

Dear Jusniar

Subject:

“Implementation of two-tier Multiple Choice Test to Assess Indonesian Elementary Students’ Higher-Order Thinking Skills in Science Learning”

We have identified you as one of the most suitable people world-wide to review the above manuscript. We hope that you will find this manuscript of interest. If you want we can send you a review certificate (If you would like a review certificate, **please specify this** when submitting your review). Thank you very much for your consideration and contributions. **Please could you confirm the receipt of this e-mail?**

Note: If it is possible, could you please send your report within two weeks? (Could you please send the files without converting them to PDF)

Sincerely yours,

Editorial

International Journal of Instruction

Implementation of two-tier Multiple Choice Test to Assess Indonesian Elementary Students' Higher-Order Thinking Skills in Science Learning

Assessten of higher-order thinking skills provide few opportunities for students to develop more in-depth knowledge, serving students' ability to identify and solve their own problems. Two-tier multiple choice test (TTMCT) is seen as an alternative instrument to measure HOTS objectively. This study aims to apply TTMCT to measure HOTS in Indonesian elementary school on learning science in concept of force, motion, and energy. This study used mixed method research using concurrent triangulation strategy with quantitative data collection technique using test and survey, while, qualitative data using interview and observation involving 227 students, 65 teachers, 5 principals. The results of study show that TTMCT applicable to Indonesian Elementary School with high and medium criteria. The results of the assessment using two-tier multiple-choice test can be a simple representation of the meaningfulness of teaching and learning in the classroom. The conclusions is the results of using TTMCT can be a simple representation of the meaningful teaching and learning process in a classroom and can be presented in the individual profiles form in regard of students' thinking skills mapping at the level of analyzing, evaluating, and creating. It would be better if the two-tier multiple-choice test instrument was developed not only as an assessment of high-level thinking skills but also as a diagnostic test of learning difficulties and student misconceptions in elementary schools.

Keywords: higher-order thinking skills, two-tier multiple choice test, elementary school, science learning

INTRODUCTION

The Competency Standards of Elementary School Graduates suggests that each student is expected to build and apply information or knowledge in a logical, critical, creative, and innovative ways; demonstrate the ability to think logically, critically, creatively, and innovatively in decision making; as well as demonstrate the ability to analyze and solve complex problems (Minister of National Education Regulation No. 54 of 2013). Higher-order thinking skills (HOTS) improves students' reading, writing, speaking and listening skills; in addition to increase the likelihood of providing better reasons in all subjects; support correct decision-making and problem solving; establish critical analysis and conclude and assess students' emotions; as well as help students to making smart choices in relationships with other fellow humans (O'Dowd's, 2007). Thus, it was important to develop high-order thinking skills in schools since each graduate needs to have competence in seeking for alternative resolution to the problems faced.

Higher-order thinking is effortful and depends on self-regulation. HOTS involves a cluster of elaborative mental activities requiring nuanced judgement and analysis of complex situations according to multiple criteria or find possible answer in perplexing situation (Resnick, 1987; Lewis & Smith, 1993). Similarly, Kings, et al., describes higher-order thinking which encompasses a variety of thinking processes applied to complex situations and reflects various variables (2010). HOTS includes critical, logical, reflective, metacognitive, and creative thinking. Brookhart (2010) classified

Commented [A1]: Compare with method and research results

Commented [A2]: incorrectly

Commented [A3]: ?

Commented [A4]: what mean

Commented [A5]: adjust to the focus of the study

Commented [A6]: ?

Commented [A7]: adjust or check with the references

HOTS into three contexts of understanding, includes (1) higher-order thinking as a transfer (students can apply their own knowledge and skills which they can further develop into a new context); (2) higher-order thinking as critical thinking (express self-reasoning, responding, and decision making without teacher's intervention); and (3) higher-order thinking as problem solving (serving students' ability to identify and solve their own problems in the work and daily lives).

