Development of A Two-Tier Multiple-Choice Instrument Based on Higher Order Thinking Skills (HOTS) on Acids, Bases, and Salts

Arif Rahmat Hidayatullah[1]*, Sri Yamtinah[1], Mohammad Masykuri[1]

[1] Postgraduate Master's Program of Chemistry Education, Faculty of Teacher Training andEeducation, Sebelas Maret University, Surakarta, Indonesia.

Abstract: This study aims to measure higher order thinking skills on Acid, Base and Salt material. All students can think, but most of the students need encouragement and guidance for higher order thinking processes The test instrument needed in this research is to encourage students to think in higher order, namely the Two-tier Multiple-Choice test instrument that encourages students to think at higher levels. This research uses survey method. The sample in this study was 140 students from 3 junior high schools in Bima City, West Nusa Tenggara, Indonesia have high, medium, and low school averages. Based on the results of the research that has been carried out, the conclusions are for the reliability value of 0.725, the difficulty of the problem is 0.560, and the discriminating power of 0.229. In the measurement of students' higher order thinking skills, it was found in the Low category of 8.8%, moderate 29.9%, good 42.7% very good 17.6%. This research is expected to provide an alternative to measure students' higher order thinking skills by using the TTMC instrument based on Higher Order Thinking Skills.

Keywords: Assessment Instruments; Two-Tier Multiple-Choice; Higher Order Thinking Skills; Acids; Bases; Salts

Introduction

Education in schools is not solely on the mastery and understanding of scientific concepts, but also on improving students' thinking abilities and skills, especially higher-order thinking skills, namely critical thinking skills (Anggraeni & Sole, 2020). Critical thinking in science education is to improve students' thinking skills and at the same time prepare students to be successful in understanding science learning and applying it in everyday life (Gultom et al., 2021). high critical thinking skills by junior high school students, students will be able to achieve competency standards that have been set in the curriculum or to be achieved in the learning process (Hikmawati et al., 2021).

Instruments that can measure higher order thinking skills train students to think critically and solve problems as well as measure their performance by using the HOTs tool which includes three aspects in Bloom's taxonomy, namely analyzing, evaluating, and creating. High Order Thinking Skills (HOTS) defined as a transfer process from a problem then the problem is sought for a solution using critical thinking (Brookhart, 2010). Approach to knowledge on High Order Thinking Skills is very important to stimulate learning (Narayanan & Adithan, 2015). In Higher Order Thinking Skills in analyzing attitudes, one must have broad knowledge through reading (Nachiappan et al., 2018). Meanwhile, according to (Brookhart, 2010) Higher order thinking skills are categorized into 3

---

* Corresponding Author: rahmatarif565@gmail.com

parts, namely (1) as a form of transfer of learning outcomes, (2) as a form of critical thinking, and (3) as a problem solving process. An assessment framework was then used to categorize all the questions recommending the use of the Revised Bloom's Taxonomy framework in the categorization of assessment items. (Mitana, 2018).

Instruments that can measure students' ability in higher-order thinking with the help of HOTS-based test items. A good question is one that pays attention to the students' thinking ability. critical thinking skills included in the High Order Thinking Skills (HOTS) group, namely focusing on questions, analyzing arguments, considering what can be trusted, considering observation reports, comparing conclusions, determining conclusions, considering induction abilities, assessing, defining concepts, defining assumptions, and describe (Maharani, and Rakhmati, 2015). good analytical, evaluation, creative, logical and reasoning skills in solving all questions, so that they are classified as having high-level thinking skills with low levels (Kurniati et al., 2016).

The quality of hots-based questions must be in accordance with Bloom's taxonomy, which includes analysis, evaluation, creation, logic and reasoning so as to be able to measure students' higher-order thinking skills by looking at the characteristics of the items produced. The characteristics of a good item are HOTS questions in accordance with the KI-KD that are possible in the subjects they teach (Hartono et al., 2022).

