EDUVELOP
Journal of English Education and Development
Nationally Accredited Journal Decree No. B/1410/E5/E5.2.1/2019

Volume 6, No. 1, September 2022
DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

ISSN 2597-713X (print)
ISSN 2597-7148 (online)

DEVELOPING HIGHER ORDER THINKING SKILLS TEST INSTRUMENTS FOR ENGLISH SPEAKING CLASSES: SING A SONG MATERIAL

Multazam Abubakar[1], Nurfatwa Putri Husain [1], Mardiana[1]
[1]Universitas Islam Negeri Alauddin Makassar
[2]Email: multazam.abubakar@uin-alauddin.ac.id

Abstract: This study aims at creating HOTS items for the "Sing a Song" material, establishing their level of validity, and determining their dependability for tenth-grade students of MA Madani Alauddin Pao-Pao. The design of this research is research and development (R&D). The improvement version this is used refers back to the Plomp version which includes five levels of development, namely Preliminary investigation, Design, Realization/Construction, Test, Evaluation and Revision (Test, Evaluation and Revision), and Implementation (Implementation). The subjects in this study are students at class X MIA 2 in MA Madani Alauddin Pao-Pao. The instruments used in data collection were validation sheets and HOTS items. Based on the findings of the study, it obtains that the extent of validity of the HOTS gadgets changed into 3.7 that is withinside the very legitimate category, the reliability of the HOTS gadgets is 0.81 suggests in superb category, the issue degree of the HOTS gadget**s** is 0.60 suggests withinside the medium category. The implications of developing HOTS items can train and encourage students to do high-level reasoning so that they are not fixated on the pattern of answers that result from memorization without understanding the concept of knowledge, and can make students think critically while working on the test in the "Sing a Song "Material so that answers can be produced based on in-depth reasoning.

Keywords: Higher Order Thinking Skill Test, Instrument Test, sing a song, Speaking skills

1. Introduction

The demand of educational system in twenty-first century is to form young generations to face the challenges of world(Sakka and Naim 2019). The students are expected to have new experiences, unique and creative ideas as well as develop collaborative attitude (Hamdi, Suganda, and Hayati 2018). Warisdiono (2017) contends that today's learning system must focus in developing the 4C's as competencies in preparing the students to face the world's challenges. The 4C are creativity, critical thinking, communication and collaboration. The four elements must be integrated in school curriculum so they can be achieved in formal way. Craig (2011) suggests that in the 21st century, students must be exposed with unusual problems, so they must be familiar with activities related to problems solving.

Today the curriculum that is implemented by the Indonesian government is the 2013 curriculum

DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

(K13). The principle of this curriculum is to prepare human resources to have the ability to live as individuals and citizens who are faithful, productive, creative, innovative, and effective to produce Indonesian people through the formation and strengthening of aspects of attitude, skills and knowledge that are integrated. Students are also expected to contribute to the life of society, nation, state and world civilization(Nazaruddin 2017)

Widana (Widana 2017) stated that the result of the research by Program for International Students Assessment (PISA) in 2015 revealed that Indonesia is in the bottom 4 of the 43 participated countries in the world. Indonesia has low rank in reading, mathematic, and scientific literacy. It is shown by the answers of the test at that research. It implies that the objective of the education in Indonesia is not relevant with the objective of 2013 Curriculum. The Government (2017) also mentioned that the student's ability is still lack in understanding complex information, theory and solving problem. So that the government makes a revision of the Curriculum.

The revised curriculum by Ministry of Education and Culture of Indonesia which known as 2013 Curriculum was revised with some completing. The completing such as the implementation on the content of the standard is minimizing the materials that are not relevant with the development of the material for student's need to thinking critically, the change of context standard, and the assessment standard, also the assessment standard in this revision is assessing the student's result of learning process that helps the student to increase the higher order thinking skill (HOTS) because higher thinking can encourage the student to think critically and more deeply about the learning (Widana 2018).