The concept of higher-order thinking skills derived from Bloom's Taxonomy (1956). There are six orders of Bloom's Taxonomy, consist of memorizing (C1), understanding (C2), applying (C3), analyzing (C4), evaluating (C5), and creating (C6). Bloom's Taxonomy classified thinking skills into higher and lower order thinking skills. Memorizing, understanding, and applying as lower-order thinking skills and analyzing, evaluating, and creating as higher-order thinking skills (Afandi & Sajidan, 2017; Ball and Garton, 2005; Anderson & Krathwohl, 2002; Airasian & Miranda, 2002). Feedback and assessment of the learning process and the existing formative assessments provide few opportunities for students to develop more in-depth knowledge (Limbach & Waugh, 2010; Cullinane, 2011). The development of formative assessment alternatives are needed to help students empowering higher-order thinking skills.

Commented [A8]: need reference

Commented [A9]: check with the references

Commented [A10]: development or implementation

Evaluation is a systematic process determining the extent to which instructional objectives are achieved by students which reflects procedures for obtaining information on student learning (Purwanto, 2010, Linn & Grondlund, 2000). Teachers should be able to choose appropriate assessment procedures to make learning decisions and use assessment results to make educational decisions (Kusaeri & Suprananto, 2012). Thus, the assessment should be well implemented, as assessment is major component of student personal development for personal students and classroom. Assessment of higher-order thinking skills can improve students' achievement and motivation (Brookhart; 2010).

The selected response and short answer is one of method in learning assesment. Students to choose the most correct answer among the already provided alternatives (Kemendiknas, 2011; Stiggins, 2004, Thorndike & Hagen, 1977). The selected-response assessment uses a scoring technique that calculates the proportion of right and wrong answers to learners. In this study, the type of assessment developed is the selected response assessment, where this assessment has an objective nature. The multiple choice item is one of the most widely applicable test items for measuring achievement (Linn & Gronlund, 2000). Multiple-choice test are comprehensive, objective scoring and easy checking, in addition to high item reliability, it can measure different levels of ability including higher-order thinking skills, the type of item can be arranged in such way that requires the ability of the test participants to distinguish different index of truth at once, it grains difficulty level that can be set by simply changing the homogeneity of alternative answers, and the information related to students' thinking skills can be more translated for teachers.

Commented [A11]: check to references

Assessment procedures must be provide oportunites to students and teachers to engange in discussions on the assessed works (Cullinane, 2011). An alternative assessment that can be developed is a modified multiple-choice question form of two-

Commented [A12]: ?

tier multiple-choice test. Two-tier multiple-choice test (TTMCT) is modifications of multiple-choice form which belong to kind of objective test. TTMCT developed consists of two levels of questions, the first tier is the content of the main question or item that has two answer choices and the second tier is the reason for the answer given on the basis of first tier. The existence of reasons at the second tier aims to improve thinking skills and see students' ability to reason (Cullinane, 2011; Treagust, 2006). TTMCT was applicable as an alternative formative assessment, to assess students' understanding, asking student to use higher-order thinking skills in giving reasons in second tier, and identify misconception that students may have, and (Halaydina & Downing, 1989; Treagust, 1995; Treagust, 2006; Sampson, 2006; Cullinane, 2011). TTMCT can be used as an insight into making a form of assessment that challenges students' knowledge, providing a technique to assess students' concepts especially in classroom learning.

The observation results in Indonesian elementary schools indicated that most teachers have shared positive perceptions and being aware of the importance of higher-order thinking skills in Elementary Schools, however, it remains difficult for teachers in establishing assessment instruments which are applicable to measuring students' HOTS. Among the difficulties faced by teachers are: 1) the difficulty in developing assessment which not only measures lower-order thinking skills but also the higher-order thinking skills; 2) higher-order thinking skills would be better measured using objective tests, such as multiple-choice tests; 3) it is most often found that teachers use multiple-choice test, however, they also realize on the difficulty in establishing distractors or effective deceiving tests; 4) the items of multiple-choice test also has a limitness. It is not able to distinguish which students answer earnestly involves higher-order thinking skills and which students answer based on guesswork. The data found in the field showed the higher-order thinking skills test are quite rarely found in the teachers' test items bank, both formative and sumative assessments.

For such elaborated matter, this study attempts to implement higher-order thinking skills in two-tier multiple-choice test forms. This study aims at conforming both quantitative and qualitative analysis towards implementing the two-tier multiple-choice test. The already developed tests will be based on the assessment test of the higher-order thinking skills in elementary schools.