Teachers' insight into global issues, skills in choosing a stimulus question, and the ability to choose the competencies to be tested are important aspects that must be considered by teachers, in order to produce quality questions (Taubah, 2019). Meanwhile, High Order Thinking Skill (HOTS) is applying the knowledge and skills that have been developed during learning in a new context (Hasyim & Andreina, 2019). HOTS-based assessment is a process of collecting and processing information to measure the competence of learners in the form of a combination of mastery of cognitive abilities with factual, conceptual, procedural, and metacognitive knowledge (Tune Sumar & Tune Sumar, 2020).

The use of instruments on questions that are able to develop higher-order thinking skills is by using multiple choice questions with reasons, namely the Two Tier Multiple Choice (TTMC) instrument. Two Tier Multiple Choice (TTMC) is a graded test instrument consisting of two levels, the first level is the main question and the second level is a choice of reasons that refers to the answers from the first level (Cengiz, 2009). The application of a two-level diagnostic test to measure students' understanding has three steps in its development, namely defining the boundaries of the test content, gathering information about students' misconceptions, and developing an instrument (Sesli & Kara, 2012).

TTMC can reduce the tendency of students to guess the answer because the questions at the first level are closely related to the second level (Nofiana et al., 2014). Two-tier Multiple Choice assessment instrument that has validity, reliability, level of difficulty, discriminatory power and a distracting index that meets the criteria as a good question and develops (Shidiq et al., 2015).

Based on the description above, the objectives of the research are " Development Of A Two-Tier Multiple-Choice Instrument Based On Higher Order Thinking Skills (Hots) On Acids, Bases, and Salts".

Method

This study aims to determine students' higher order thinking skills. This study uses the ADDIE model. The instrument used in this research is the development of a HOTS-based Two-Tier Multiple Choice instrument to measure students' creative thinking skills. This research was conducted in three schools, SMPN 2, SMPN 7 and SMPN 11, Bima, West Nusa Tenggara Indonesia, with a limited-scale trial subject of 140 students. Consisting of 3 schools, each school consisted of 46 students who had an average of high, medium and low grades. Instrument validation was carried out by 8 expert validators, namely material, discuss, evaluation, for the instrument validation stage using AIKEN'V. Validation on AIKEN'V validation signs as a reference for the validity of an instrument (R. Aiken, 1985). Validation is done to determine the feasibility of an instrument (Azwar, 2016) Validation is done to determine the feasibility of an instrument. The presentation of a grid of questions on the HOTS-based Two-tier Multiple-Choice instrument to measure students' creative thinking abilities presented in Table 1.

Table 1. Grid of Two-Tier Multiple Choice questions base on HOTS

No	Question Indicator	Question	Answer key and Reason	Reinforcement Dimension
2 4	Students can explain the properties around them in the form of acids,bases, and salts in everyday life	A can find salt in everyday life, one of which is table salt which is used as a food seasoning. In addition to table salt, examples of salt in everyday life..... A. Orange juice and shampoo B. Baking soda and batteries C. Pepsodend and Baking Soda D. Starfruit and soap To support the answer above, please choose the reason below..... A. Because citrate and calcite are examples of Salt B. Because baking soda and calcite are examples of salt C. Because potassium hydroxide and baking soda are examples of Salt D. Because Citrate and Calcium Hydroxide are examples of Salts	Answer B Reason B	C4
1	Students can classify which properties include acids, bases and salts.	One day Dina was in a hurry to go to school, so she forgot to eat breakfast. During the lesson, Dina felt pain in her stomach. To overcome this, what Dina must consume is A. Orange fruit B. Stomach medicine C. ORS D. Stomach pain medicine Based on the answers above, the most appropriate reason is... A. Helping digestion in the stomach B. Gives a feeling of fullness to the stomach C. Can neutralize stomach acid D. Overcoming bacteria in the stomach	Answer B Reason C	C5

Results and Discussion

The validation stage is carried out by 8 expert validators, namely language material experts, and evaluation, using the AIKEN 'V formula validation results using the AIKEN'V formula In the table of signs with 8 expert validators with a score of 0.75 (R. Aiken 1985). As a benchmark for the feasibility of the instrument. The assessment of the feasibility of the instrument includes 5 aspects. The results of instrument validation are presented in Table 2.