Higher order thinking skills (HOTS) are the ability to think which is activated when the students encounter unfamiliar problems, dilemmas over the answer of a question which make them not only memorize the solutions but also think critically and creatively in solving the problems (King, F. J., Goodson, L., and Rohani 2010). Ngah et al. (Ngah et al. 2017) defined HOTS as the ability to think in high level that involve complex thought process and thinking process towards higher level, not merely memorizing and repeating learning. Higher order thinking skills are important because they can make the students become achievers and give contributions to the society (Conklink 2012). They can also make the students develop their careers, have more achievement, have social skills, self-control, be creative, responsible work hard and be able to solve problems by making the right decision (Wang and Wang 2014). To develop the ability of critical thinking, there are five lesson can be taken, they are (1) determine the learning objectives, (2) teach through inquiry, (3) practice, (4) review, refine and improve understanding, and (5) practice feedback and assess learning (Limbach and Waugh 2010).

The purpose of education will be successful achieved if the implementation is released that based on government's certainty. The quality of education is reflected in the national education standard in the country (Yonisa, 1, and Lestari 2019). It is following the assessment system for the students based on the government's rules. The assessment research of education consists of the result of study

DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

by the teacher, the result of the study by the unit of educator and the result of study by the government. The result of the learner's assessment that it is seen by the teacher by using some technique such as test, observation, individual task or group, and the other form which appropriates with the characteristic of the competency and the stage of the learner's development (Salamah 2018). Assessment for learning is an assessment which that is designed to provide information for teachers to modify the learning activities, differentiate, and understand the way students approach learning (Earl 2006). Thus, assessment is a valuable instrument to make teaching decision (Welle 2007).

A test as an instrument of learning outcomes measures student skills at various levels, ranging from low levels of thinking to high levels of thinking. Therefore, it is necessary to pay attention to each proportion of the level of thinking ability in each question problem that will affect student learning patterns. The fact, in making instruments questions tend to be dominated by questions with low-level thinking skills so students prefer to learn by memorizing and working on the questions contained in the book rather than developing their ability at a higher level of thinking when trying to solve a problem.

Problems that occur in schools, questions tend to be more numerous testing aspects of memory that lack training in higher order thinking skills students, the ability to think of Indonesian children scientifically is still considered low seen one of the factors causing it, among others because students in Indonesia are less trained in solving problems that are measure HOTS, and the problem faced by the teacher is the ability of the teacher in developing HOTS assessment instruments are still lacking and not yet the availability of assessment instruments specifically designed to train HOTS, so that HOTS assessment instruments need to be developed. Development of thinking skills a high level of students will result in: students' proficiency in problem solving strategies become good, the level of confidence of students in mathematics increases, and students' learning achievement on non-routine problems which requires higher-order thinking skills to increase.

Based on the explanation presented above, it is clear that the students need to increase their thinking skills by giving the instrument of test based on the HOTS concept in 2013 curriculum. Thus, it is important to conduct the research about designing the instrument test based on the English material which is applied HOTS concept. This study aims at creating HOTS items for the "Sing a Song" material, establishing their level of validity, and determining their dependability for tenth-grade students of MA Madani Alauddin Pao-Pao.

2. Method

The design of this research is Research and Development (R&D). R&D is defined as a research design that involves the classroom problems, studying recent theories of educational product development, developing the educational products, validating the product to experts, and field testing the product. It consists of a number of steps commonly known as the "R&D cycle", which involves studying the research findings related to the product to be developed,

DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

developing the product based on the findings, conducting field tests in the environment where it will eventually be used, and making revisions that can be corrected found during the field test phase (Borg and Gall 2003). In this research, the Tjeerd Plomp model was used. The Plomp model is shown to be more adaptable and versatile than the Four-D model since it includes development activities at each step that may be tailored to the peculiarities of the research. Preliminary research, for example, may be included in the preliminary investigation. This preliminary research can also take the form of preliminary research whose findings serve as the foundation for future development; in this case, the research design is a research and development design (L. Azizah, S. Mariani 2012). This model has five stages, they are (1) preliminary investigation, (2) design, (3) realization / construction (realization / construction), (4) test, evaluation, and revision (test, evaluation, and revision), and (5) implementation (Implementation). The Plomp development model was chosen because it is more versatile, as each step involves bespoke development activities based on the research's characteristics.