METHOD

This study is a mixed method research using concurrent triangulation strategy. The researchers collect quantitative and qualitative data concurrently (simultaneously), then compare these two databases to determine convergence, differences, or some combination found (Creswell, 2012). In this strategy, quantitative and qualitative data collection is are conducted simultaneously within a single research stage. Quantitative data were collected from the survey results of the assessment needs of 49 teachers and the results of the validity, reliability, and item analysis of the two-tier multiple-choice test instrument involving 227 students. Qualitative data about implementation of two-tier multiple-choice test was obtained from interview with 16 teachers and 5 principals and also observations in 5 different Elementary Schools in Purbalingga Regency,

Commented [A13]: give a clearly reasons or argument based on previously research, why the two-tier test can improve the students' HOTS

Commented [A14]: the aims of study must be clearly

Commented [A15]: ?

Central Java, Indonesia. The schools involved are schools that are categorized as high, medium, and low criteria.

Data collection techniques used questionnaire and test. The questionnaire was conducted to find out the needs and problems in the assessment of HOTS in elementary schools. Questionnaire results are used to investigate the testing and feasibility of TTMCT. Test used to find out the TTMCT validity, reliability, and item analysis. The validity test used is the content validity test based on Aikens' Formula involving experts in Basic Education. The test aims to analyze the ability of each item question in the two-tier multiple-choice test to measure indicators of higher-order thinking skills. Reliability testing and item analysis is done by entering the data of students work into Iteman 3.0.

Commented [A16]: provide clear and consistent information for each instrument and its use in this study.

Qualitative data collection techniques were conducted with in-depth interviews and observations. Interviews were conducted involving teachers and principals, involved consisted of 16 Elementary Schools teachers who had at least 10 years of teaching experience in higher class, and a principal who has master education degree in education. The purpose of in-depth interviews with teachers and principals is to find out about the teachers' responses to the practical implications of implementing two-tier multiple-choice tests. While the observation technique performed during pre-implementation until the implementation process of tests instruments to know the advantages and disadvantages of test instrument. Qualitative data analysis is done by interactive analysis with data collection step, data reduction, display data, and conclusion (Miles and Huberman, 2007).

TTMCT scoring is not too much different with scoring on the multiple-choice test which refers to the correct answer in the first tier and correct answer in the second tier (Adesoji & Omilani, 2012). TTMCT scoring in this study refer to Yamtinah (2015) can be seen in Table 1.

Table 1
Scoring Two-Tier Multiple-Choice Test

1 st Tier (Answer)	2 nd Tier (Reason)	Score
Correct	Correct	3
Correct	Incorrect	2
Incorrect	Correct	1
Incorrect	Incorrect	0

Commented [A17]: see the IJI guided

The scoring based on table 1 provides a different scoring process of the students who answered incorrectly on the first tier but answered correctly on the second tier and answered wrongly on both tier. The item analysis is performed using Microsoft Excel 2010 and Iteman 3.0 software.

RESULTS AND DISCUSSION

Development of Test Instruments

In developing two-tier multiple-choice test instrument to measure higher-order thinking skills requires certain criteria. Two-tier multiple-choice test is a form of multiple-choice

Commented [A18]: see IJI guided for author

questions that not only requires the students to choose several option given as answers, but also to choose the reason for the answer which has been selected in first tier. Inclusion of reasons for answers in this form aims to improve students' thinking ability.

Commented [A19]: implement or development.

The development of two-tier multiple-choice test used in this study aims to measure the success of achieving cognitive indicators in higher-order thinking skills developed by Anderson & Krathwoh (2002) and Airasian & Miranda (2002) that covering the skills of analyzing, evaluating, and creating. Stem writing questions adapted to operational verbs that represent cognitive higher-order thinking skills. The reasoning as a second tier of the question is given directly under the question.

The two-tier multiple-choice test instruments that has been developed consists of 25 items covering 3 competency indicators for 5th grade students at elementary school such as: 1) identify the gravity force, frictional force, and magnetic force and its utilization in everyday life; 2) analyze the relationship between force, motion, and energy; 3) formulate problem solving related to gravity force, friction force, and magnetic force or simple machine. Then, the three indicators of competence are described into 25 indicators of test forms which combine operational verbs of higher-order thinking skills as follows: 1) identifying, analyzing, describing, and defining features (C4); 2) criticizing, clarifying, and interpreting (C5); 3) making generalization, connecting hypothesis, predicting, and proposing hypothesis (C6).