Table 2. Instrument Validation Result

Assessment Aspect	Average	Aiken Validity Coefficient	Description
Aspects of Science Competence	0,89	0,75	Valid
Science Subject Knowledge	082	0,75	Valid
question context	0,80	0,75	Valid
Spelling and Grammar	0.82	0,75	Valid
Clarity of Question Commands	0,85	0,75	Valid

The data obtained from the test results used a HOTS-based Two-Tier Multiple-Choice instrument. Then processed (analysis) using the ITEMAN 4.3 software. The analysis of students' higher order thinking skills is based on the students' correct and incorrect answers. The range of classification of answers on the HOTS-based Two-Tier Multiple-Choice test questions shown in Table 3.

Table 3. Two-Tier Multiple Choice Scoring

Criteria	Score
No answer	0
Wrong Answer and Wrong Reason	0
Correct Answer and Wrong Reason	1
Wrong Answer and Right Reason	1
Correct Answer and Correct Reason	2

Reliability test

Software Results Of Reliability Test Using ITEMAN 4.3. Data Is Provided In Table 4.

Table 4. Instrument Reliability Test Results

Reliability	Question	Description
0.725	30	Tall

The results of the reliability test of Cronbach's

Alpha on 50 respondents for the limited scale test, the results obtained for the Cronbach's Alpha for the limited test with a score of 0.725 and on Cronbach's Alpha as many as 30 questions. The test can use the Cronbach Alpha method. The questionnaire is said to be reliable if the value of Cronbach's Alpha is greater than the value of r table (E, 2011). *Cronbach's Alpha* of 0.05 The instrument can be said to be reliable (Azwar, 2016). The value of the validity and reliability of an instrument is influenced by the subject being measured, validity and reliability must always be tested before the instrument is used (Yusup, 2018). Reliability test can be processed with SPSS to determine the coefficient alpha (Azwar, 2019).

Difficulty of Problems and Differences

The test results for the difficulty test of the HOTS-based TTMC instrument, in the difficult category there is 1 question, the medium category is 27 questions and the easy category is 2. If a question has a balanced level of difficulty, it can be said that the question is good (Arifin, 2013). The level of difficulty of the questions can affect the distribution of the total test scores (Sutaryat, 2014) The results of the discriminatory power test on the developed HOTS-based two-tier instrument. In this study, to test the discriminatory power in the bad category as many as 9 questions, the sufficient category as many as 15 questions, the good category as many as 6 questions. In the difference power test, the measurement of the extent to which an item is able to distinguish students who master competence (Arifin, 2013). To distinguish the ability of students with high and low abilities (Widiyanto, 2018). There are several questions that must be revised before being tested on a wide scale.

Table 5. The Results Of The Recapitulation Of The Analysis Of The Differentiating Power Level And The Level Of Difficulty Of The Test Items On A Limited Scale.

Test classification	Category	Question Number
Different power test	Bad	1,2,3,7,8,19,20,22,26
	Enough	6,10,11,12,13,16,17,18,21,23,25,27,28,30
	Good	4,5,9,14,15,24,29
	Very Good	0
Test the difficulty of the question	currently	7
	currently	1,4,5,6,8,9,10,11,12,13,14,15,16,17,19,20,21,2223,24 25, 26, 27, 28
	easy	29,30

Distractor index runs well if it has been chosen at least 0.05% of all the options that have been answered by the test taker. (Widiyanto, 2018). In the distractor index (option) the answers to the questions did not go well below 0.05, namely 0.016, 0.017 and 0.027. The distractor index question is more functional than the answer key so that students have high, medium and low abilities compared to the answer key.

Comparison of students' higher order thinking skills on a limited scale trial

The results of the comparison are measured from the answers of students who have moderate high and low levels of ability in answering multiple choice questions based on reasons in the first tier are multiple choice answers while in the second tier are the reasons for the answers. The results of the item analysis on a limited scale test using the iteman software 4.3 indicate that the level of students' higher-order thinking skills. Presented in Table 6.

