

Developing the Two-Tiers Multiple Choice Test Based on Content, Construct and Language Feasibility in Identifying Students' High-Order Thinking Skills

Suyedi Hendra Yanto

Department of Biology Education
Postgraduate, Universitas Negeri
Medan

hendra.suyedi54@gmail.com

Herbert Sipahutar

Department of Biology Education
Postgraduate, Universitas Negeri
Medan

Syahmi Edi

Department of Biology Education
Postgraduate, Universitas Negeri
Medan

Abstract – This type of research is a Research and Development (R & D) referred to Borg & Gall model. The aims of this study were to find out the content, construct, and language feasibility of the two tiers multiple choice test that have been developed. The data obtained was adjusted to the objectives and development models used to the type of data to be collected in this development is quantitative description data as the main data, collected through items or two tiers multiple choice instruments that students have answered using scoring guidelines from Bayrak (2013). The results showed that the validation of material/content feasibility indicators assessed by material/content validators indicated that the assessment score obtained was 10, the average score was 3.33, with a very feasible category. The validation of construct feasibility indicators assessed by the construction validators indicated that the assessment score obtained was 33, the average score was 3.44, with a feasible category. The validation of language feasibility indicators assessed by language validators indicated that the assessment score obtained was 15, the average value was 3.75, with a very feasible category. Based on the average score of indicator assessed by the validator teams consisting of 3 indicators; material/content, construct and language feasibility obtained the average value of 4 with a very feasible category.

Keywords – *Two-Tiers Multiple Choice Test; High-Order Thinking Skill; Human Circulatory System*

I. INTRODUCTION

Biology learning is science learning that requires investigation or experimentation as part of scientific work. Scientific work emphasizes students to think creatively, critically, analytically, and divergently (Depdiknas, 2006). The ability of students to think critically and creatively includes in the form of high-order thinking skills. The importance of mastering high-order thinking skills is found in several points of Middle School Graduates Competency Standards. The expected points are students can build and apply information or knowledge logically, critically, creatively, and innovatively; demonstrate the ability to think

logically, critically, creatively, and innovatively in decision making; and demonstrate the ability to analyze and solve complex problems (Permendiknas No. 23, 2006).

The low level of thinking of students is one of the problems of education in Indonesia. This can be seen from the results of international studies; TIMSS (Trends in Mathematics and Science Study). There are three aspects of TIMSS cognitive abilities that are measured, namely, knowing (understanding), applying (application) and analyzing (reasoning). The study results reported by TIMSS, 2015 showed that the scores achieved by Indonesian students were still below the international average score of 500. Indonesia was still ranked 45th out of 48 participating countries with an average score of 397 (IEA, 2016). Likewise with the results of studies for science conducted by the PISA (Program for International Student Assessment) in the last two years 2012 and 2015, according to the Organization for Economic Cooperation and Development (OECD) Indonesia is ranked 64th out of 65 participating countries with a score of 382 in 2012, with an international average score of 500, whereas in 2015 the average score of Indonesian students 'achievement for science was ranked 62nd out of 70 participating countries with a score of 403, meaning that students' abilities in Indonesia were still low compared to other countries and is below the international average level (OECD, 2015).

TTMC is a diagnostic test in the form of a two-tiered MCQ that was first developed by David F. Treagust in 1988 which was further developed in 2006. The TTMC is a two-tiered multiple choice, where the first level contains questions about the concepts tested while the second level contains the reasons for each answer to the question at the first level as a form of diagnostic tests.

The study conducted by Tuysuz (2009) shows that many students only memorize the subject matter, where the level of student learning occurs in the low level cognitive domain so that they are unable to deal with the two tiers multiple choice instrument that requires high cognitive skills. Students can also easily guess the answers in traditional multiple choice

questions with 5 choice answer options and the possibility of guessing the correct answer is 20 percent. But on the two tiers multiple choice instrument is 4 percent, students usually face exams in the form of traditional multiple choice tests so students learn problem solving techniques rather than studying the subject of the test. These results were obtained from the research conducted on 9th grade students by giving 15 questions of two tiers multiple choices to 141 students. The traditional multiple choice test ($X = 9.40$) was higher than the average of the two-level test ($X = 5.48$), this shows that students are more successful in traditional multiple choice tests because they easily guess answers and plagiarism.

