2	3	3	3	1	2,4
14	3	3	3	3	3	3,0	3	3	3	3	3	3,0	2	3	3	3	1	2,4
15	3	3	3	3	3	3,0	3	3	3	3	3	3,0	2	3	3	3	1	2,4
16	3	3	3	3	3	3,0	3	3	3	3	3	3,0	2	3	3	3	1	2,4
17	3	3	3	3	3	3,0	3	3	3	3	3	3,0	2	3	3	3	1	2,4
18	3	3	3	3	3	3,0	3	3	3	3	3	3,0	2	3	3	3	1	2,4
19	3	3	3	3	3	3,0	3	3	3	3	3	3,0	3	2	3	3	3	2,8
20	3	3	3	3	3	3,0	3	3	3	3	3	3,0	3	3	3	3	3	3,0
21	3	3	3	3	3	3,0	3	3	3	3	3	3,0	3	3	3	3	3	3,0
22	3	3	3	3	3	3,0	3	3	3	3	3	3,0	3	3	3	3	3	3,0
23	3	3	3	3	3	3,0	3	3	3	3	3	3,0	3	3	3	3	3	3,0
24	3	3	3	3	3	3,0	3	3	3	3	3	3,0	2	3	3	3	3	2,8
25	3	3	3	3	3	3,0	3	3	3	3	3	3,0	2	3	3	3	3	2,8

In terms of the authenticity of the developed test item, the result of the analysis shows that the three raters provided different scores related to all the authenticity criteria in the test items. The data show that the majority of the test items can fulfill the relevancy criteria which account for 21 items (84%), 18 items meet the contextualized criteria (72%) and 16 items can fulfill the thematic organization criteria (64%). On the other hand, there are only 8 items that meet the criteria of representing the real word task (32%), while the criteria of natural language use only account for 6 items (24%).

Overall, the analysis of the authenticity can be seen in the table below:

Table 8. The Result of Authenticity Analysis 2

Item	Score			Average
number	Rater1	Rater 2	Rater 3	
1	2,6	2,6	1,2	2,1
2	2,6	2,6	1,2	2,1
3	2,6	2,6	1,2	2,1
4	2,6	2,6	1,2	2,1

5	3,0	3,0	2,4	2,8
6	2,6	2,6	2,0	2,4
7	2,6	2,6	2,2	2,5
8	3,0	3,0	2,2	2,7
9	3,0	2,8	1,8	2,5
10	3,0	3,0	2,8	2,9
11	3,0	3,0	2,4	2,8
12	3,0	3,0	2,4	2,8
13	3,0	3,0	2,4	2,8
14	3,0	3,0	2,4	2,8
15	3,0	3,0	2,4	2,8
16	3,0	3,0	2,4	2,8
17	3,0	3,0	2,4	2,8
18	3,0	3,0	2,4	2,8
19	3,0	3,0	2,8	2,9
20	3,0	3,0	3,0	3,0
21	3,0	3,0	3,0	3,0
22	3,0	3,0	3,0	3,0
23	3,0	3,0	3,0	3,0
24	3,0	3,0	2,8	2,9
25	3,0	3,0	2,8	2,9

The existing data show that only 4 items are considered to meet the authenticity criteria which are shown through the natural language, contextualized, meaningful topics, thematic organization and represent, or closely approximate to the real-world task. The items assessed meet the five criteria are items number 20, 21, 22, and 23. Therefore, those 4 items (16%) can be declared authentic.

On the other hand, there are 21 items (84%) that are considered less authentic by the three raters. Although they differ in the results of the analysis, they do not show a significant difference in score as shown in the table below.

Table 9. Descriptive Statistic

Range	Min.	Max.	Mean	SD	N=
.90	2.10	3.00	2.69	.30	25

The minimum score given by the raters was 2.10, but the maximum score was 3.00 and the mean was 2.69. Furthermore, the results of the analysis of the three raters were then correlated using SPSS application to measure the extent of the level of agreement of the three raters to the authenticity of the test. The correlation results are shown in the following table:

Table 10. The Results of Authenticity Correlation Analysis

Correlations

		Rater_1	Rater_2	Rater_3
Rater_1	Pearson Correlation	1	.974[**]	.794[**]
	Sig. (2-tailed)		.000	.000

		25	25	25
Rater_2	Pearson Correlation	.974**	1	.836**
	Sig. (2-tailed)	.000		.000
	N	25	25	25
Rater_3	Pearson Correlation	.794**	.836**	1
	Sig. (2-tailed)	.000	.000	
	N	25	25	25

**. Correlation is significant at the 0.01 level (2-tailed).

The coefficient correlation between rater 1 and rater 2 is 0.974, therefore we can say that the level of the correlation is very strong. The coefficient correlation between rater 1 and rater 3 is 0.794, so we can say that the level of the correlation is strong. The coefficient correlation between rater 2 and rater 3 is 0.836, it is categorized as very strong.

DISCUSSION

Content knowledge and language skills are two elements that can not be separated from language learning. These two elements support one another to be a proficient language user. Content knowledge such as discourse, linguistic and interactional knowledge, to a certain degree, play an important role in reading, speaking, and writing. From the analysis, it is evidence that content knowledge such as discourse, linguistic and interactional knowledge are integrated in speaking, reading and writing.