This research took place at MA Madani Alauddin Pao-pao, South Sulawesi. It involved several participants. They are selected based on Purposive sampling in order to conduct data collection Therefore, this research employs multiple data sources by inviting learners and teachers of MA Madani Alauddin Pao-pao and expert. To be more detail, the research subjects are elaborated as follows:

1. Students (Target Group): 20 students became the subject or the people whom this development is addressed to.
2. Teachers (Audience Group): 2 teachers.
3. Experts: 2 experts be involved in this research due to make sure the appropriateness of the materials.

The instruments of this research were as follows:

1. File

   During the research process the researchers collected qualitative documents. Specifically, the literature in this study is to train teachers to train students' higher-order thinking ability. In the assessment of high-level thinking skills, it is necessary to understand the forms and aspects of thinking, which are mainly used by teachers to diversify students. A document analysis guide is needed to help researchers analyze document data.

2. Interview

   This research used semi-structured interviews. In the process, the researchers prepared and asked questions based on the research questions. The interview aimed at investigating how if the English teachers use higher-order thinking skills assessments in their test for their students. The interview guide is required to help the researcher conduct the interview.

3. Test

   The research process is research and development, the researchers need test questions used by teachers so that they can be developed into types of test questions that fulfill the HOTS

DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

standards. The purpose of knowing the types of previous questions also helps researchers to analyze the form of previous test so that they can be developed as much as possible. This also helps researchers to see the media and types of questions used by teachers so that in collecting data, researchers will find it easier to collect.

4. Validation Checklist

   Validation checklist was used as a benchmark and given to both of expert in evaluating the product. This validation checklist used to measure the level of validity of each items.

3. Findings and Discussion

A. Findings

The Existing of Test Instrument

Documents, interviews, and teacher's test items are the components to be analyzed at this section. In this case, the document consists of the Syllabus, Lesson Plan, and teaching materials.

1. Document analysis
   a. Syllabus

      Language learning cannot be separated from the requirements of learning such as how to plan, how to design, and how to obtain learning resources. All of the above are closely related to the syllabus. The syllabus contains a framework or details of how teachers and students should act in the classroom. The syllabus also becomes a guide or instructions for organizing activities in the classroom.

      The available syllabus was in accordance with the provisions of the 2013 curriculum because it already consists of several components that must exist in the development of the syllabus, including: competency standards, basic competencies, learning materials, indicators, assessments, time allocation, and learning resources developed by each educational unit.

   b. Lesson Plan

      Lesson plan is mandatory for every subject teacher in every education unit, including the X grade English teacher at MAN Madani Pao-Pao. RPP is described from the syllabus to direct student learning activities in an effort to achieve basic competence (KD). Every teacher in the education unit is obliged to prepare a complete and systematic lesson plan so that learning takes place interactively, inspiring, fun, challenging, motivating students to participate actively, and providing sufficient space for initiative, creativity, and independence according to talent, interests, and the physical and psychological development of students.

   c. Textbook

      Textbook is the primary source that contain learning materials for achieving basic and core competencies. Competence is intended to be related to increasing faith, piety, noble character and personality, mastery of science and technology, increasing sensitivity and aesthetic ability, increasing kinesthetic ability, and improving health, as measured by national education standards. The National Education Standards Agency (BNSP) has

DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

established that an excellent book is one that meets four criteria for eligibility: material feasibility, linguistic feasibility, presentation feasibility, and graphic feasibility.