Halaydina & Downing (1989) and Treagust (2006) suggest that one of the advantages of the form of the two tiers multiple choice problem is being able to measure students' cognitive abilities at a higher level (high-order thinking). The question of two tiers multiple choices can be used to test students' understanding and identify the high-order thinking skills students have. Cullinane (2011) suggests that the inclusion of reasons at the second level can be used to improve high-order thinking skills and see students' ability to give reasons. In addition, the inclusion of reasons for the second level of this question can be used to reduce the occurrence of chancy answers that are often a weakness of the usual MCQ form. Objective assessment, which is objective, easy and fast, is the advantage of two-tiers multiple choices compared to other high-order thinking skills, for example essays. The model of the assessment of two tiers multiple choices proved to be very successful in educational settings (Boo, 2002; Kwen & Cheng, 2005; Adodo, 2013). Two tier multiple choice diagnostic tests have been used to determine students' understanding of concepts in several fields of study such as biology, chemistry and physics. Research on two-tier multiple choice diagnostic tests has been carried out on several topics of biology, such as photosynthesis (Griffard & Wandersee, 2001), transport in plants and the circulatory system in humans (Wang, 2004), genetics (Kılıç & Sa lam, 2009), cell division and reproduction (Sesli & Kara, 2012). In Indonesia, research on the development of two-tier multiple choices has been carried out on several biological materials, including those on Archaeobacteria and Eubacteria (Septiana et al., 2014) and kingdom plantae (Nofiana, 2016).

The results of the TIMS, PISA and several researchers above showed that high-order thinking of Indonesian students is still relatively low. Likewise, interviews with teachers as well as preliminary observations have been made at SMA Negeri (Public Senior High School) 15 Medan 15 where 30 multiple choice midterms have the composition of instruments at levels C1, C2, C3, C4 which are 20%, 53.3%, 16.7%, 10% respectively cognitive domain of Bloom and only 13.3% of students were able to correctly answer three C4 questions from 30 questions, whereas ideally the formative tests carried out by the teacher 80% must include high-level thinking skills (C4-C6) (BAN Assessment Standards, 2012).

This shows the lack of use of test instruments to uncover high-order thinking skills of students, it shows that students do not yet have the ability to solve non-routine problems or

questions that are required to think higher. Given the challenges of improving the quality of education in various aspects of life, it becomes very important to instill and improve students' high-order thinking skills. This can be achieved if education in schools is directed not only at the ability to memorize and understand scientific concepts, but also on improving students' abilities and thinking skills specifically for high-order thinking skills. Test instruments used to measure high-order thinking skills can use assessment types such as modified multiple choice. According to Ramirez & Ganaden (2008), one of the modified multiple choice alternatives that can be used to measure high-order thinking skills is a form of two-tiers multiple choice.

Therefore, it is necessary to develop a second-tier multiple choice diagnostic test (TTMC) that can identify students' high-order thinking skills in order to create a good learning environment even students can find their own knowledge or an answer by constructing the knowledge they have acquired. One such learning environment is a test. A good test will help students construct their knowledge, because the test not only measures student achievement but the test instrument must be able to be a good learning tool for students, with the right test will help students master learning. With TTMC, high-order thinking skills will be increasingly stimulated to grow or vice versa without using TTMC in the learning process so that a student will find it difficult to have high-order thinking skills.

II. RESEARCH METHOD

The research has been conducted in May to June 2018. The trial test was carried out to the eleventh grade students of SMAN 15 Medan at Jalan Sekolah Pembangunan, Medan Selayang District, Medan City, North Sumatera. This type of research is a Research and Development (R & D) referred to Borg & Gall (1987). The development of these test instruments was carried out based on the need analysis that teachers in schools required some evaluation instruments that were capable of measuring high-order thinking skills.