Commented [A20]: Whether the force, friction, and energy represent science learning in elementary school? .

Two-tier multiple-choice test emphasizes the higher-order thinking skills of C4, C5, and C6, for that reason, if such percentage is converted, it will get 50% about C4 8 (analysis), 25% about C5 (evaluation), and 25% about C6 (creating) by comarison C4:C5:C6 equals to 2:1:1. The two-tier multiple-choice test aims to connect the multiple-choice test weakness used by teachers in schools turns to be an alternative instrument to measure higher-order thinking skills. It has objective, valid, reliable, good difficulty index, good different strenght. It also receives good responses from students and teachers and can show individual profiles to higher-order thinking skills of each student.

Commented [A21]: ??

Two-tier Multiple-Choice Test Feasibility

The developed two-tier multiple-choice test instrument was tested using content validity test by eight validators consisting of five Lecturers in the Elementary School Teacher Education Study Program and three teachers who have had more then ten years teaching experience. The content validity procedure used in this analysis is the Aiken's V formula for calculating content coefficient validity based on the results of validator assessment on each item to know how exactly the item describes the indicator being measured (Azwar, 2012). On the table of coefficient validity with 8 validator and 4 rating scale, the item is valid if its coefficient validity is $\geq 0,75$. Test results show 17 valid items without having required revision while 8 item are valid after the revision.

Commented [A22]: Where the research data

The two-tier multiple-choice test tool analyzed the suitability of its application separately based on students' ability level using Item 3.0. out of 100 students who take on the two-tier multiple-choice test in operational field testing, 30 students were came from predefined "high ability" Elementary School, 40 students from predefined

Commented [A23]: Showed the summary of results of the content validity

“medium-ability” elementary school, and 30 students from predefined “low ability” elementary school. School predicate is determined based on accreditation and National Examination rank in Purbalingga Regency in the academic year 2016/2017. It aims to know the two-tier multiple-choice test whether or not it is applicable to schools with appropriate categories.

Commented [A24]: Adjust with abstract and method

Table 2
Analysis result items based on the elementary school’s criteria

Analysis		School Criteria					
		High-Ability		Medium-Ability		Low-Ability	
		Result	Interpretation	Result	Interpretation	Result	Interpretation
Reliability	Answer (1st tier)	0,892	Reliable	0,788	Reliable	0,527	Not Reliable
	Reason (2nd tier)	0,903	Reliable	0,801	Reliable	0,709	Not Reliable
Mean P	Answer (1st tier)	0,428	Fairly Difficult	0,369	Fairly Difficult	0,198	Difficult
	Reason (2nd tier)	0,492	Fairly Difficult	0,449	Fairly Difficult	0,264	Fairly Difficult
Mean Item-Tot	Answer (1st tier)	0,525	Excellent	0,380	Good	0,282	Fairly enough
	Reason (2nd tier)	0,542	Excellent	0,564	Excellent	0,344	Good

The reliability coefficient, difficulty index (Mean P), and determination power (Mean Item-Tot) were analyzed separately based on school criteria and it get different and significant results. The test instrument is reliable when applied to students in schools with “high” and “medium” criteria, while for schools with “low” criteria it is “unreliable” to apply. Judging from the mean of the difficulty index (Mean P) value, for schools with “high” and “medium” criteria, the test instrument has a “Fairly Difficult” interpretation, while for “low” schools criteria, the test instrumen has a “difficult” interpretation for the answer and “quite difficult” for the reason. Based on Mean Item-Tot (Determination Index) two-tier multiple-choice test has a “Excellent” interpretation, however, on different school with “low” criteria it has a “Fairly Good” interpretation.

