International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2021 International Journals of Sciences and High Technologies
http://ijpsat.ijsht-journals.org
Vol. 28 No. 2 September 2021, pp.619-626

Development of Test Instrument Two Tier Based on Higher Order Thinking Skills (HOTS)on Volta Cell Material for SMA/MA Students

Genia Tarina[1], Andromeda[1*]

[1]Department of Chemistry, Faculty of Mathematics and Sciences,
Universitas Negeri Padang
Padang, Indonesia

Abstract— HOTS is a thinking ability that can lead students to transfer and connect between concepts, process information, solve problems based on the information obtained and examine creative, innovative and critical ideas. With this HOTS, it can improve students' thinking and creativity skills and prepare their abilities to face the industrial revolution 4.0 in the 21st century. The type of research used is research and development (R&D) with the Plomp development model. Based on the results of the content validity research consisting of questions, stimulus questions, answers to questions (first tier) and reasoning questions (second tier) the CVI value was obtained. 0.95, 0.96, 0.95 and 1. The instrument was tested on class XII students of SMAN 3 Payakumbu and 19 items of first tier and second tier questions were empirically valid. Instrument reliability for first tier and second tier respectively namely 0.97 and 0.91. The practicality of the test instrument by the teacher was obtained by a score of 83.33, the practicality of the student obtained a score of 75.11. Based on the data obtained, the developed test instrument is valid, has very high reliability and is practical.

Keywords— Test Instrument, HOTS, Two Tier, Voltaic Cell, Plomp.

I. Introduction

In an effort to improve the quality of learning, the Ministry of Education and Culture through the Directorate General of Teachers and Education Personnel (Ditjen GTK) has developed a learning program oriented to Higher Order Thinking Skills (HOTS) [1]. Based on Bloom's Taxonomy revision of Higher Order Thinking Skills (HOTS) includes analytical and synthesis skills (C4), evaluating (C5) and creating or creativity (C6) [2].

HOTS is a student's ability to solve a problem to be able to think analytically, critically and creatively and to equip students in the four competencies that exist in the 2013 curriculum, the 2013 curriculum is designed with various improvements, however, from the results of the evaluation of the 2013 curriculum implementation, it is known that the teacher's understanding of the assessment of student outcomes is the main problem [3]. The assessment should be carried out referring to the achievement of Basic Competence (KD) by comparing the achievement of students with predetermined competency criteria [4].

Based on the results of the Basic Competency (KD) analysis that has been carried out, the results show that as many as 42.86% of KD demand a HOTS-based learning system. However, from the results of a questionnaire distributed through Google Forms conducted to eight chemistry teachers from several schools in West Sumatra Province, it was found that only 25% of teachers had implemented the HOTS-based learning system. Then from the questionnaire, it was also found that 62.5% of teachers used LOTS questions for evaluating chemistry learning. This happens because the evaluation instruments used by teachers are still sourced from the internet, printed books and textbooks, causing the lack of HOTS-based test instruments in schools. The results of the questionnaire also showed that as many as 75% of teachers needed the HOTS question instrument and 25% of the teachers really needed the HOTS question instrument to measure students' higher-order thinking skills on the voltaic cell material.

One of the instruments that can be used to measure students' higher-order thinking skills and meet the demands of KD 3.4 at the analysis level is a test instrument in the two-tier form of multiple choice questions. Two tier is a form of evaluation question with two levels where the first tier is in the form of the content of the questions on the material being taught while the second tier is the cause or reason for the content of the questions answered [5]. The second level of the two tier is also used to reduce the occurrence of guessing answers on ordinary multiple choice questions [6].

II. Methodology

The type of research used in this research is Research and Development (R & D). This type of research is research that is used to create and produce a particular product and to test the practicality and effectiveness of the resulting product [7]. The development model used in this study is the Plomp model developed by Tjreed Plomp which consists of three stages, namely preliminary research, prototyping stage and assessment phase [8]. This research was conducted at SMA Negeri 3 Payakumbuh in the odd semester of the 2021/2022 academic year. The subjects of this study were lecturers of the Department of Chemistry, FMIPA UNP, chemistry teachers and class XII students of SMA/MA. The object of this research is a collection ofTest Instruments Two Tier based on Higher Order Thinking Skills (HOTS) on Voltaic Cell material for SMA/MA students.

The type of data in this study is primary data because it is obtained directly from lecturers, teachers and students. In addition, there is also secondary data obtained from literature studies from books, journals, theses and other sources related to research. The instruments used to collect data in this study were interview questionnaire sheets, instrument validity sheets and practicality sheets.