2. Interview analysis

The interview technique was used by determining the representative sources in line with the research questions. The responders in this study were English teachers to find out the teacher's comprehension of HOTS while developing questions. The results of the interview is as follows:

"Both of them., Essays are easy to measure students' competency; while, the multiple choice is determined by the school"

Next, it was investigated about whether or not to include opening texts, pictures, and scenarios while creating questions. The teacher responded:

"There are introductory texts, visuals. There are a lot because it's like matching the picture with the description of the picture. What is the description of a picture, describe something, and scenarios also exist"

In response to further question about the aim of implementing these test items, she stated:

"That definitely measures whether the students understand what they been learning all this time or not"

3. Teacher's Test Items Analysis

The following test was given by the teacher to students through Google form along with the analysis from researchers which will then be developed by researcher. The researcher used the cognitive level theory to assess the questions posed by the teacher. Anderson & Krathwohl (Anderson and Krathwohl 2001) categorize thought process dimensions as follows:

Table 1.
Thought Process Dimension

LOTS (Lower Order Thinking Skill)	C1 (Remembering)	• Recalling. • Verbs: remember, register, repeat, imitate.
MOTS (Middle Order Thinking Skill)	C2 (Understanding)	• Explain ideas/concepts. • Verbs: explain, classify, accept, report.
	C3 (Applying)	• Using information on a different domain • Verbs: use, demonstrate, illustrate, operate
HOTS (Higher Order Thinking Skill)	C4 (Analyzing)	• Specify aspects/elements • Verbs: compare, examine, , criticize, test.
	C5 (Evaluating)	• Take your own decisions • Verbs: evaluate, judge, refute, decide, choose, support
	C6 (Creating)	• Create your own ideas/ideas • Verbs: construct, design, create, develop, write, formulate

DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

B. *Developing of HOTS Test Instrument*

Plump development model is used in the construction of the test item. This model has five stages of development: (1) preliminary study, (2) design, (3) realization/construction, (4) test, evaluation, and revision, and (5) implementation (Implementation). This study focuses on generating HOTS questions includes a grid of questions, questions, and grading standards (rubrics). The development of items pertains to the validity and dependability requirements.

1. *Preliminary Investigation Phase*

   The main activity at this stage is analysis the documents for the development of new product and analysis the feasibility and requirements of developing HOTS test items. Therefore, the researchers conducted preliminary observations based on document analysis and interviews analysis with the 10th grade English teachersof MA Madani Alauddin Pao-Pao.

2. *Design Phase*

   a. Collecting

      To get to this point, various papers must be prepared, including the syllabus, lesson plans, teaching materials, and teacher-created questions. The researcher evaluated these materials as part of the process of creating the questions into HOTS questions.

   b. Organizing

      Several findings of the study of these documents were derived from the data collection acquired. The syllabus is utilized as a guideline while creating a question grid, which is also used to produce the questions to be created. In addition to the syllabus, teaching materials serve as a guideline when developing questions. This teaching material is divided into ten chapters, with the questions provided by the researcher focusing on the chapter "Why Were They Famous?" This chapter focuses on reading skills that discuss about recount text, particularly in biographies. Several biographies of prominent people are offered in this content, together with their literary structures. There are also some quizzes and vocabulary concerning the biography itself. Based on these materials, the content of the questions must be in sync with what students have learnt about this material while generating the substance of the questions.

   c. Creating

      The next stage is to create after all of the documents have been organized. At this stage, files that did not previously exist were created to assist the production of a question product. For example, lattices, cover questions, item grading rules, and expert validation tools.

3. Construction/Realization

   At this step, a prototype, namely the primary design based on the initial concept, is created. In the context of education, the second and third steps mentioned above are referred to as the production stage. It means, the HOTS test instrument is created. There are 10 numbers

DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

of multiple choice and 5 numbers of Essay. The researchers develop the test being multiple choice and essay.

