Advances in Social Science, Education and Humanities Research, volume 384
4th Annual International Seminar on Transformative Education and Educational Leadership (AISTEEL 2019)

Development of Two-tier Multiple Choice Instrument to Measure Higher Order Thinking Skills

Rifka Annisa Girsang
Program of Postgraduate,
Universitas Negeri Medan
Medan, Indonesia
nissagirs16@gmail.com

Wawan Bunawan
Program of Postgraduate,
Universitas Negeri Medan
Medan, Indonesia
wanbunawan@gmail.com

Rita Juliani
Program of Postgraduate,
Universitas Negeri Medan
Medan, Indonesia
julianiunimed@gmail.com

***Abstract*— The study aims to develop Two-tier Multiple Choice (TTMC) test to measure students' Higher Order Thinking Skills (HOTS) in material Momentum and Impulse to standard qualifications of good test based in validity, reliability, difficulty index, discrimination index and effectiveness distractor. The type of research is developmental research, using the ADDIE model. The data analysis technique used is qualitative and quantitative technique. The results of this study based on qualitative analysis showed that quality of the test questions was very good with a percentage of validity of 91,33%. Quantitative analysis of the quality of the TTMC test is good. Analysis of 27 items in the small group trials showed test obtained 24 items received and 3 items were rejected. The large group test was obtained: (1) 19 valid items (71%), 5 invalid items (21%). (2) The test having enough reliability is 0.587. (3) 3 (13%) easy items, 12 (50%) medium items and 9 (38%) difficult items. (4) Discrimination index of the questions was obtained by 14 (58%) questions in the excellent category, 5 (21%) questions in the good category and 5 (21%) questions in the bad category. (5) 19 (79%) were effective items and 5 (21%) were ineffective items. Items received were 12 (50%) questions, 7 (29%) items revised and 5 (21%) items rejected.**

Keywords*— *Development, Two-tier Multiple Choice, Validity, Reliabilitty

I. INTRODUCTION

Educational goals can be measured through evaluation activities in school [1] Evaluation activities in learning are balanced with the application of the curriculum. The curriculum in force in Indonesia is the 2017 revised curriculum in 2017 which integrates strengthening character education in learning including religious, nationalist, independent, mutual cooperation, and integrity. Integrating 21st century skills or termed 4C (creative, critical thinking, communicative, and collaborative) and High Order Thinking Skills (HOTS) [2] High Order Thinking Skills (HOTS) were applied following a low ranking Program for International Mathematics and Sciense Study (PISA) and Trends of the International Mathematics and Science Study (TIMSS) compared to other countries, so that the standard national examination questions are tried to be improved to catch up [3].

The results of the PISA test in 2012, Indonesia is in 64th position from 65 countries. The average science score of Indonesian students is 382 and the average OECD science score is 501, successive results have occurred over the past ten years, the new 2015 TIMSS results published in December 2016 are not far from the results in 2012, the achievements of Indonesian students in science got ranked 46th out of 51 countries with a score of 397. Teaching and learning activities require assessment to determine the level of understanding and success of students. In research more focused on the assessment of knowledge. Knowledge assessment was measured using tests in the form of questions that covered the cognitive domains C1 to C6 based on the revised Bloom taxonomy. Anderson & Krathwohl [4] stated that the cognitive domains of questions C1, C2, and C3 are categorized as low-level thinking skills (LOTs) while the cognitive domains of the questions C4, C5, and C6 are classified as high order thinking skills (HOTs).

Barnett & Francis [5] argue that high-level thinking questions can encourage students to think deeply about subject matter, so that it can be said that high-level thinking ability tests can provide stimuli to students to develop high-level thinking. Developing a standardized test of Higher Order Thinking Skills (HOTS) needs to be done because it can train and familiarize students with the questions in the form of HOTS. Reality in the field, questions tend to test more aspects of memory included in LOT (Lower Order Thinking). This test aims to show the level of ability and success of students in solving problems at a high level. Brookhart explains the basic requirements for testing high-level thinking skills is requiring tasks that require the use of knowledge and skills in new situations. To carry out tests on high-level thinking skills must use new materials, one of which is to use a two-tier multiple-choice (TTMC) instrument. The diagnostic two-choice choice (TTMC) test is a diagnostic test in the form of a two-tiered multiple choice question that was first developed by David F. Treagust in 1988 [6].