In discourse competence, the students closely engage with how to construct the written or spoken text. On one hand, they need to have the relevant schemata in terms of selecting, sequencing, and also arranging the words, phrases, clauses, sentences, and also utterances in order to create unified information to convey in the context of communication. Discourse knowledge which deal with coherence integrated into students' writing, speaking and reading skills. Cohesion and coherence are also essential for students to interpret the text well. It is in line with a study conducted by Pan (2021) which found that in many international tests such as IELTS and TOEFL, coherence is an important marking criterion for writing (written or spoken text) therefore students are required to write English compositions coherently.

Linguistic knowledge is also related to students' writing skills as stated in Menggo et al (2019) that English writing skill encourages students to employ their understanding of micro linguistics, i.e., morphology, syntax, lexicon and semantics that have already been learned in English class.

Whereas, actional competence is integrated in speaking and writing skill since language is an effective means of expressing ideas and feelings, asking and giving information both in spoken and written form to communicate or interact with other people (Anggraini,2012). Moreover, it gives students a way to communicate and help them socialize in society.

In terms of students' level cognitive ability, the multiple choice items developed by the teacher can measure students' cognitive level include level C1, C2, C3, and C4, while there is

no item to measure students' higher order thinking skills C5 and C6. It is difficult for the teacher to assess the students' high order thinking skills which cover students' creativity. Carneson et al (2016) argued that creativity cannot easily be tested by using multiple choice questions. Discursive questions, such as the "Essay-type" question, are ideal for testing creativity. However, Scully (2017) argued that multiple choice items have the capacity to assess higher-order thinking skills. Therefore, teachers need to learn the strategies for constructing multiple choice items to assess students' higher-order thinking.

Based on the result analysis of the authenticity of the test, this research found that the four multiple choices which are considered authentic meet five criteria as proposed by Brown. The researcher recognized that most of the test tasks had problems fulfilling the naturalness of language used in the test instructions, stems and the optional answers. Despite the fact that the language test was not designed to measure specific grammatical or lexical issues, the teacher should minimize linguistic errors in order to provide a highly authentic reading test. In order to minimize test takers' difficulty in comprehending the test instructions, there should be no linguistic faults in the test tasks, such as typographical errors, lexis, word ordering, grammar (syntactic concerns), and diction. It is also found that most test texts face problems to fulfill the naturalness of language used in the test passages and the real-world representativeness. Even though the topics of the passages were reasonable and based on real world context, almost all of the passages included in the reading exam failed to portray the real world context. Authentic reading passages are derived from real-world sources, however the teacher who developed the English test items did not mention the sources where the passages were taken from.

CONCLUSION AND SUGGESTION

Based on the results of the research, the researcher concludes that the content knowledge which cover discourse competence, linguistic competence and interactional competence are three main content knowledge tested for students of grade seven in junior high school since it is in accordance to the curriculum of junior high school. In terms of measuring the students' cognitive level, the test items only measure the lower level to the middle level thinking skill, while the higher order thinking skill is not found in the test item. Therefore, teachers need to learn the strategies for constructing multiple choice items to assess students' higher-order thinking which cover analytical, critical or creative thinking. Concerning to the authenticity of the test, most of the problem appear in the test items are the naturalness of language used in the test passages and the real-world representativeness. Therefore, teachers need to avoid typographical mistakes, some lexical problems, and the unknown sources of reading passages in order to avoid test takers' confusion in understanding the test tasks.

REFERENCES

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Complete Edition*. New York: Longman

Anggraini, (2012). Improving Students' Speaking Skill Through CLT An Action Research. *Wanastra* Vol X No.1,17-23.

Arikunto, S. (2005). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.

Ary, Donald, et al. (2010). *Introduction to Research in Education (8th ed.)*. Canada: Nelson Education, Ltd.

Brown, H.D. (2003). *Language assessment: Principles and classroom practices*. New York: Longman.

Carneson, John & Delpierre, Georges & Masters, Ken. (2016). Designing and Managing Multiple Choice Questions (2nd Ed). 10.13140/RG.2.2.22028.31369.

Celce-Murcia M. (2008) Rethinking the Role of Communicative Competence in Language Teaching. In: Soler E.A., Jordà M.S. (eds) *Intercultural Language Use and Language Learning*. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-5639-0_3

Menggo, S., Suastra, I. M., Budiarsa, M., & Padmadewi, N. N. (2019). Speaking for academic purposes course: An analysis of language functions. *E-Journal of Linguistics*, 13(2), 314–332. https://doi.org/10.24843/e-jl.2019.v13.i02.p10

Napitupulu, Darmawan, et al. (2018). Analysis of Student Satisfaction Toward Quality of Service Facility. *Journal of Physics Conference Series*. 954. 10.1088/1742-6596/954/1/012019.

Pan, Hui. (2021). On Enhancing Students' Discourse Competence in Reading. *Journal of Higher Education Research*. 2. 10.32629/jher.v2i1.252.

Widoyoko, E., P. (2012). *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta: Pustaka Pelajar.