Table 6. The Recapitulation of Item Analysis Comparison

Items	Tier	Difficulty Level	Distinguishing Power	Option			
Item 001	1	0.56	-0.069	0.14	0.56	0.12	0.18
Item 002	1	0.54	0.138	0.14	0.54	0.16	0.16
Item 007	1	0.429	0.178	0.245	0.12	0.24	0.49
Item 008	1	0.52	0.067	0.18	0.52	0.12	0.18
Item 010	1	0.48	0.24	0.027	0.24	0.16	0.48
Item 012	1	0.56	0.288	0.2	0.08	0.56	0.16
Item 014	1	0.54	0.316	0.16	0.54	0.12	0.18
Item 015	1	0.56	0.354	0.16	0.1	0.18	0.56
Item 016	1	0.56	0.21	0.16	0.56	0.016	0.12
Item 026	1	0.58	0.126	0.18	0.08	0.58	0.16
Item 027	1	0.54	0.238	0.54	0.06	0.2	0.2
Item 028	1	0.54	0.2	0.54	0.16	0.1	**0.017**
Item 003	2	0.48	0.22	0.16	0.14	0.48	0.035
Item 007	2	0.5	0.283	0.5	0.24	0.08	0.18
Item 010	2	0.5	0.24	0.14	0.012	0.5	0.24
Item 013	2	0.5	0.255	0.5	0.06	0.28	0.16
Item 014	2	0.54	0.288	0.26	0.12	0.08	0.54
Item 021	2	0.54	0.285	0.54	0.22	0.18	0.06

The data from the analysis of higher order thinking ability test questions in multiple choice based on reasons for the tier, on item analysis for reason-based answers on the average value of students' answers with a score of 0.560. The results of the analysis of items on discriminating power which are given in blue indicate that it is 0.2%, so that the question is not able to distinguish students' abilities in working on questions, it can be seen at 1,2,7,8,26.

The results of the item analysis on the distractor index (option) which are colored yellow are below 0.05%, so the distractor index does not work effectively. In tier 1, where the distractor value does not work well, there is number 4 for answer A, in answer option B there are questions 12.26 and 27, answer option C is at number 16, in answer option D is number 13 tier 2 for open reasoned answers, for options that do not work well, answer B is listed at numbers 10 and 13, answer option C is listed at numbers 7 and 17, answer option D is listed at numbers 3 and 21.

The results of the test with a limited scale were analyzed, namely the difficulty of the questions, the discriminatory power and the distractor index (options) of the multiple-choice answers based on reasons. If a question has a balanced level of difficulty, then it can be said that bad luck is good. The test questions should not be too difficult or not too easy (Arifin, 2013). On Distinguishing power to see high-ability and low-ability students (Trisnamansyah, 2014). The student's answer option is in the bad category if the value of the student's answer option is below 0.05 (Widiyanto, 2018). The results of the analysis of items in the bad category are in options 0.017 and 0.028 because the distractor index is below 0.05, the items are categorized as bad, so they need revision.

Results of student responses to questions

The results of the response show that the quality of the questions can develop higher-order thinking skills using the Likers scale. The results of the recapitulation of student responses to the questions are presented in table 8.

Table 8. Results of the Recapitulation of Student Responses to The Items

Question	r- count	r- Table	Description
1	0.580	0,195	Valid
2	0.493	0,195	Valid
3	0.494	0,195	Valid
4	0.258	0,195	Valid
5	0.478	0,195	Valid
6	0.446	0,195	Valid
7	0.859	0,195	Valid
8	0.297	0,195	Valid
9	0.710	0,195	Valid
10	0.378	0,195	Valid

Question	r- count	r- Table	Description
11	0.360	0,195	Valid
12	0.839	0,195	Valid
13	0.429	0,195	Valid
14	0.674	0,195	Valid
15	0.717	0,195	Valid
16	0.287	0,195	Valid
17	0.732	0,195	Valid
18	0.464	0,195	Valid
19	0.709	0,195	Valid
20	0.295	0,195	Valid
21	0.607	0,195	Valid
22	0.747	0,195	Valid
23	0.539	0,195	Valid
24	0.560	0,195	Valid
25	0.639	0,195	Valid
26	0.555	0,195	Valid
27	0.523	0,195	Valid
28	0.455	0,195	Valid
29	0.330	0,195	Valid
30	0.458	0,195	Valid

The questions in the material developed in acid-base and salt solutions can make students motivated in learning chemistry so that students can develop higher-order thinking skills so that learning in the classroom becomes active.