Treagust has founded a conceptual test composed of two-tier multiple-choice questions to find out student misconceptions because of two big benefits of multiple choice

questions. First, they make it possible to investigate two aspects of the same phenomenon. Students are asked to predict the results of a particular situation at the first stage and to give their reasons on the second level, the reason students give details of their alternative concepts. Second, they reduce the uncertainty of measurement from guesses of students' guesses. Students have a 25% chance to guess correctly in a question with four choices, in a two-level question, students must respond correctly at both levels, so they only have a 6.25% chance to guess correctly. The development of Two-Tier Multiple Choice (TMCC) instruments developed by researchers used the ADDIE design, a design of educational product development and other learning resources consisting of Analyze, design, development, implementation and evaluation (ADDIE). ADDIE is a product development concept established to build performance-based learning.

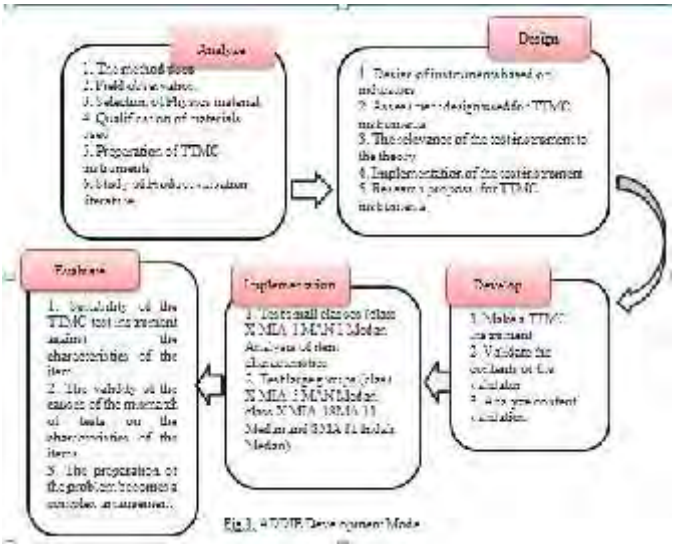
Products made using the ADDIE process are effective because they function as a framework guide for complex situations in developing educational products and learning resources [7]

The ADDIE concept is a strategy that is distant, limiting, passive, and singular from a didactic design model then shifts to a more active and multifunctional design model which is an inspirational approach to learning [7]. Assessment is important to measure the higher order thinking skills of students examined by Istiyono, dkk [8] states that high-level physics thinking instruments (PhysTHOTS) meet the requirements used to measure high-level physics thinking skills of high school students. Kusuma, et al, [8] showed that the results of his research, namely the HOTS instrument that had been developed, could help students practice their high-level thinking skills as an assessment for learning. Field trials to train HOTS students, it can be seen that students with HOTS abilities are categorized as good.

Barniol and Zavala [9] modified multiple choice questions and discussed the reasons behind not relying on others and obtained good results and could be used by teachers and researchers to assess students' understanding of mechanical waves. The method developed by Barniol and Zavala proved to be of good quality. Kamcharean & Wattanakasiwich

II. . RESEARCH METHOD

This research is a development research. The developed product is instrument assessment to train student's higher order thinking (HOTS). development type is adapted from ADDIE type which consists of 5 development steps. However in this research is used 5 steps only, which consist of: 1) Analyze, 2) Design, 3) Development, 4) Implementaion, 5) Evaluate. The steps of research and development with ADDIE model are shown in the following figure 1.



The instruments used in this study include learning device instruments and data collection instruments. Then, Instrument test which has been arranged is used to do limited try out. Result of limited try out that will determine how the question parameter developed is, such as reliability, distinguishing power, and difficulty level. Limited try out is done in MAN 1 Medan with 25 number of samples. Then, Questions which have been known reliability, distinguishing power and difficulty level are arranged into early product which are used for field try out. Field try out Then, Instrument assessment which has been revised is tried out 3 Senior High School in Medan. the number of samples used is one class for each school with the number of students shown in Table 1.

TABLE 1. THE NUMBER OF STUDENTS USED IN THE RESEACH

No	School	Number Of Student
1.	MAN 2 Medan	38
2.	SMAN 11 Medan	35
3.	SMA IT Indah Medan	20
Total		93

Table 1 the number of students used in the research No School Number of students 1) MAN 2 Medan 38, 2) SMAN 11 Medan 35, 3) SMA IT indah Medan 20 total 93. The result of field try out is done to see the instrument assessment which has been developed as assessment for learning for students in training their HOTS. The data needed in this study are qualitative data and quantitative data.