Table 3
The recommendation of two-tier multiple-choice test implementation

No.	School Criteria	Feasibility (Implementation)			Conclusion
		Reliability	Difficulty Index	Different Strength	
1.	High	√	√	√	applicable
2.	Medium	√	√	√	applicable
3.	Low	-	-	√	inapplicable

Teachers’ Responses

Teacher’s responses related to the two-tier multiple-choice test instrument are identifiable as follows: 1) Teachers are well-received development of two-tier multiple-choice test instrumen although some adjustments are still needed; 2) developing questions which measure the high-level thinking skills, this will lead to providing assessment for students’ concepts of understanding and identifying students’ learning meaningfulness as well as their ability to relate the received materials according to the contextual surrounding environment; 3) being a more functional measuring alternative

of multiple choice in general since it reduces students' chance in guessing the tests answers and simultaneously measures their ability in understanding the concept and its relation to the environment; 4) applicable through taking into account the learning methods given to the students, the scope of material, and the students' sufficient average ability if only requiring optimal results.

The teacher's response to the two-tier multiple-choice test was in line with Sampson's (2006) study. Sampson (2006) notes the advantages of using a two-tier multiple-choice test for teachers as an insight to create assessment form which challenges students' knowledge and provides a way to assess their concepts of understanding. Treagust in his study (1995) mentioned that two-tier multiple-choice tests are more effective for determining students' alternative concepts on the tested material and applicable to see whether or not a learning activity is meaningful.

Other studies in line with the results of this study are Tuysuz's (2009) research which has proved that the form of two-tier multiple-choice test helps teachers to teach better in addition to help students to gain a better learning. The two-tier multiple-choice test is easily applicable for teachers to improve students' knowledge and maintain students' alternative understanding to be compared with the usual multiple-choice test form.

The Implementation of TMCT

The two-tier multiple-choice test implementation in this study yielded several findings related to both advantages and disadvantages of the two-tier multiple-choice test instruments, all of which concern the previous studies and those obtained from field observations. The advantages of two-tier multiple-choice test among others to measure the level of high-order thinking skills (analysis, evaluation, and creation) which are commonly difficult carried out by common double choice (Cullinane, 2011; Halaydina dan Downing, 1989; Treagust, 2006); scoring becomes easy, fast, and objective, in addition to be applicable to determining the teachers' learning effectiveness (Cullinane, 2011); applicable to measuring both problem-solving skills and critical thinking (O'Dowd, 2007); applicable to diagnose material understanding and detect possible misconceptions that students can make (Cullinane, 2011; Sampson, 2006).

In addition to the advantages of two-tier multiple-choice test instruments which shared similar report with the previous studies, the researchers also explored new result findings of two-tier multiple-choice test analysis in the field.

First, the material scope contributing to the tests is relatively more than that material arranged into the description tests. The two-tier multiple-choice test is sort of modified multiple-choice test, however, both share similar advantages as the multiple-choice test found in common. The observation results showed that in the developing process of the two-tier multiple-choice test, it enables the researchers to more freely determine the material scope at once along with the skills indicator to be tested according to the assessment needs. This is in line with an opinion which suggests multiple-choice tests within a short period of time can effectively cover broad material and numerous test items (Susetyo, 2015, p.13). It can effectively measure various types of knowledge and complex learning outcomes (Linn & Gronlund, 2000). Thus it enables the two-tier

Commented [A25]: Where the results of their implementation TMCT

Commented [A26]: where the discussion about the implementation of TMCT and its relationship with HOTS

multiple-choice test instruments play a role as a more effective formative and summative assessments instead of the other assessments.

Second, the relative tests reliability is higher than the description. The quality test items is good; for they can be analyzed empirically, including validity and reliability (Susetyo, 2015, p.13). A good reliability coefficient is above 0.75 (Susetyo, 2015). The reliability of a two-tier multiple-choice test instrument can reach a figure above 0.80 in main field testing and achieve 0.90 in field testing operations. Hopkins and Antes (1989) further argue that reliability refers to a consistent observations obtained from recurrent records on both a subject and a number of subjects. Linn & Gronlund (2000) note "the multiple choice item is one of the most widely applicable test items for measuring achievement". Multiple-choice tests have advantages including: comprehensive, objective scoring and easy checking; good test items quality (high reliability test items). A measuring device is reliable, in case the test results are unchanged or relatively share similarities when repeated testing is done. This measuring instruments are, therefore, called reliable. Reliability is a coefficient that shows the consistency of measurement results of a test (Mardapi, 2012). Thus the two-tier multiple-choice test reflects the nature of reliability and consistency as an instrument of students' skill test.