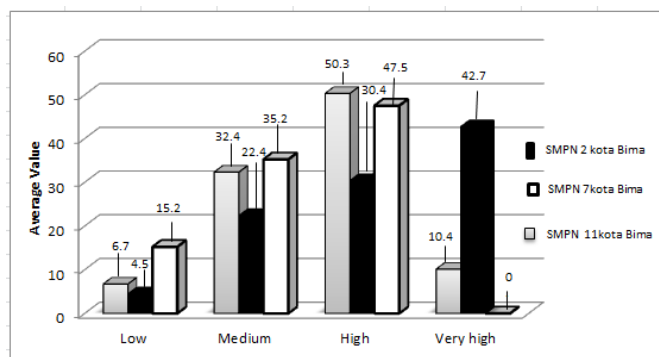### The results of the analysis of higher order thinking skills

Based on this research, table 3 has been carried out to analyze higher order thinking skills with three high, medium and low categories presented in table 9.

Table 9. Results of recapitulation of higher order thinking skills

| Assessment Aspect | SCHOOL | | |
	SMPN 2 Bima City	SMPN 7 Bima City	SMPN 11 Bima City
Low	4.5	15.2	6.7
Currently	22.4	35.2	32.3
Tall	30.4	47.5	50.3
Very high	42.7	0	10.1

The results of the analysis of higher order thinking skills at the three schools for the low category were found at SMPN 7 Kota Bima West Nusa Tenggara Indonesia with an average score of 24.47%, the medium category was at SMPN 11 Kota Bima West Nusa Tenggara Indonesia with an average score of 24.87%, SMPN 2 Kota Bima West Nusa Tenggara Indonesia with an average score of 24.87%. Average Value 25%.
The results of tests conducted in class VII at SMPN 2, SMPN 7 and SMPN 11 Bima city WestNusa Tenggara Indonesia which has a high, medium, and low average. This test was conducted to see the difference in HOTS at each level. The results in the table which are the high category at SMPN 2 Bima City, show the results of the HOTS percentage for the high category 4.5%, medium

22.4%, high 30.4% very high 42.7%. SMPN 7 Kota Bima for the HOTS test results in the low category with a percentage score of 15.2% low, 35.2% moderate47.5% high, and 0% very high. SMPN 11 Kota Bima for the HOTS test results in the medium category with a percentage score of 6.7% low, 32.3% moderate, 50.3% high and very high 10.1%. This shows that the use of HOTS questions in schools is very effective in improving students' high thinking skills with multi-level multiple choice questions. The HOTS comparison chart for the three schools is presented in Figure 1.

Figure 1: Comparison of the results of the Analysis of Students' Higher Order Thinking Abilit

The results of the assessment of higher order thinking can be seen a significant difference in Figure 1 above. HOTS (Higher Order Thinking Skills) is able to rovide changes in providing information in the form of changes in increasing critical thinking skills and creative thinking skills (Ahmad, 2020). effective for measuring students' ability to answer questions (Hidayatullah et al., 2020) an interaction learning with peers who have high, medium and low abilities and high learning motivation has a direct effect on students' HOTS (Lu et al., 2021).

The question is 0.560%, and discriminatory power is 0.229%. In measuring higher order thinking skills, 3 junior high school students in Bima City, West Nusa Tenggara, Indonesia with altitude the average score has a very good high order thinking skills, compared to schools with medium and low abilities percentage of low score is 6.7% moderate, 35.2% high 47.5% very high 10.1%. Other researchers can use this research as a basic reference for assessing students' higher order thinking skills with or higher models such as Tree-Tier or four-tier.