Scoring guidelines used in research use two-tier multiple choice instruments that refer to Beyrak [10] Scoring has been adapted as a Table 2.

TABLE 2. SCORING GUIDLINES

criteria	Skor
No answer	0
Answer more than one	0
One correct answer in the second tier	0
One correct answer in First Tier	1
Two correct answer to First Tier and second Tier	2

III. RESULT AND DISCUSSION

A. Analyze

The results of the analysis by conducting interviews, distributing questionnaires and initial ability tests. Interview with one of the study teachers who said that they still use the usual multiple-choice and essay evaluation forms. The results of the questionnaire distributed indicate that the average interest of students in learning physics is included in the medium category. The results of the initial ability test state that students from four schools have the same initial abilities, so it is feasible to conduct an instrument trial

B. Design

The design of the Two-tier Multiple Choice instrument (TTMC) is to create a question grid that refers to indicators of achievement of learning competencies. The instruments developed are instruments based on Higher Order Thinking Skills (HOTS). The indicators developed are indicators of problem solving and critical thinking. Higher order Thinking Skills (HOTS) instruments carried out are in the form of a Two-Tier Multiple Choice (TTMC) or two-tier multiple choice. Indicators of problem solving and critical thinking Level are adjusted to Bloom's taxonomy cognitive level starting from levels C4 to C6.

C. Development

Instrument validation is done so that the product developed is valid and suitable for use. The material expert in this study was Mrs. Dr. Derlina, M. Sc and Mr Dr. Nurdin Siregar. M.Sc, namely lecturers from the Physics Education Study Program, Postgraduate Program, Medan State University and senior teacher Mr Pandapotan Harahap, M. Pd, M. PFis. Validation is carried out related to the three aspects, namely material, construction and language aspects. The results of the instrument validity assessment can be seen in Table 3

TABLE 3. THE RESULTS OF THE INSTRUMENT VALIDITY ASSESSMENT

No.	Indicator	Skor %	criteria
A.	**Material**		
1.	According to the indicator	90	Exellent
2.	Make one correct or correct answer	91	Exellent
3.	Content of the material in accordance with the measurement objectives	91	Exellent
4.	Logical illiterate and functioning	92	Exellent
B.	**Construction**		Exellent
5.	The subject matter is clearly formulated	93	Exellent
6.	The formulation of the questions and choices is clearly formulated	94	Exellent
7.	The subject matter does not shoe towards the correct anser	90	Exellent
8.	The subject matter does not contain multiple negative statements	93	Exellent
9.	Homogeneous and logical answer choices	90	Exellent
10.	The length of the formula is relatively the same	91	Exellent
11.	Answer choices don't use statements " all the answers above are correct or all of the answers above are wrong"	89	Good
12.	Answer choice in the form of numbers are arranged in sequence, while the answer choice in the form chronologically	95	Exellent
13.	Pictures, graphics, tables and diagrams contained in the problem clearly and functioning	92	Exellent
14.	The item does not depend on the previous answer	89	Exellent
C.	**Language**		
15.	Formulation of communicative sentence	88	Good
16.	The question of using languages that are in accordances with Bahasa	92	Exellent
17.	The sentence foemulation does not give rise to multiple interpretations or misconception	93	Exellent
18.	Not use local language	91	Exellent
Total		91.33	Exellent

Table 3 shows the average percentage of the TTMC instrument rating is 91.33% or "very good". The assessment by the lecturer and teacher aims to determine the feasibility of the instrument before use at school. Improvements have been made according to the advice of senior lecturers and teachers

D. Implementation

a. Trial Small group. Small group trials are conducted after obtaining approval from experts. The trial was conducted in class X MIA-1 MAN 1 Medan with 25 students. The results of empirical analysis of small group trials of critical thinking problems and problem solving abilities can be seen in Table 4 and Table 5.