Third, chances to guess the answers are smaller since between the choice of answers and the choice of reasons are interrelated. On multiple-choice tests, some students have the opportunity to guess answers without paying a lot of attention to questions or answer options. In the scoring process, teachers normally unable to distinguish which students answering questions based on good conceptual thinking and understanding and which students answering questions by guessing. However, in a two-tier multiple-choice test, in addition to choosing an answer to a question, students are also to provide appropriate reasons. Thus, students' understanding measurable through looking at the pattern of conformity between the choices of answering and reasoning. This is in accordance with the results of Cullinane's study which elaborates that the use of reasoning on the second level aims to improve thinking skills and pay a close attention to students' ability in giving reasons (2011).

Fourth, the two-tier multiple-choice test instrument provides a challenge for both teachers and students to engage in a higher-level assessing process. Students are more motivated to explore subject matters since these typical test forms are more challenging to their thinking ability, while the teachers can develop their skills strategically in arranging the student's skill pattern and linking it to establish answer and reason choices in the 1st tier and 2nd tier options. This is in accordance with Sampson's (2006) and Treagust's (1985) results of observation.

The two-tier multiple-choice test also reflect some drawbacks in referring to the field observations. First, a two-tier multiple-choice test instrument cannot be used to measure verbal capabilities similar to that of Linn & Gronlund's (2000). Thus, when teachers apply a two-tier multiple-choice test to assess students' thinking skills comprehensively, similarly teachers need other assessment alternatives to assess students' verbal skills too. Second, in preparing a good two-tier multiple-choice test requires a relatively longer time rather than the other. The multiple-choice types preparation takes a lot of time and

Commented [A27]: adjust to the focus of this study

Commented [A28]: where is the discussion about HOTS and science learning?

effort (Susetyo, 2015). Third, learning practice in Indonesian elementary schools' level, both students and teachers have not yet accustomed to using two-tier multiple-choice test assessment. The two-tier multiple-choice test implementation needs to be well-prepared since at the stage of preparing instructional planning, learning indicators and assessment indicators, model selection and learning methods, and formative authentic assessment which apply to stimulate students' thinking skills. Teachers are not yet accustomed to compiling and using two-tier multiple-choice tests, therefore, guidance, development, and direction are required if a similar assessment needs to be applied.

Once it is implemented in Elementary Schools, the two-tier multiple-choice test provides several practical impacts and implications, including: 1) the assessment results using two-tier multiple-choice test applicable as material for evaluation and follow-up in an individual profile form of mapping high-ability thinking skills from several activities such as analyzing, evaluating, and creating; 2) applying two-tier multiple-choice test can be a simple representation of the meaningful teaching and learning processes within a classroom as well as showing an undergoing learning effectiveness. The representation will be more illustrated if the test instrument is used as a training medium to streamline the learning objectives; 3) a two-tier multiple-choice test applicable as a basis for developing schools' teaching materials which encompass what appropriate teaching materials should be provided to student's needs, what materials should be deepened, not only at the knowledge level but also at technical mastery level; 4) the result of two-tier multiple-choice test is one of the alternatives in collecting data on the basis of follow-up planning in solving efforts form according to the already identified problems or difficulties after applying test instruments, thus at least it not only measures students' high-order thinking skills but also a diagnostic test of learning difficulties and misconceptions towards a concept.

CONCLUSION

The conclusions is the results of using TTMCT can be a simple representation of the meaningful teaching and learning process in a classroom and can be presented in the individual profiles form in regard of students' thinking skills mapping at the level of analyzing, evaluating, and creating. It would be better if the two-tier multiple-choice test instrument was developed not only as an assessment of high-level thinking skills but also as a diagnostic test of learning difficulties and student misconceptions in Elementary Schools. TTMCT requires relatively longer time rather than the other test materials, in the Indonesian contextual learning practice of elementary school, both students and teachers remain unaccustomed to using the two-tier multiple-choice test assessments. technically this approach requires further guidance and development.

Commented [A29]: basically from where

REFERENCES

- Adesoji & Omilani.(2012). A Comparison of Secondary Schools Students Levels of Conception of Qualitative and Qualitative Inorganic Analysis. *American Journal of Scientific and Industrial Research*, 3(2), 56-61.
- Airasian, P., & Miranda, H. (2002). The Role Assessment in The Revised Taxonomy. *Theory Into Practice*, 41(4), 249-254.