To validate the content of this test instrument, it is measured using the content validity coefficient of Lawshe's CVR (content validity ratio). Lawshe said that for each subject matter expert (SME) or assessor consisting of a panel of experts to answer questions from each item with three answer choices, namely (3) essential, (2) useful but not essential, (1) not required. The CVR score is also determined by the number of validators or SME. If the CVR score exceeds the critical limit or minimum value of CVR then the test item is declared valid and vice versa if the CVR value is less than the critical limit or minimum value then the test item is declared invalid. The critical limit for 5 validators is 0.736 and for 6 validators is 0.672 [9].

The mathematical equation for Lawshe's proposed formula is:

$$CVR = \frac{2ne}{N} - 1$$

Description:

CVR : content validity ratio

ne : Number of SME giving a score of +1

N : Number of SME [10]

Meanwhile, construct validity was analyzed usingIndex Aiken's V. The construct validity test uses 4 scales consisting of: not relevant (TR), less relevant (KR), quite relevant (CR) and relevant (R). Theindex Aiken's V for assessing the construct validity test is formulated as follows:

$$V = \frac{r - l_0}{n(c-1)} = \frac{\Sigma s}{n(c-1)}$$

Description:

V : rater agreement index

s : $r - l_0$

lo : the lowest score of validity (in this case = 1) c : the highest number of validity assessments (in this case = 4)

r : the number given by the raterraters

n : the number of raters

Table 1. Value of Aiken Scale Validity Coefficient V [11]

Raters	**Number of Rating Categories**	
	4	
	v	**p**
5	0.93	0.006
5	0.87	0.021
6	0.89	0.007
6	0.78	0.050

After validation of the test instrument, it was tested on a small group consisting of 20 students of SMAN 3 Payakumbuh. The test of this instrument aims to determine the empirical validity, reliability, discriminating power, difficulty index and distractor function of the question. Item analysis was carried out using the Anates application. Anates is software aprogram computer that is used in analyzing the items. Anates was developed by Mr. Drs. Karno To, M.Pd. a lecturer in Psychology at UPI and Mr. Yudi Wibisono ST [12].

After the small group test was carried out, field tests were carried out on 30 students and 2 chemistry teachers at SMAN 3 Payakumbuh. This test was conducted to determine the practical value of the developed instrument. The percentage of practicality level is calculated using the calculation formula:

$$practical\ value = \frac{average\ score}{maximum\ score} x100\%$$

Table 2. Practicality Level Category Scale [13]

Interval	**Category**
0% - 20%	Impractical
20% - 40%	Less Practical
40% - 60%	Practical Enough
60% - 80%	Practical
80% - 100%	Very Practical

III. Result And Discussion

A. Result

The overall research results for each stage are described as follows.

1) Preliminary Research

a) Needs Analysis

The needs analysis was carried out by distributing questionnaires using google form with 8 chemistry teachers regarding the development ofbased test instruments Higher Order Thinking Skill (HOTS)from several schools in West Sumatra. Based on the distributed questionnaires, it was found that all schools had implemented the 2013 curriculum, but in the learning process and the types of test instruments used, most of them still used Lower Order Thinking Skills (LOTS) due to the lack of availability of HOTS instruments.

b) Context Analysis

The context analysis process is carried out by analyzing the curriculum used in schools. Based on the results of the questionnaire at the time of analysis of curriculum content used in schools is the 2013 curriculum. As stated in Ministerial Regulation No. 37 of 2018 there are 42.86% of basic competencies (KD) that demand based learning system Higher Order Thinking Skills (HOTS)and to develop The test instrument requires material that is at the C4-C6 level, one of which is the voltaic cell material at the analytical level or C4.

c) Literature Study

The literature study aims to find and understand the sources or references needed to developtest instrument two-tier a HOTS-basedfrom various sources such as books, journals and other sources on the internet.

d) Conceptual Framework Development

At this stage the things that are done are identifying, detailing and compiling the main concepts needed for the test instrument derived from the IPK on the voltaic cell material. Then the IPK is reduced to a question indicator and after that it is written in the form of a grid and a question table. In the developed test instrument there are 30 questions where each indicator has 10 questions.