TABLE 4. INTERPRETATION OF THE RESULTS OF SMALL GROUP TRIAL QUESTIONS OF CRITICAL THINKING

No.	Validity	Level of Difficulty	Appropri-ateness	Discriminati on index	Conclusi on
1.	Valid	Medium	Good	Efektif	Accepted
2.	Valid	Difficult	Exellent	Efektif	Revised
3.	Valid	Difficult	Good	Efektif	Revised
4.	Valid	Easy	Good	Efektif	Revised
5.	Invalid	Easy	Enough	Inefektif	Rejected
6.	Valid	Medium	Exellent	Efektif	Accepted
7.	Valid	Medium	Exellent	Efektif	Accepted

8.	Invalid	Easy	Ugly	Infektif	Ditolak
9.	Valid	Difficult	Exellent	Efektif	Revised
10.	Valid	Easy	Good	Efektif	Revised
11.	Valid	Medium	Good	Efektif	Accepted
12.	Valid	Difficult	Good	Efektif	Revised
13.	Valid	Medium	Good	Efektif	Accepted
14.	Valid	Medium	Good	Efektif	Accepted

TABLE 5. INTERPRETATION OF SMALL GROUP TEST RESULTS PROBLEM SOLVING PROBLEMS

No.	Validity	Level of Difficulty	Appropri-ateness	Discrimination index	Conclusion
1.	Valid	Medium	Exellent	Efektif	Accepted
2.	Valid	Medium	Enough	Efektif	Revised
3.	Invalid	Difficult	Ugly	Efektif	Rejected
4.	Valid	Medium	Exellent	Efektif	Accepted
5.	Valid	Medium	Good	Inefektif	Accepted
6.	Valid	Difficult	Cukup	Efektif	Revised
7.	Valid	Medium	Good	Efektif	Accepted
8.	Valid	Difficult	Enough	Inefektif	Revised
9.	Valid	Difficult	Enough	Efektif	Revised
10.	Valid	Medium	Exellent	Efektif	Accepted
11.	Valid	Difficult	Good	Efektif	Revised
12.	Valid	Medium	Good	Efektif	Accepted
13.	Valid	Medium	Exellent	Efektif	Accepted

b. Trial of Large Groups. Large group trials were carried out after revision of the analysis questions in small groups. Trial large groups use 24 revised questions. The results of the validity, reliability, level of difficulty and deception effectiveness are briefly presented in table 6 and table 7

TABLE 6. INTERPRETATION OF THE RESULTS OF SMALL GROUP TRIAL QUESTIONS OF CRITICAL THINKING

No	Validity	Level of Difficulty	Appropri-ateness	Discrimination index	Conclusion
1.	Valid	Medium	Exellent	Effektive	Accepted
2.	Valid	Difficult	Good	Effektive	Revised
3.	Valid	Medium	Exellent	Effektive	Accepted
4.	Invalid	Easy	Ugly	Ineffektive	Ditolak
5.	Valid	Medium	Exellent	Effektive	Accepted
6.	Invalid	Difficult	Ugly	Ineffektive	Rejected
7.	Valid	Difficult	Good	Effektive	Revised
8.	Valid	Medium	Exellent	Effektive	Revised
9.	Valid	Medium	Exellent	Effektive	Accepted
10.	Valid	Difficult	Exellent	Effektive	Revised
11.	Valid	Medium	Exellent	Effektive	Accepted
12.	Valid	Difficult	Exellent	Effektive	Revised

TABLE 7. INTERPRETATION OF SMALL GROUP TEST RESULTS PROBLEM SOLVING PROBLEMS

No. Soal	Validity	Level of Difficulty	Appropri-ateness	Discrimination index	Conclusion
1.	Valid	Medium	Exellent	Effektive	Accepted
2.	Valid	Medium	Exellent	Effektive	Accepted
3.	Valid	Medium	Good	Effektive	Accepted
4.	Invalid	Difficult	Ugly	Effektive	Revised
5.	Valid	Difficult	Good	Ineffektive	Revised
6.	Valid	Medium	Good	Effektive	Accepted
7.	Valid	Medium	Exellent	Effektive	Accepted
8.	Invalid	Easy	Ugly	Ineffektive	Rejected
9.	Valid	Medium	Exellent	Effektive	Accepted
10.	Valid	Difficult	Exellent	Effektive	Revised
11.	Invalid	Difficult	Ugly	Effektive	Revised
12.	Valid	Difficult	Exellent	Effektive	Revised

E. Implementasion

The results of the large group trial showed that the questions received were 6 questions (50%), the questions needed to be revised were 3 questions (25%) and the questions that had to be rejected were 3 questions (25%). The results of the quantitative analysis which includes the analysis of validity, reliability, differentiation, level of difficulty and effectiveness of deception, need to be followed up on the items in question. If all four are good, then the item is appropriate to be used as an evaluation tool. If there is one aspect or more than four aspects that are not fulfilled, then the item in question must be corrected. There are 3 possible follow-up actions, including being accepted, revised, rejected. Good items can be stored in the question bank for later use in future tests. The poor questions were revised and tested again in the upcoming tests. Items that are not well discarded.