The development of test instruments consists of several stages, namely: (1) Problem Identification (analysis of problems and needs/collecting information: literature review, observation of test instruments, and identifying problems); (2) Planning (product development planning in the form of a test instrument for students); (3) Development of the Preliminary Form of the Product (developing the initial form of the product by applying the Borg & Gall model); (4) Preliminary Field Test (initial field trial); (5) Revision Product (revised test instrument based on expert validator); (6) Field Test (wider field trial); (7) Revision of the Product (revision based on reference to suggestions, criticisms and previous assessments); (8) Main Field Test (Final field trial); (9) Revision (Revision or repair if needed). The results of the two tiers multiple choices obtained in the first trial were used to test the validity, item validity, reliability, and item analysis on the TTMC.

Assessment was carried out by paying attention to students' answers to the two tiers presented, first tier and second tier. The following are the assessment criteria for

student answer scoring on TTMC items as revealed by Tuysuz, in which (1) two correct answers to first and second tiers are categorized as understanding; (2) only one correct answer in first tiers is categorized as a misconception; (3) only one correct answer on the second tiers is categorized as guessing; (3) two wrong answers, namely first and second tiers categorized as not understanding; (4) answering more than one choice/no answer at all categorized as not understanding.

The data obtained was adjusted to the objectives and development models used so the type of data to be collected in this development is quantitative description data as the main data, collected through items or two tiers multiple choice instruments that students have answered using scoring guidelines from Bayrak (2013).

III. RESEARCH RESULTS

3.1. Results of Developing Instruments for High-Order Thinking Skill Tests

3.1.1. Validator Team Assessment

3.1.1.1. Assessment Based on Construction Validators (Construct)

Evaluation of the material validator on construction feasibility consisting of 9 assessment indicators, namely: (1) classification of questions (2) the subject matter is formulated briefly, clearly and firmly (3) the formulation of the subject matter and choice of answers is the only statement needed does not give a clue to the answer key (5) the subject matter does not use multiple negative statements; (6) graphic/table/diagram/images are clear and functioning; (7) the length of the formulation of the question is relatively the same; (8) choice of answers does not use the statement "all the correct answers" or "all answers are wrong"; (9) question items do not depend on the answers to the previous questions.

TABLE 1. THE RESULTS OF CONSTRUCTION VALIDATOR ASSESSMENT

CONSTRUCTION (Construct)	Validator Assessment I
1. ** Problem classification	3
2. The subject matter is formulated briefly, clearly, and firmly	3
3. The formulation of the subject matter and choice of answers are only necessary statements	4
4. The subject matter does not give an answer key	4
5. The subject matter does not use multiple negative statements	4
6. Images/graphs/tables/diagrams/etc are clear and functional	3
7. The length of the formulation of the questions is relatively the same	3
8. Choice of answers does not use the statement "all the correct answers" or "all the answers are wrong"	4
9. The question item does not depend on the answer to the previous question	3
Total Value	31
Average	3.44
Category	Feasible

Information:

1: Not feasible
2: Less feasible

3: Feasible
4: Very feasible

3.1.1.2. Assessment Based on Language Validators (Readability)

Evaluation of material validator on construction feasibility consisting of 4 assessment indicators, namely: (1) in accordance with the rules of the Indonesian language; (2) communicative language; (3) answer choices do not repeat; (4) do not use taboo language.

TABLE 2. THE RESULTS OF LANGUAGE VALIDATOR ASSESSMENT

LANGUAGE (Readability)	Validator Assessment II
1. Use language that is in accordance with the rules of Indonesian language	3
2. Using communicative language	4
3. Choice of answers does not repeat the same word / group of words, unless it is a unified understanding	4
4. Using standard language	4
Total Value	15
Average	3.75
Category	Very Feasible

Information:

1: Not feasible
2: Less feasible

3: Feasible
4: Very feasible

3.1.1.3. Assessment Based on Material/Content Validator

Evaluation of the material validator about the material feasibility consisting of 4 assessment indicators, namely: (1) in accordance with the material in eleventh grade students; (2) distractors must function; (3) there is only one right answer key.