2) Prototyping Stage

a) Prototype 1

The development of prototype 1 begins with designing questions based on a grid of questions derived from indicators of achievement of competence (IPK) and basic competence (KD), preparing stimulus questions, making answers (first tier) and reasons (second tier), answer keys to questions, the identity of the question to the stage of making the design of the cover of the test instrument.

b) Prototype 2

In prototype 2, acarried out self-evaluation was to improve prototype 1 by using asystem checklist in the questionnaire used based on the characteristics of the HOTS-based test instrument developed.

c) Prototype 3

At this stage an evaluation is carried out in the form of an expert review (expert review) and one-to-one evaluation (one to one evaluation) and on the prototype 2. At the expert review stage, an analysis of test instruments is carried out which consists of an analysis of content validity and construct validity. Content validity was analyzed using the CVR method, while construct validity was analyzed using Aiken V. The number of validators in the analysis of developing test instruments was 6 people consisting of 3 chemistry lecturers, FMIPA UNP and 3 chemistry teachers. In content validity there are 4 forms of assessment, namely on questions, stimulus questions, answers to questions (first tier) and reasons for questions (second tier). Based on the results of the validity of the test instrument from 6 validators on 30 items, there are 4 valid questions, namely items number 18, 20, 24 and 27 and the remaining 26 items are not valid because they do not reach the minimum value of the CVR. Based on the content validity analysis data prior to revision, it can be seen in the graph below:

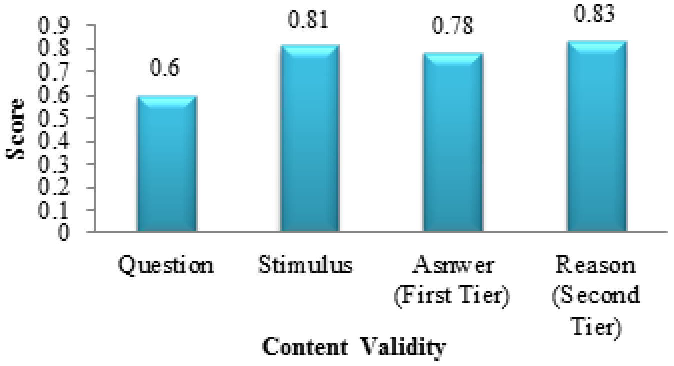

Fig. 1 . Graph of Average Content Validity Before Revision

Items that are not valid are revised according to the suggestions given by the validator. After the revision of the content validity was carried out, the average CVI value of the test instrument was 1 with the category of all items being valid. In construct validity, an assessment of material aspects, presentation, language and additional rules of the developed test instrument was carried out. Construct validity was analyzed using the Aiken V formula. The results obtained from construct validity were 0.95 for the material aspect, 0.96 for the presentation aspect, 0.95 for the language aspect and 1 for additional rules with valid categories.

In the stage one-to-one evaluation , a trial was conducted on 3 class XII students with different abilities related to the test instrument two-tier HOTS-based on the resulting voltaic cell material. At this stage the student is given a questionnaire to ask questions related to the instrument including aspects of the language used, understanding of the instructions for working on the questions, readability of tables, data, pictures on the test instrument, the time needed to complete the test, and the obstacles experienced when doing the test. two tier based HOTS on voltaic cell material.

d) ***Prototype 4***

Phase 4 of the formation of prototypes carried out small-scale trials (small group) of 20 students. The item analysis was analyzed using the anates application. Each item is divided into two forms of analysis, namely the answer to the question (first tier) and the reason for the question (second tier). The results of the analysis obtained in the first tier and second tier there are 19 items that are empirically valid, 1 question is not legible and 10 questions are not valid. Distractors or decoys were analyzed using anates application. Distractors are said to be effective when selected by at least 5% of test participants [14].

Table 3. Reliability Score

Test Instrument	Reliability
First Tier	0.97
Second Tier	0..91

Table 4. Question Difficulty Level

Test Instrument	Difficulty Level Persentage				
	Very Difficult	*Difficult*	*Medium*	*Easy*	*Very Easy*
First Tier	13,33%	10%	40%	30%	6.67%
Second Tier	6.67%	6.67%	56.67%	23.33%	6.67%

Tabel 5. The Persentage of Different Power Question

Test Instrument	Power Persentage Difference				
	Very Ugly	***Ugly***	***Medium***	***Good***	***Very Good***
First Tier	13.33%	16.67%	6.67%	30%	33.33%
Second Tier	13.33%	10%	20%	16.67%	40%

3) Assessment Phase

Assessment stage is the stage of field trials(field tests) to test instruments, involving 30 students and 2 teachers of chemistry to look at the practicalities of the prototype level 4 is generated.

a) Practicality by the teacher

The average practicality result from the teacher is 83.33 with a very practical category.