Based on the findings of the analysis of the teacher's questions, the researcher created questions based on the HOTS criteria by Alderson.

4. Test, Evaluation and Revision Phase

This stage is concerned with the quality of the design that will be created. Make judgments after deep reflection as well. The process of gathering, processing, and evaluating information in a systematic manner is referred to as evaluation. The first step at this stage was all of the items were validated by experts before the items were used. The suggestions were given by the experts can be seen below :

Table 2.
Revision from Expert

Revision from Expert I
Do not use short question narration
Improve your question naration to be detail question
Used word question (How and Why) do not use (What, When, Who)
Use a various question type

Revision from Expert 2
Cannot. It is not separated
Pay attention on the use of punctuation.
Use the word "construct".
Do not give a question. Give a command to retell about the song.

After knowing what was would be revised, the researcher re-make the product based on the revision. The new product was named Prototype II. The product/prototype II was the test instrument based on the HOTS concept with revision by the expert. The new product can be seen in this link: https://drive.google.com/drive/folders/1aDzRGgeOPHte9xksl-px__Zuc4dSamKT?usp=sharing

The classification of the product above based on cognitive level can be seen as follows:

DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

Table 3:
Analysis of Multiple Choice Test Instrument

Test Items	Cognitive Level	Category
1st Question	C4 (Analyze)	HOTS
2nd Question	C4 (Analyze)	HOTS
3rd Question	C4 (Analyze)	HOTS
4th Question	C4 (Analyze)	HOTS
5th Question	C4 (Analyze)	HOTS
6th Question	C4 (Analyze)	HOTS
7th Question	C4 (Analyze)	HOTS
8th Question	C4 (Analyze)	HOTS
9th Question	C5 (Evaluating)	HOTS
10th Question	C4 (Analyze)	HOTS

Table 4 :
Analysis of Essay Test Instrument

Test Items	Cognitive Level	Category
1st Question	C6 (Creating)	HOTS
2nd Question	C4 (Analyze)	HOTS
3rd Question	C5 (Evaluating)	HOTS
4th Question	C5 (Evaluating)	HOTS
5th Question	C6 (Creating)	HOTS

After the new product was created, the next phase was test trial to the students. It aims at measuring the empirical validity and reliability of the product.

5. Implementation Phase

   After analyzing the items developed, the quality of the items was obtained which will be implemented by the English teacher of class X MIA 2 MA Madani Alauddin Pao-Pao with the address at Paccinongan, Somba Opu, Gowa, South Sulawesi. At this stage, it will be implemented in schools as a *questions bank* based on HOTS criteria.

DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

C. *The Level of Validity and Reliability of Test Items*

1. Expert Validity

Table 5.
The result of Validation sheet by expert

Item No.	Examined Aspects	Validator Assessment	
		Validation 1	Validation 2
	MATERIAL OF QUESTIONS PRESENTED		
1.	The relationship between the questions and the question indicators	3	4
2.	The relationship of the question to the level of intellectual development of students	3	4
3.	Questions vary according to cognitive level	3	4
	Mean	3	5,3
	CONSTRUCTION		
1.	The questions are stated clearly	3	4
	Mean	3	4
	LANGUAGE		
1.	Using communicative sentences	3	3
2.	Using language that is in accordance with the correct Indonesian language rules	3	3
3.	Use simple and easy-to-understand language	3	3
4.	Sentences in each item of the question do not cause multiple interpretations	3	3
	Mean	3	3
	THE COVER OF QUESTION BANK		
1	Use an attractive cover design	3	3
	Mean	3	3

2. Empirical Validity

The analysis of the empirical validity and reliability of the HOTS items that have been developed can be seen from the test of the validity of the questions using correlation by *Statistical Product and Service Solution (*SPSS) program. Empirical validity testing, the researcher immediately tried it out instrument to students and then the results are analyzed with the validity test by using of the SPSS program. Based on

DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

the results of the students' work, the level of test validity can be calculated. The following data is the result of the calculation of the validity of the test based on SPSS.