F. Discussion

This research is the development of Two-tier Multiple Choice Instrumen on Momentum and Impulse material for participants student in class X high school. Development of TTMC instrumen through five stages of development, namely Analyze, design, develop, implementation and Evaluation.

The results of the item analysis of the Two-tier Multiple Choice Test (TTMC) on the material Momentum and Impulse in High School have a validity score of 91.37% which is included in the very high category in line with the research of Viana and Subroto [11]. The data above shows that 8 items can be received and stored in the Two-tier test questions multiple choice on momentum and impulse material in high school because they have met the validity, level of difficulty, differentiation, effectiveness of good deception. There are 8 items still need improvement because they have not fulfilled a good differentiating power. While the 5 items are rejected and cannot be used because they do not meet any criteria of validity, level of difficulty, differentiation, and effectiveness

of good deception [12]. The average level of difficulty reaching the critical thinking questions 0.402 means that the level of difficulty of the two-tier multiple choice test (TTMC) is in the moderate category. While the average differentiation reaches 0.506, this means that the differentiation of the two-tier multiple choice test (TTMC) test on momentum material and Impulse is in a good category [13]. The revised question can be used as a question bank for two-tier multiple choice test (TTMC) where the number of questions is reduced to 19 questions.

Research uses classic test theory where the conditions and characteristics of test participants will influence the results of the study. The conditions referred to are the internal and external factors of the test takers. Internal factors include the level of intelligence, motivation, health and so on. The conducivity of the test room is one of the external factors that also affects the condition of the test takers. The more conducive the exam room, the better the results will be. The lower the ability of the test group, the more difficult the test items [14].

The results showed that the two-tier Multiple Choice test instrument (TTMC) developed could measure high-level thinking skills (HOTS). The results of field trials were obtained by students with HOTS abilities with a high category of 3%, enough by 30% and less than 67% according to the study kusuma.

The TTMC HOTS instrument that has been developed is very effective in measuring higher order thinking skills of students because this instrument is a two-level multiple choice evaluation, the first level is related to knowledge statements. The second level resembles the first level question format but aims to encourage thinking and higher order thinking skills thinking. The results of the large group on critical thinking questions amounted to 28.46% and the problem solving ability problem was 28.66%. Syahrul and Setyarsih [15] state the causes of misconceptions experienced by students are identified from students' mistakes in choosing reasons that are not right at the second level (two-tier). Impostors provided in the two-tier section are specifically designed to be able to describe the causes of misconceptions from preconception to intuition. Kurniasih and Haka [16] state the category of misconception seen from the type of correct answer at both levels of the question.

IV. CONCLUSION

The Two-tier Multiple Choice Test Instrument on Momentum and Impulse Materials in SMA is declared feasible and fulfills the criteria as a valid and effective question with the results of content validation having an average ideal of 91.33%. Characteristics of the Two-tier Multiple Choice test instrument on material Momentum and Impulse in High School is Good. Analysis of 27 items obtained 8 (30%) items received, 11 (40%) items were revised, 8 (30%) items were rejected. The validity of the items obtained was 19 (70.4%) valid items and 8 (29.6%) items were invalid. The question has sufficient reliability which is 0.587. The level of difficulty, as many as 3 easy questions (15.5%), 14 moderate questions (51.8%), and 7 difficult questions (25.9%). Based on the differentiating power, there were 12 questions including the excellent category (44.4%), 7 questions including the good category (25.98%), 2 questions including the adequate category (3.7%), and 3 questions including the bad category (11.1%). The effectiveness of deception, there are 19 effective questions (70.4%) and 8 ineffective questions (29.6%).