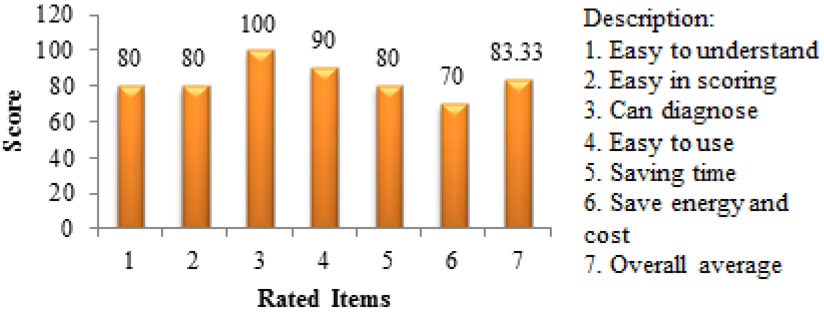

Fig. 2 . Practicality Chart by Teacher

b) Practicality by the studens

The average practicality result of students is 75.11 in the practical category.

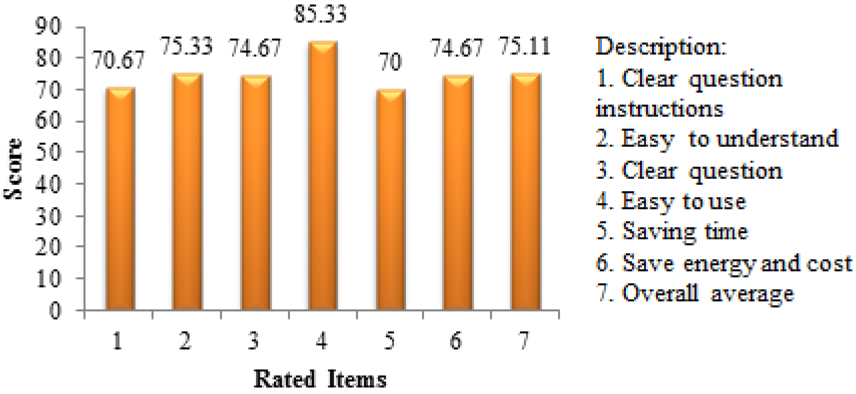

Fig. 3 . Practicality Chart by Students

B. Discussion

The Plomp model consists of three stages, namely preliminary research, prototyping phase and assessment phase. To get a valid test instrument on the developed test instrument, content validity is carried out which consists of questions, stimulus questions, answers to questions (first tier). , the reason for the question (second tier) and construct validity consisting of material aspects, presentation aspects, language aspects and additional rules that are provisions in a test instrument. Content validity or content validity is carried out to obtain a valid test instrument in terms of stimulus questions, stimulus questions, answers to questions (first tier) and reasoning questions (second tier). Content validity is done 2 times. In the first validation there were 4 valid items while 26 items were declared invalid because the items did not reach the specified minimum CVI limit. The items that are not yet valid are because there are still many questions that are still at the C3 level, the questions are not quite right, the lack of stimulus in the questions and the reasons for the questions are not quite right so it is necessary to make revisions based on the suggestions given by the validator. The results of the revision were then re-validated to obtain a valid test instrument. The results of the second validation showed that all items were valid and the CVI value for all items was 1.

Construct validity was carried out to obtain a valid test instrument in terms of material aspects, presentation aspects, language aspects and additional rules of test instruments. The analytical technique used is the Aiken V formula and the scale used consists of 4 components, namely irrelevant (TR), less relevant (KR), quite relevant (CR) and relevant (R). In the analysis of material aspects related to the suitability of the question indicators with IPK and KD, the stimulus questions used, logical and homogeneous answer choices. The results of the material aspect analysis obtained a value of 0.95 with a valid category. Construct validity in the presentation aspect consists of questions formulated effectively, questions do not provide clues to the answer key and questions do not depend on answers to other questions. The results of construct validity from the presentation aspect obtained a value of 0.96 with a valid category.

In the analysis of the construct validity of the language aspect in terms of the suitability of the language of the question with the rules of Indonesian, a language that is communicative and does not repeat words or groups of words. The results of construct validity from the language aspect obtained a value of 0.95 with a valid category. Furthermore, the construct validity of the additional rule aspect relates to the questions formulated that do not contain SARAPPPK (ethnicity, religion, race, intergroup, pornography, politics, propaganda and violence. The result of construct validity from the additional rule aspect is obtained with a value of 1 with a valid category.

Empirical validity was analyzed by using anates application. Each item is divided into two forms of analysis, namely the answer to the question (first tier) and the reason for the question (second tier). The results of the analysis at the stage small group for the first tier and second tier there are 19 items that are empirically valid because they have a positive correlation value, 1 item is not legible because it has a correlation value of 0.00 and 10 items are invalid because they have a negative correlation value.