Table 5 :
The Result of Multiple Choices and Essays validity

No	R count	R Table	Category
1	0,745	0,444	Valid
2	0,664	0.444	Valid
3	0,615	0,444	Valid
4	0,517	0,444	Valid
5	0,686	0,444	Valid
6	0,766	0,444	Valid
7	0,522	0,444	Valid
8	0,565	0,444	Valid
9	0,881	0,444	Valid
10	0,673	0,444	Valid

No	R count	R table	Category
1	0,813	0,444	Valid
2	0,767	0.444	Valid
3	0,564	0,444	Valid
4	0,599	0,444	Valid
5	0,565	0,444	Valid

1. Hasil Validasi Butir Soal Hots

Nomor Item	Aspek yang Dinilai	Penilaian Validator	
		Validation 1	Validation 2
	MATERI SOAL YANG DISAJIKAN		
1.	Keterkaitan antara soal dengan indikator soal	3	4
2.	Keterkaitan soal dengan tingkat perkembangan intelektual peserta didik	3	4
3.	Soal bervariasi sesuai dengan tingkatan kognitif	3	4
	Rata-rata	3	5,3
	KONSTRUKSI		
1.	Pertanyaan pada soal dinyatakan dengan jelas	3	4
	Rata-rata	3	4
	BAHASA		
1.	Menggunakan kalimat yang komunikatif	3	3
2.	Menggunakan bahasa yang sesuai dengan kaidah bahasan Indonesia yang benar	3	3
3.	Menggunakan bahasa yang sederhana dan mudah dimengerti	3	3
4.	Kalimat dalam setiap item soal tidak menimbulkan penafsiran ganda	3	3
	Rata-rata	3	3
	SAMPUL BANK SOAL		
1	Menggunakan desain sampul yang menarik	3	3
	Rata-rata	3	3

3. 1. Reliability Analysis

This reliability test is based on the results of trials validation test previously which all the test items are valid. So that the level of test reliability can be calculated by using SPSS application.

DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

a. Multiple Choice

Table 6 :
The Result of Multiple Choices Reliability

Case Processing Summary

Reliability

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	20	100.0
	Excluded[a]	0	.0
	Total	20	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.840	10

b. Essay

Table 6 :
The Result of Essays Reliability

Reliability

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	20	100.0
	Excluded[a]	0	.0
	Total	20	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.754	6

4. Dicussion

Based on the findings, there were three aspects that need to be summarized. The first was the existing of the test instruments. Both of syllabus and lesson plan were not suitable with 2013 curriculum, but the textbook was suitable. The result of interview analysis indicated that the teacher had one of the HOTS criteria, namely the use of a variety of questions in the creation of items, multiple choice, and essays. It is generally seen that the number of questions she made for the final exam per semester is 35 numbers, where each chapter represents 7 numbers, but depending on the number of chapters, she divides the numbers as evenly as possible for each chapter. Furthermore, for the questions, she created her own, but it was still related with the textbook, and some pictures were taken from it.

There were 5 teacher's test items, the first question can be classified into low level based on the HOTS (Higher Order Thinking Skill) concept because the question just applied remembering skill (C1). It did not need analysis and comprehension

DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

skill in which these two skill are classified in HOTS level. Related to the first question, the second question can be classified into low level, because it just need a remembering skill (C1). The students just need memorizing the material about verb 2. The third question, the fourth question and the fifth question were included in the category of HOTS questions because they are included in the evaluation stage where the characteristics of the HOTS questions are 3, namely C6 (creating) C5 (analyzing) and one of the criteria for the HOTS questions, namely C5 is in the question above.

Based on the findings of the analysis of the teacher's questions, the researcher created questions according to the HOTS criteria, there were 10 questions of multiple choice and 5 questions of essay. These test items had been revised by the expert and the researcher made a new product based on the suggestion. After the new product was created, the next phase was test trial to the students. The trial aims at measuring the empirical validity and reliability of the product.