Conclusion

The conclusion of this research is the reliability value is 0.725, the difficulty of the question is 0.560, and the discriminatory power is 0.229. In measuring higher-order thinking skills, 3 junior high school students in

Bima City, West Nusa Tenggara, Indonesia with high average scores have very good higher-order thinking skills, compared to schools with medium and low abilities. The percentage of low score 6.7% moderate, 35.2% high 47.5% very high 10.1%. Other researchers can use this research as a basic reference to assess the high-level thinking ability of students with similar or higher models such as Tree-Tier or four-tier.

Acknowledgements

References

Ahmad, D. N. (2020). Analisis Sistem Penilaian Hots (Higher Order Thinking Skills) Dalam Mengukur Kemampuan Berpikir Kritis Dan Kreatif. *BIOTIK: Jurnal Ilmiah Biologi Teknologi Dan Kependidikan*, *8*(1), 11. https://doi.org/10.22373/biotik.v8i1.6600

Anggraeni, D. M., & Sole, F. B. (2020). Analysis of Science Teachers' Understanding of High Order Thinking Skills (HOTS) and Their Implementation in Learning. *Jurnal Penelitian Pendidikan IPA*, *6*(2), 210. https://doi.org/10.29303/jppipa.v6i2.411

Arifin, Z. (2013). *Evaluasi Pembelajaran* (pendidikan agama Islam (ed.); pp. 47–57). Rosda jakarta.

Azwar, S. (2016). *Penyusun Skala Psikologi*. In pusaka belajar (Ed.), Pusaka Belajar (pusaka bel, pp. 115–120). yogyakarta.

Azwar, S. (2019). *Reliabilitas dan Validitas*. In Pusaka Belajar.

Brookhart. (2010). Assess higher-order thingking skills in your classroom. Alexandria : ASCD. In *Journal of Education* (Vol. 88, Issue 18). https://doi.org/10.1177/002205741808801819

Cengiz, T. ysuuml; z. (2009). Development of two-tier diagnostic instrument and assess students' understanding in chemistry. *Scientific Research and Essays*, *4*(6), 626–631.

E, R. W. (2011). Uji Validitas dan Reliabilitas dalam Penelitian Epidemiologi Kedokteran Gigi. *Stomatognatic (J.K.G. Unej)*, *8*(1), 27–34.

Gultom, S. T., Siahaan, P., & Suhandi, A. (2021). Effect of PBL Hybrid Learning on the Higher Order Thinking Skills of Seventh Grade Students in Global Warming and Their Environmental Care Attitudes. *Jurnal Penelitian Pendidikan IPA*, *7*(SpecialIssue), 272–280. https://doi.org/10.29303/jppipa.v7ispecialissue.1012

Hartono, H., Indra Putri, R. I., Inderawati, R., & Ariska,

M. (2022). The strategy of Science Learning in Curriculum 2013 to Increase the Value of Science's Program for International Student Assessment (PISA). *Jurnal Penelitian Pendidikan IPA*, *8*(1), 79–85. https://doi.org/10.29303/jppipa.v8i1.1185

Hasyim, M., & Andreina, F. K. (2019). Analisis High Order Thinking Skill (Hots) Siswa Dalam Menyelesaikan Soal Open Ended Matematika. *FIBONACCI: Jurnal Pendidikan Matematika Dan Matematika*, *5*(1), 55. https://doi.org/10.24853/fbc.5.1.55-64

Hidayatullah, A. R., Yamtinah, S., & Masyukri, M. (2020). Pengembangan Instrument Two-Tier Multiple Choise Berbasis Hots Pada Materi Ipa Terpadu Asam. *Seminar Nasional Kimia Dan Pendidikan Kimia Xii (Sn-Kpk Xii) Pendahuluan*, 50–59.

Hikmawati, H., Suastra, I. W., Suma, K., Sudiatmika, A. A. I. A. R., & Rohani, R. (2021). Effect of Problem-Based Learning Integrated Local Wisdom on Student Hots and Scientific Attitude. *Jurnal Penelitian Pendidikan IPA*, *7*(SpecialIssue), 233–239. https://doi.org/10.29303/jppipa.v7ispecialissue.1118

Kurniati, D., Harimukti, R., & Jamil, N. A. (2016). *Kemampuan Berpikir Tingkat Tinggi Siswa Smp Di Kabupaten Jember Dalam Menyelesaikan Soal Berstandar Pisa*. *20*(2).