TABLE 3. THE RESULTS OF MATERIAL VALIDATOR ASSESSMENT

MATERIAL (Content)	Validator Assessment III
1. The question must be in accordance with the material in eleventh grade students	3
2. The distractors must function	3
3. There is only one correct answer key	4
Total Value	10
Average	3.33
Category	Feasible

Information:

1: Not feasible
2: Less feasible

3: Feasible
4: Very feasible

The level of two multiple choice questions have been assessed by 3 validators, the aspects being assessed are content/material, construct, and language feasibility. In the material feasibility, validator 1 gives a score of 3.33 and the researcher makes improvements or revisions to the test instrument until the test instrument is feasible to use with the final score of 4 categories very feasible. In the construct

feasibility, validator 2 gives a score of 3.44 with a feasible category, and the researcher makes revisions to obtain a score of 4 with a very feasible category. In the language feasibility, the validator gives a score of 3.75 with a very feasible category, and the researcher makes revisions to obtain a score of 4 with a very feasible category.

TABLE 4. THE RESULTS OF THE COMPARISON OF VALIDATOR TEAM ASSESSMENT

Team of Validator	Assessed Aspect	Score	Score (Revised)	Category (Revised)
Validator 1	Content Feasibility	3.33	4	Very Feasible
Validator 2	Construct Feasibility	3.44	4	Very Feasible
Validator 3	Language Feasibility	3.75	4	Very Feasible
Total Value		12		
Average		4		
		Very Feasible		

IV. DISCUSSION

The parameters used to obtain good quality questions, besides fulfilling validity and reliability, are the balance of the difficulty level of the problem. A good question is a question that is not too easy or not too difficult (Siswanto, 2017). Questions that are too easy do not stimulate students to enhance their efforts to solve them. Conversely the problem that is too difficult will cause students to become discouraged and not have the enthusiasm to try again, because it is out of reach. Difficulty levels can be defined as the proportion of test takers who answer correctly. The index of difficulty is between 0.00 and 1.0. This difficulty index shows the level of difficulty of the question. Problems with difficulty index 0.0 indicate that the problem is too difficult, whereas the index 1.0 indicates that the problem is too easy (Arikunto, 2012).

Based on the above theory it can be concluded that the cognitive level and difficulty level of the problem are different things. Cognitive level refers to the level of reasoning used, so we know the term HOTS (high-order thinking skills) and LOTS (low-order thinking skills). In Bloom's theory, HOTS is a skill that requires high-level reasoning, which is generally associated with levels C4, C5 and C6 while LOTS, are generally aligned with C1 and C2. While the level of difficulty of the problem or the level of difficulty is a number that shows difficult and easy questions indicated by what percentage of all test participants answered the question correctly. The level of difficulty of the problem is associated with the complexity of the workmanship of the question, although it actually does not require high-level reasoning (HOTS).

The teacher can make complicated Mathematical questions even if only the C3 level. Multiplication, division with large numbers or complicated fractions will make the problem difficult to solve. So the problem has a high level of difficulty, but the cognitive level is low. Conversely, if the teacher presents illustrated questions with several glass beakers that have different concentrations of solvents and contain a plant cell then ask which one is the faster the osmosis process. To

answer that question, analytical skills are needed, that the osmosis event correlates with the concentration of the solution and the extent of the spreading area. If the logic is mastered by students, it is easy to answer it. Conversely, if the analytical ability is not owned by students, it will be difficult to do it. So the problem has a low level of difficulty but a high cognitive level.

Likewise with the following studies: (1) Budiman & Jailani (2014) in the Mathematics Education Research Journal proves that: 1. the final product in this study produced HOTS assessment instruments to measure the high-order thinking skills of mathematics in eighth grade students of Junior High School, 2. the instrument is feasible to use and it is obtained that the multiple choice questions have a moderate level of difficulty, good distinguishing power, all the distractors function well, and the problem description has a moderate level of difficulty with a good distinguishing power. (2) Based on the results of Nofiana's research, *et al.* (2014) in the Inquiry Journal, it can be concluded that the evaluation instrument of development results has the difficulty level of the questions with the medium category, which is 15% easy: 80% moderate: 5% difficult. In which the level of problem difficulty is calculated through the proportion of students' overall answers. (3) It is even inversely proportional to the results of Shidiq, *et al.* (2014) where in his research the development of two-tier multiple choice instruments to measure students' high-order thinking skills is known that the questions used are moderate difficulty categories but more than half of the results based on research results of the total number of students having very low level of thinking that is as much as 55.4% students think very low, and as many as 44.6% of students have a medium-high level of thinking.