The result of validation by expert developed in this study is in a score of 3.7which is very valid, based on predetermined criteria. HOTS items with aspects of the material then presented with a value of 3.5 are in the veryvalid category, the construction aspect shows a value of 4 in the very valid category, then the language aspect shows a value of 3.5 which is in the very valid category. The higher the VR (close to 4 or equal to 4), the higher the validity value of an item/item, and the lower the VR number (close to 0 or equal to 0), the lower of validity value of an item.

The result of empirical validity, the analysis of the empirical validity and reliability of the HOTS items that have been developed can be seen from the test of the validity of the questions using correlation by *Statistical Product and Service Solution (*SPSS) program. Technique that used by researchers is the value of 5% significance with a critical value. In other words, it can compare between *Rcount* with *Rtable*. Test results of the validity data shows that it is valid if *Rcount* is larger than *Rtable*. The next is the reliability analysis. The level of test reliability can be calculated by using SPSS application. Meanwhile, the tool that is used to get the realibility in this research by using Alpha Cornbach, that is SPSS program. The criteria of the reliabilty test is if the coefisient Alpha is larger than the significance 0,6, then the test are reliable. Based on the data above, the result of reliablility in multiple choice test by Cronbach's alpha is 0,840 which means all the test ietms are reliable. Beside, the essay test instrument shows score in 0,754 which means all the essay items test are reliable.

Furthermore, in order to be effective, the test should not only be valid and accurate, but it should also follow some criteria. The researcher mentioned certain language assessment principles. According to Brown (H.douglas brown 2013), an effective test must include some language assessment standards. There are several types of washback: practicality, reliability, validity, authenticity, and washback. The following are the test instruments based on the findings of this study; Practicality. The student does not require a large sum of money based on the results of the test instrument. In this situation, the test is not prohibitively expensive. The form of the test was created by the researcher using a Google form

DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

that was created online. The test does not need to be printed by the teacher. The test then does not require much time to complete. The test may then be administered conveniently and automatically in terms of scoring and grading. The next, reliability. The test is reliable and consistent. Using the SPSS tool, the researcher examined the reliability of the results of the respondents' answers. As a result, the Cronbach's alpha result for reliability in multiple choice tests is 0.840, indicating that all test ietms are reliable. Furthermore, the essay test instrument yields a score of 0.754, indicating that all of the essay items tested are reliable.

The third is validity. According to the SPSS program's results, the 10 multiple-choice questions and 5 essay questions are declared valid because all of the calculated rhitung are more than the rtabel values. With a total of 20 student respondents, it is declared that the 10 multiple choice questions assessed and 5 essays are valid. Next, Authenticity. It is a test that is based on the evaluation or construction of the test. The test instrument was created using the lesson plan, syllabus, and material. Furthermore, the test is based on the HOTS category and has been evaluated by ELT teachers. And the last Washback. The final option is washback. The test instrument is being returned to the students. The final test was reviewed and confirmed while being built. The test was carried out by the researcher using the HOTS category. In this situation, the test referred to HOTS characteristics as defined by Jurnal Penyusunan Soal HOTS (2017). It is possible to conclude that the test is washback since the test can cause pupils to think more critically while analyzing the test.

CONCLUSION

Based on the findings of the study, it is found that the extent of validity of the HOTS gadgets changed into 3.7 that is within side the very legitimate category, the reliability of the HOTS gadgets is 0.81 suggests in superb category, the issue degree of the HOTS gadget**s** is 0.60 suggests within side the medium category. The implications of developing HOTS items can train and encourage students to do high-level reasoning so that they are not fixated on the pattern of answers that result from memorization without understanding the concept of knowledge, and can make students think critically while working on the test in the "Sing a Song "Material so that answers can be produced based on in-depth reasoning.