Anderson, L.W., & Krathwohl, D.R. (2010). *A Taxonomy for Learning Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objective A Bridged Edition: Addison Wesley Longman, Inc. 2001* (A. Prihantoro, Terj.). Yogyakarta: Pustaka Pelajar.

Azwar, S. (2015). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.

Azwar, S. (2015). *Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar*. Yogyakarta: Pustaka Pelajar.

Bloom, B.S. (1956). *Taxonomy of Educational Objective: The Classification of Educational Goals Handbook Cognitive Domain*. London: Longmans Inc.

Brookhart, S.M. (2010). *How to Assess Higher-Order Thinking Skills in Your Classroom*. Virginia : ASCD.

Cresswell, J.,W.(2012). *Research Design: Qualitative, Quantitative, and Mixed Methods*. Yogyakarta: Pustaka Pelajar.

Cullinane, A., & Maeve, L. (2010). *Two- Tier Multiple Choice Question: an Alternative Methods of Formative Assessment for First Year Undergraduate Biology Students*. Limerick: National Center for Excellence in Mathematics and Education Science Teaching and Learning (NCE-MSTL).

Grondlund, E., & Linn, R.L. (1990). *Measurement and Evaluation in Teaching*. Milan: Macmillan Publishin Company.

Halaydina, T.M., & Downing, S.M. (1989). A Taxonomy Multiple-Choice Item-Writing-Rules. *Applied Measurement in Education*, 2(1), 51-78.

Halaydina, T.M., Downing, S.M. & Rogriguez, T. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, 15(3), 309-334.

Hopkins, C.D., & Antes, R.L. (1989). *Classroom Testing: Construction*. F.E. Peacock Publisher.

King, F.J., Goodson, L., & Rohani, F. (2010). Higher-Order Thinking Skills: Definition, Teaching Strategies, and Assessment. Retrieved from www.caa.fsu.edu/files/higher_order_tinking_skills.pdf.

Kusaeri & Suprananto. (2012). *Education Measurement and Evaluation*. Yogyakarta: Graha Ilmu.

Kutluay, Y. (2005). *Diagnosis of Eleventh Grade Students' Misconceptions about Geometric Optic by A Three-Tier Test* (Master's Thesis, Medium East Technical University, 2005). Turkey: METU Library E-Theses.

Lewis, A., & Smith, D. (1993). Defining Higher-Order Thinking. *Theory Into Practice*, 32(3), 131-137.

Commented [A30]: see the IJI guided for Author

- Limbach, B. dan Waugh, W. (2010). Developing Higher-Order Thinking. *Journal of Instructional Paedagogies* 3 (1) 1-9. Retrieved from <http://www.aabri.com/jip.html>.
- Mardapi, D. (2012). *Education Measurement, Assessment, and Evaluation*. Yogyakarta: Nuha Medika.
- Miles, M.B., & Huberman, A.M. (2007). *Qualitative Data Analysis: References Book About New Methods*. Jakarta: Universitas Indonesia Press.
- O'Dowd, G. V. (2007). Developing Higher-Order Thinking Skills in Medical Students. *Hamamatu University School of Medicine* , 21, 21-33.
- Permendiknas Nomor 54 Tahun 2013 tentang Standar Kompetensi Lulusan Pendidikan Dasar dan Pendidikan Menengah.
- PPPPTK Matematika. (2011). *Pengembangan Instrumen Penilaian Pembelajaran Matematika SD/SMP*. Kementerian Pendidikan Nasional.
- Purwanto, N.. (2010). *Prinsip-prinsip dan Teknik Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya.
- Resnick, L. B. (1987). *Education and Learning to Think*. Washington DC: National Academy Press.
- Sampson, V. (2006). *Two-tier Assessment: Teacher Toolkit*. Arizona: College of Education at Arizona State University in Tempe.
- Stiggins, R.J, Arter, A., Chappuis, J., Chappuis S. (2004). *Classroom Assessment for Student Learning: Doing it Right -- Using it Well*. Assessment Training InSTITUTE.
- Susetyo, B.. (2015). *Prosedur Penyusunan dan Analisis Tes untuk Penilaian Hasil Belajar Bidang Kognitif*. Bandung: Refika Aditama.
- Thorndike, L. R., & Hagen, E. P. (1977). *Measurement and Evaluation in Psychology and Education*. John Wiley Publication.
- Treagust, D. F. (2006). Diagnostic Assessment in Science as A Means to Improving Teaching, Learning, and Retention. *Invited Speaker Paper in Universe Science Assessment Symposium Proceedings*, 1-12. The University of Sydney, 28 September 2006.
- Treagust, D.F. (1998). Development and Use Diagnostic Test to Evaluate Student's Misconception in Science. *International Journal of Science Education*, 10(2), 159-169.
- Tuysuz, C. (2009). Development of Two-Tier Diagnostic Instrument and Assess Students Understanding in Chemistry. *Scientific Research and Essay*, 4(6), 626-631.
- Yamtinah, S. (2015). Development of Diagnostic Instruments Learning Difficulties in Chemistry Learning in Higher-Education. *Jurnal Penelitian dan Evaluasi Pendidikan*, 19(1), 69-91.