Lu, K., Yang, H. H., Shi, Y., & Wang, X. (2021). Examining the key influencing factors on college students' higher-order thinking skills in the smart classroom environment. *International Journal of Educational Technology in Higher Education*, *18*(1), 1–13. https://doi.org/10.1186/s41239-020-00238-7

Maharani, Y., & Rakhmat1, C.A.S. (2015). Analisis Hots (High Order Thinking Skills) Pada Soal Objektif Tes Dalam Mata Pelajaran Ilmu Pengetahuan Sosial (Ips) Kelas V Sd Negeri 7 Ciamis. *PEDADIDAKTIKA: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, *2*(2), 187–195.

Mitana, J. (2018). Assessment of higher order thinking skills: A case of Uganda Primary Leaving Examinations. *African Educational Research Journal*, *6*(4), 240–249. https://doi.org/10.30918/aerj.64.18.083

Nachiappan, S., Ahmad Damahuri, A., Ganaprakasam, C., & Suffian, S. (2018). Application of Higher Order Thinking Skills (HOTS) in teaching and learning through communication component and spiritual, attitudes and values component in preschool. *Southeast Asia Early Childhood Journal*, *7*, 24–32. https://doi.org/10.37134/saecj.vol7.3.2018

Narayanan, S., & Adithan, M. (2015). Analysis Of Question Papers In Engineering Courses With Respect To Hots (Higher Order Thinking Skills).

American Journal of Engineering Education (AJEE), *6*(1), 1–10. https://doi.org/10.19030/ajee.v6i1.9247

Nofiana, M., Sajidan, S., & Puguh, P. (2014). Pengembangan instrumen evaluasi two-tier multiple choice question untuk mengukur keterampilan berpikir tingkat tinggi pada materi kingdom plantae. *Jurnal Inkuiri*, *3*(2), 60–74. https://doi.org/10.20961/inkuiri.v3i2.9694

R. Aiken. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement, V45 N1 P131-42 Spr 1985*, *1*(1), 1–11. https://doi.org/https://doi.org/10.1177/0013164485451012

Sesli, E., & Kara, Y. (2012). Development and application of a two-tier multiple-choice diagnostic test for high school students' understanding of cell division and reproduction. *Journal of Biological Education*, *46*(4), 214–225. https://doi.org/10.1080/00219266.2012.688849

Shidiq, A. S., Masykuri, M., & Susanti, E. (2015). Analisis Higher Order Thinking Skills (HOTS) Menggunakan Instrumen Two-Tier Multiple Choice Pada Materi Kelarutan Dan Hasil Kali Kelarutan Untuk Siswa Kelas Xi Sma N 1 Surakarta. *Prosiding Seminar Nasional Pendidikan Sains*, *November*, 2015–2159.

Sutaryat, T. (2014). *Evaluasi Pembelajaran*. pustaka setia.

Taubah, M. (2019). Penilaian HOTS dan Penerapannya di SD/MI. *ELEMENTARY: Islamic Teacher Journal*, *7*(2), 197. https://doi.org/10.21043/elementary.v7i2.6368

Trisnamansyah, S. (2014). evaluasi pembelajaran. In E. Ratna (Ed.), *evaluasi pembelajaran* (pp. 1–339). pustaka setia.

Tune Sumar, W., & Tune Sumar, S. (2020). Implementasi Program Pengembangan Keprofesian Berkelanjutan Guru melalui Peningkatan Kompetensi Pembelajaran Berbasis Zonasi. *Pedagogika*, *10*(2), 84–94. https://doi.org/10.37411/pedagogika.v10i2.60

Widiyanto, J. (2018). Evaluasi Pembelajaran (Sesuai dengan Kurikulum 2013) Konsep, Prinsip & Prosedur. In *Unipma Press* (Vol. 53, Issue 9).

Yusup, F. (2018). Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif. *Jurnal Tarbiyah : Jurnal Ilmiah Kependidikan*, *7*(1), 17–23. https://doi.org/10.18592/tarbiyah.v7i1.2100