Likewise in this study, the results of the two trials carried out were individual trials and small group tests, there were several items that were revised even though the power of the test was still in the good category had not reached a very good category, because in this study the main test package requirements were developed can be used if it meets valid and reliable criteria. Anderson (*in* Arikunto, 2012) states that the requirements that are very important for tests, namely validity and reliability. If the two conditions are met, then the package of high-order thinking skill tests can be used because it can measure what is to be measured, namely the students' high-order thinking skills and provide the data which is in line with the reality. As Arikunto (2012) stated, good instruments are instruments that can provide data that is consistent with reality.

V. CONCLUSION

Based on the objectives, results, and discussion from the research development of the two-tier multiple choice instrument for high school students, it can be concluded that (1) the results of the material/content expert validation of the two tier multiple choice items developed are included in the feasible category, (2) the results of the construct expert validation of the two tier multiple choice items developed are included in the feasible category, and (3) the results of the

language expert validation of the two tier multiple choice items developed are included in the very feasible category.

ACKNOWLEDGMENTS

Thank you for my head of study program, students and teachers who helped me a lot in completing this article and also principal who has allowed the author to do research at SMA Negeri 15 Medan that helped me a lot in collecting data for the completion of this article.

REFERENCES

- [1] Al-Mahbasi, A. A., Noor, N. M & Amir, Z, *The Effect Of Multiple Intelligences on DDL Vocabulary Learning*, (International Journal of Applied Linguistics and English Literature, 2017), pp.182-191.
- [2] Amalia, F. R, *Pengembangan Multiple Intelligences Siswa Oleh Guru Melalui Kegiatan Pembelajaran Pendidikan Agama Islam*. Skripsi, UIN Sunan Kali Jaga, Yogyakarta, 2015.
- [3] Andini, Y, *Effect Of Application of Multiple Intelligences Multiple Learning Models to Mathematical Learning Results*. (Prosiding Diskusi Panel, 2014), pp.1-8.
- [4] Armstrong, T, 2009, *Multiple Intelligences in The Classroom 3rd Edition*, USA : ASCD.
- [5] Bire A, L. & Bire, J, *Pengaruh Gaya Belajar Visual, Auditorial, Dan Kinestetik Terhadap Prestasi Belajar Siswa*, (Jurnal Kependidikan, 2014) pp. 158 – 174.
- [6] Depdiknas. 2006. Petunjuk Teknis Pengembangan Silabus dan Contoh/ Model Silabus SMA/MA. Jakarta: Departemen Pendidikan Nasional.
- [7] Fauzi, A and Farichah, E, *The Implementation of Multiple Intelligences on Learning Process at SMP IT and SD IT Madani Islamic School of Cilegon Banten Province Indonesia*, (International of Journal Home Science, Banten Indonesia, 2016), pp. 50-57.
- [8] Gangadevi, R, *Multiple Intelligence Based Curriculum To Enhance Inclusive Education To Bring Out Human Potential*, (International Journal of Advanced Research, 2016) pp. 619-626.
- [9] Gardner, H. & Hatch, T, *Multiple Intelligences Go To School: Educational Implications Of The Theory Of Multiple Intelligences*, (Journal American Educational Research Association, 1989) pp. 4-10.
- [10] Komala & Asri, C, *Efektivitas Program Pembelajaran Dengan Menggunakan Permainan Tradisional Untuk Mengembangkan Kecerdasan Majemuk Anak Usia Dini*. (Jurnal Ilmiah UPT P2M STKIP Siliwangi, 2016) pp. 1-8.
- [11] Kotsis, A., Kiss, M & Kun, A. I, *The Relationship Between Intelligence, Emotional Intelligence, Personality Styles and Academic Success*, (Journal Business Education & Accreditation, 2014) pp. 23-34.