Commented [A31]: add a new references



International Journal of Instruction

Certificate of Reviewing

Date: January 01, 2020

Ref. No.: IJI-20-1009

Subject: The review certificate

This is to certify that

Jusniar

Subject: “Implementation of two-tier Multiple Choice Test to Assess Indonesian Elementary Students’ Higher-Order Thinking Skills in Science Learning”

We received your review for this article on January 01, 2020. Thank you very much for your contributions.

Regards,

Prof. Asim Ari

Editor in Chief

p-ISSN: 1694-609X

e-ISSN: 1308-1470

<http://www.e-iji.net> & <http://www.gateacademy.ch>

E-mail: editor.eiji@gmail.com

IJI is indexing in ESCI, SCOPUS and ERIC





International Journal of Instruction Article Evaluation Form

Mr. /Mrs. Jusniar

It is to acknowledge you that the Executive Committee of *International Journal of Instruction* has decided that the article mentioned below would be reviewed by you. Thank you very much for your contributions.

Asim ARI
Editor in Chief

Name of the article: Implementation of two-tier Multiple Choice Test to Assess Indonesian Elementary Students' Higher-Order Thinking Skills in Science Learning

After reviewing the attached article, please read each item carefully and select the response that best reflects your opinion. To register your response, please **mark** or **type in** the appropriate block.

	Yes	Partially	No
Do you think the title is appropriate?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Does the abstract summarize the article clearly and effectively?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Are the objectives set clearly?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Is the issue stated clearly?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Is the literature review adequate?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Is the design of the research appropriate, and the exemplary, if any, suitable?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Is the methodology consistent with the practice?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Are the findings expressed clearly?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Is the presentation of the findings adequate and consistent?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Are the tables, if any, arranged well?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Are the conclusions and generalizations based on the findings?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Are the suggestions meaningful, valid, and based on the findings?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Are the references adequate?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Is the language clear and understandable?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Is cohesion achieved throughout the article?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Is the work contributing to the field?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

- Evaluation:**
- The article can be published as it is.
 - The article can be published after some revision.
 - The article must undergo a major revision before it can be resubmitted to the journal.
 - The article cannot be published.

Would you like to see the revised article if you have suggested any revisions? Yes No

Please write your report either on this paper or on a spare paper.

REPORT

Section of the Manuscript	Comments and Notes
Title- Abstract-Summary	This research whether developing or implementation two-tier test
Introduction and Literature Review	Need additional to the reason the authors use two-tiered instruments to improve students' HOTS compared to other instruments such as essay, multiple choice and others.
Research Methods	The method is not in accordance with the results of the study. somewhat confusing because the article title is the implementation of TTMCT, but on the presentation of the results the author begins with the development of the instrument and its feasibility.

Research Findings	Information about the instruments used in this study is unclear Display of research results is still lacking, especially on the validity
Discussion	An explanation of HOTS from the results of the implementation of TTMCT is still lacking. support data on student work and interview results has not been displayed.
Conclusion and Suggestions	adjust to the title and focus of the research
References and Citation	References need to be added in accordance with the fields of science and research in the last 5 years.
Language	Many words written incorrectly and some sentences are not simple or repeatable.
Other issues	Whether the force, friction, and energy represent science learning in elementary school? Because the title to asses HOTS in science learning.