References

Anderson, L. W., and D. R. Krathwohl. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives.* Boston: Allyn & Bacon.

Borg, W.R, and M.D. Gall. 2003. *Educational Research: An Introduction*. 4th Editio. London: Longman Inc.

Conklink, W. 2012. *Higher Order Thinking Skills to Develop 21st Century Learners*. Hungtinton Beach: Shell Educational Publishing, Inc.

Earl, L. 2006. "Assessment - A Powerful Level for Learning." *Brock Education* 16 (1).

Hamdi, Syukrul, Iin Aulia Suganda, and Nila Hayati. 2018. "Developing Higher-Order Thinking Skill (HOTS) Test Instrument Using Lombok Local Cultures as Contexts for Junior Secondary School Mathematics." *REID (Research and Evaluation in Education)* 4 (2): 126–35. https://doi.org/10.21831/REID.V4I2.22089.

H.douglas brown. 2013. *Language Assessment Principle and Classroom Practice*. *Journal of Petrology*.

DOI: https://doi.org/10.31605/eduvelop.v6i1.1870

King, F. J., Goodson, L., and F Rohani. 2010. *Assessment & Evaluation Educational Services Program: Higher-Order Thinking Skills.* Washington: : A publication of the Educational Services Program.

L. Azizah, S. Mariani, Rochmad. 2012. "Pengembangan Perangkat Pembelajaran Model Core Bernuansa Konstruktivistik Untuk Meningkatkan Kemampuan Koneksi Matematis." *Unnes Journal of Mathematics Education Research*.

Limbach, Barbara, and Wendy Waugh. 2010. "Developing Higher Level Thinking." *Journal of Instructional Pedagogies*, 1–9.

Nazaruddin, Rezkiawati. 2017. "Content Analysis Speaking Materials in English Textbook Based on 2013 Curriculum for the First Grade Student at Vocational High School." *Eduvelop: Journal of English Education and Development* 1 (1): 49–56. https://doi.org/10.31605/eduvelop.v1i1.2.

Ngah, Norulbiah, Zaleha Ismail, Zaidatun Tasir, and Mohd Nihra Haruzuan Mohamad Said. 2017. "Students' Higher Order Thinking Skills and Their Relationships with Problem Posing Ability." *Advanced Science Letters* 23 (4): 2876–79. https://doi.org/10.1166/ASL.2017.7589.

Sakka, Wahyuni, and Rezky Naim. 2019. "Application of A Joyful Learning Strategy Based on Humor Communication to Improve the Interests and Achievements of Learning English | Eduvelop: Journal of English Education and Development." 2019. https://ojs.unsulbar.ac.id/index.php/eduvelop/article/view/423.

Salamah, Umi. 2018. "PENJAMINAN MUTU PENILAIAN PENDIDIKAN." *Journal EVALUASI*. https://doi.org/10.32478/evaluasi.v2i1.79.

Wang, S., and H. Wang. 2014. "Teaching and Learning Higher Order Thinking." *International Journal of Arts and Sciences* 7 (2): 179–87.

Welle, Van de. 2007. *Elementary and Middle School Mathematics: Teaching Developmentally*. 5th Editio. Boston: Allyn and Bacon.

Widana, I Wayan. 2017. "Higher Order Thinking Skills Assessment (HOTS)." *Journal of Indonesian Student Assessment and Evaluation* 3 (2): 32.

Widana, I Wayan. 2018. "Higher Order Thinking Skills Assessment towards Critical Thinking on Mathematics Lesson." *International Journal of Social Sciences and Humanities (IJSSH)*. https://doi.org/10.29332/ijssh.v2n1.74.

Yonisa, Riza, Kurniawan I, and Dian Lestari. 2019. "The Development Assessment Instruments of Higher Order Thinking Skills on Economic Subject." *Dinamika Pendidikan* 14 (1): 102–15. https://doi.org/10.15294/DP.V14I1.19226.