The test instrument is said to be reliable if the test instrument gives the same results if it is repeated to evaluate the same object [14]. The reliability of the items was analyzed using the anates application which was divided into two parts, namely first tier and second tier. Test the reliability of the questions in the small group was carried out on 20 students. The reliability results in thetest small group for the first tier obtained a reliability value of 0.97 with a very high category and the second tier reliability value of 0.91 with a very high category.

Item analysis of the items was carried out to improve the quality of the questions. Items that have an effective difficulty index if the questions can be answered correctly by all upper, middle and lower groups, in other words, the items are not too easy and not too difficult or have a moderate level of difficulty. Differential power analysis was carried out using the anates application. The discriminatory power of the items is the ability of the items to be able to distinguish between students in the upper group and students in the lower group. Analysis of distractors or distractors were analyzed using anates application. The test instrument developed has five answer choices with one correct answer. An effective distractor if at least 5% of all test takers are selected.

An evaluation instrument must be equipped with instructions for working, uncomplicated and in simple language. Practical learning outcomes tests are learning outcomes tests whose implementation does not require a long time of effort and large costs [14]. The results of the practical analysis of the test instrument two-tier HOTS-based by 2 chemistry teachers at SMAN 3 Payakumbuh obtained a practicality value of 83.33 with a very practical category. While the value of practicality by 30 students of SMAN 3 Payakumbuh obtained a practicality value of 75.11 with a practical category.

IV. Conclusion

Based on the results of the research and data analysis that has been done, it can be concluded that the instrument two-tier HOTS-based on voltaic cell material for SMA/MA students is valid, has very high reliability and is practical to use.

REFERENCES

[1] Ariyana, Y., Bestary, R., Yogyakarta, U. N., & Mohandas, R. (2018). Buku Pegangan Pembelajaran Berorientasi pada Keterampilan Berpikir Tingkat Tinggi. Direktorat Jenderal Guru dan Tenaga Kependidikan Kementerian Pendidikan dan Kebudayaan.

[2] Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). A taxonomy for learning, teaching, and assessing: a revision of Bloom’s taxonomy of educational objectives. 41(4), 352.

[3] Direktorat Pembinaan SMA Ditjen Pendidikan Dasar dan Menengah. Panduan Penilaian oleh Pendidik dan Satuan Pendidikan untuk Sekolah Menengah Pertama. Jakarta: Kemdikbud; 2017.

[4] Andromeda, Z. F. and F. Q. ‘Aini. (2020). Evaluasi Kompetensi Pedagogik Guru Kimia Dalam Menyusun Instrumen Penilaian Higher Order Thinking Skill (HOTS) Siswa SMA Evaluation of Pedagogy Competence of Chemistry Teacher. 2(2), 91–95.

[5] Treagust, D. F. (2006). Diagnostic Assesment In Science as A Means to Improving Teaching, Learning, and Retention. UniServe Science Assesment Symposium Proceedings. The University of Sydney, 28 September 2006.

[6] Cullinane, A. dan M. L. (2011). Two-tier Multiple Choice Question: An Alternative Method of Formatif Assessment for First Year Undergraduate Biology Students. National Center for Excellence In Mathematics and Education Science Teaching and Learning (NCE-MSTL).

[7] Sugiyono. 2017.Metode Penelitian Kuantitatif, Kualitatif dan R&D.Bandung:CV. Alfabeta.

[8] Plomp, Tjreed. 2007. Educational Design Research: An Indroduction to Educational Research Enschede. Netherland : National Institute For Curriculum Development.

[9] Wilson, F. R, et al. 2012. “Recalculationof the Critical Values For Lawshe’s Content Validity Ratio”. Measurement and Evaluation In Counseling and Development. 45(3): 197-210.

[10] Lawshe. 1985. A Quantitative Approach to Content Validity. Personel Psyclogy.

[11] Aiken, L. R. 1985. Three Coefficients foe Analyzing The Reliability, and Validity of Ratings. Educational and Psychological Measurement, 45, 131-142.

[12] Sari, P. I., & Yudha, R. I. (2020). Pemanfaatan Penerapan Media Berbasis Software Anates pada Mata Kuliah Evaluasi Pembelajaran di Universitas Batanghari Jambi. 20(1), 81–85.

[13] Riduwan. 2010. Skala Pengukuran Variabel-Variabel Penelitian. Bandung: Alfabeta.

[14] Latisma. 2011.Evaluasi Pendidikan. Padang: UNP Press.