TTMC instruments that have been developed are effective for measuring students' understanding in studying Higher Order Thinking Skills (HOTS) questions and honing their ability to have Higher Order Thingking Skills (HOTS) so that assessment of instruments can be used as assessment for learning for students. students in solving the Two-tier Multiple Choice (TTMC) problem in each sub concept on momentum material and deep impulses with an average of 24.73% on critical thinking questions and 38.3% on problem solving skills.

References

[1] Anggraini, N dan Wasis. (2014). Pengembangan Soal IPA-Fisika Model TIMMS (Trends International Mathematics and Science Study). Jurnal Inovasi Pendidikan Fisika, 3(1): 15-18

[2] Rachman, D. A. (2018) .https://nasional.kompas .com/read/2018/05/28/08485331/mendikbud-pastikan-hots-tetap-dipakai-dalam-ujian-nasional-tahun-depan. (Retrieved on January 02 2019)

[3] Ariyana, Y., Pudjiastuti, A., Bestari, R dan Zamroni. (2018). Buku Pegangan Pembelajaran Berorientasi pada Keterampilan Berfikir Tingkat Tinggi. Jakarta: Direktorat Jenderal Guru dan Tenaga Kependidikan Kementrian Pendidikan dan Kebudayaan.

[4] Anderson, L & Krathwohl, D. R. (2015). Kerangka Landasan Untuk Pembelajaran, Pengajaran, dan Assesmen: Revisi Taksonomi Bloom, Pustaka, Yogyakarta.

[5] Barnett, J. E and Francis, A.L. (2012). Using Higher Thinking Questions to Foster Critical.

[6] Tuysuz, C. 2009. Development of Two-Tier Diagnostic Instrument and Assess Students Instruments's Understanding in Chemistry. Scientific Reseach and Esssay, 4(6). 626-631.

[7] Branch, R. M. 2009. Instructional Design-The ADDIE Approach. New York: Springer.

[8] Kusuma, M. D., Rosidin, U., Abdurrahman dan Suyatna, A. (2017). The Development of Higher Order Thinking Skill (Hots) Instrument Assessment in Physics Study. IOSR-JRME, 1(7) 26-32

[9] Barniol, P dan Zavala, G. (2016). Mechanical Waves Conceptual Survey: Its Modification and Conversion to a Standart Multiple-Choice Test. Physical Review Physics Education Reseach, 12: 1-12

[10] Shidiq, A.S., Masykuri, M., Susanti, E. (2014). Pengembangan Instrumen Two-Tier Multiple Choice untuk mengukur keterampilan berfikir tingkat tinggi(HOTS) pada materi kelarutan dan hasil kali kelarutan untuk siswa SMA/MA kelas XI, JPK 3(4) : 83-92

[11] Viana, R. V dan Subroto. (2016). Pengembangan Sistem Assessment Dalam Pembelajaran Materi Usaha dan Energi Berbasis Media Audio Visual di SMA Negeri 1 Prambanan. Jurnal Pendidikan Fisika, 5(5): 311-319

[12] Muslim, Suhandi, A., & Nugraha, G. (2017). Development of Reasoning Test Instruments Based on TIMSS Framework for Measuring Reasoning Ability of Senior High School Student on the Physics Concept. Journal of Physics, 812 (1). IOP Publishing

[13] Kara, F., & Celikler, D. (2015). Development of Achievment Test: Validity and Reliability Study for Achievment Test on Matter Changing. Journal of Educatoin and Practice, 6(24): 21-26

[14] Adeleke, A. A., & Joshua, E. O. (2015). Development and Validation of Scientific Literacy Achievment Test to Asses Senior Secondary School Stidents' Literacy Acquisition in Physics. Journal of Education and Practice, 6(7), 28-42

[15] Syahrul, D. A dan Setyarsih, W. (2015). Identifikasi Miskonsepsi dan Penyebab Miskonsepsi Siswa dengan Three-tier Diagnostik Test Pada Materi Dinamika Rotasi. Jurnal Inovasi Pendidikan Fisika (JIPF), 4(3): 67-70

[16] Kurniasih, N dan Haka, N., B. (2017). Penggunaan Tes Diagnostik Two-Tier Multiple Choice Untuk menganalisis Miskonsepsi Siswa Kelas X Pada Materi Archaebacteria dan Eubacteria. Biosfer Jurnal Tadris Pendidikan Biologi, 8(1): 114-127.