- [12] Koura, A.A & Al-Hebaishi, M, *The Relationship Between Multiple Intelligences, Self Efficacy And Academic Achievement of Saudi Gifted And Regular Intermediated Students*, (Journal Educational Research International, 2014) pp. 48-70.
- [13] Kuncel, N. R., Hazlet, S. A and Ones, D. S, *Academic Performance, Career Potential, Creativity, and Job Performance: Can One Construct Predict Them All?*, (Journal of Personality and Social Psychology, 2004), pp. 148-161.
- [14] Lesmana, F. 2016. *Pengaruh Kecerdasan Naturalis terhadap Hasil Belajar IPA Materi Tumbuhan dan Hewan di Al-Khoiriyah 2 Semarang*. Skripsi, FTIK, UIN Walisongo : Semarang.
- [15] Lukman, A and Patahuddin, *Hubungan Intelegensi Jamak (Multiple Intelligences) Dengan Hasil Belajar Ekonomi Akuntansi Ranah Kognitif Siswa SMA Negeri 1 Watan Soppeng*, (Jurnal Penelitian dan Evaluasi Pendidikan, 2005), pp. 1-9.
- [16] Lwin, M, Khoo, A, Lyen, K & Sim, C. 2004, *How to Multiply Your Child's Intelligence* : (Pener. Christine Sujana) (Cara Mengembangkan Berbagai Komponen Kecerdasan). Indonesia : PT. Indeks
- [17] Maftuh, *Intelegensi Sebagai Faktor Belajar*, (Jurnal MIYAH, 2015) pp. 168-179.
- [18] Mahasneh, A.M, *The Relationship Between Multiple Intelligence And Self Efficacy Among Sample of Hasemithe University Students*, (International Journal of Education and Research, 2013), pp. 11-12.
- [19] Morgan, J.A & Carmen, F, *Multiple Intelligences Theory And Foreign Language Learning : A Brain-Based Perspective*, (International Journal of English Studies, 2004) pp. 119-136.
- [20] Murdiyani, I, *Pembelajaran Biologi Menggunakan Metode E-Learning Berbasis Multiple Intelligences Pada Materi Sistem Gerak Manusia*. (Innovative Journal of Curriculum and Educational Technology, 2012) pp. 45-52.
- [21] Murray, S & Kylie, M, *Inclusion Trough Multiple Intelligences*, (Journal of Student Engagement: Education Matters, 2012) pp. : 42-48.
- [22] Pratama, A. T., Syamsuri, I, Adi, and Aloysius, D. C, *The Contribution of Intelligence Quotient (IQ) on Biology Academic Achievement of Senior High School Students In Medan, Indonesia* (International Journal of Educational Policy Research and Review, 2015) pp.141-147.
- [23] Purwatiningsih, "Pembelajaran Biologi Berbasis Multiple Intelegensi Berpendekatan Observation Based Learning", Prosiding Seminar Nasional Pendidikan Biologi 2015, FKIP Univ. Muhammadiyah Malang, 2015.
- [24] Richards, *The Integration Of The Multiple Intelligence Theory Into The Early Childhood Curriculum*, (American Journal of Educational Research, 2016) pp. 1096-1099.
- [25] Samsinar, *Korelasi Strategi Multiple Intelligences Dengan Kualitas Pembelajaran Pendidikan Agama Islam Pada SMPN Di Watampone*, (Jurnal Lentera Pendidikan, 2014) pp. 58-71.
- [26] Sholiha, I., Karyanto, P., Sugiharto, B, *Kekuatan Dan Arah Kemampuan Metakognisi, Kecerdasan Verbal, Dan Kecerdasan Interpersonal Hubungan Dengan Hasil Belajar Biologi Siswa Kelas XI IPA SMA Negeri 3 Sukoharjo*. (Jurnal Pendidikan Biologi, 2012) pp. 31-39.
- [27] Spirovska, E, *Integrating Multiple Intelligences In Teaching English As A Foreign Language-Seeu Experiences And Practices*, (Journal Versitas, 2013)pp. 1-12.
- [28] Uzho & Salame. 2014. *Use of Bodily-Kinesthetic Intelligence Theory to Improve the Creative Learning of the English Language in Students of 1st EGB at Undidad Educativa Bilingue Instituto Particular Abdon Caderon*. Research Project, Educativa Bilingue Instituto Particular Abdon Caderon : Ecuador.
- [29] Zohar, D, *Spiritual Intelligent Leadership*. No.38, 2005, ocw.knu.edu.tw/sysdata/doc/0/0615c5ca0e84309a/pdf.pdf, (28September 2017)