The Development of HOTS Test of Physics Based on the Modern Test Theory: Question Modeling through E-learning of Moodle LMS

by Sri Wahyu Widyaningsih

The Development of HOTS Test of Physics Based on the Modern Test Theory: Question Modeling through E-learning of Moodle LMS

Sri Wahyu Widyaningsih
Assist. Prof., Faculty of Teacher Training and Education, Universitas Papua, Indonesia, *s.widyaningsih@unipa.ac.id*

Irfan Yusuf
Assist. Prof., Faculty of Teacher Training and Education, Universitas Papua, Indonesia, *i.yusuf@unipa.ac.id*

Zuhdan Kun Prasetyo
Prof., Faculty of Science, Universitas Negeri Yogyakarta, Indonesia, *zuhdan@uny.ac.id*

Edi Istiyono
Prof., Graduate School, Universitas Negeri Yogyakarta, Indonesia, *edi_istiyono@uny.ac.id*

The present study discussed the development of higher-order thinking skills (HOTS) test of physics based on the modern test theory. HOTS questions were designed and presented in the e-learning with the Moodle learning management system (LMS) that could be accessed online. This study employed the ADDIE model with analysis, design, development, implementation, and evaluation stages. The instrument consisted of 24 multiple choice physics questions regarding the direct current circuit topic; the questions were designed by following the aspects and sub-aspects of HOTS and had been validated by the experts of measurement, physics education, physics, and practitioners. Moreover, validity analysis was based on the V Aiken formula, in which every aspect was confirmed valid. The validated instrument was then tried out to all 34 students at the Department of Physics Education, Universitas Papua, who participated in the basic physics subject. Dichotomy data analysis used the Rasch Model (RM) 1-PL through the Quest program, and the test characteristics comprised item fitness, reliability, and difficulty. The trial result obtained the criteria of INFIT MNSQ mean and standard deviation of 1.0 and 0.0, respectively, showing that the items fitted the RM1-PL. In addition, the value of item reliability based on the value summary of the item estimate arrived at 0.66; meanwhile the case reliability under the summary of the case estimate accounted for 0.85. The reliability value in the range of 0.67- 0.80 was categorized as quite reliable. As based on the criteria of minimum and maximum INFIT MNSQ of 0.77 and 1.30, 24 question items fitted the RM 1-PL model. The result of the Quest output also revealed that the average values of Thresholds and its standard deviation were 0.00 ± 0.71, or in the acceptance range of -2 to 2. All in all, all 24

question items that had been tried out had fitted the model with a good category in order that they could be utilized in HOTS measurement.

Keywords: E-learning, HOTS Test, and Modern Test Theory.

INTRODUCTION

Assessment, particularly in the cognitive domain, is central to the learning process and should be carried out accurately and in compliance with the subject to be assessed measured. Students' cognitive skills in the learning process can be categorized into lower-order thinking (LOT) and higher-order thinking (HOT). The LOTS include remembering, understanding, and applying; the HOTS, on the other hand, encompass analyzing, evaluating, and creating. HOTS are thinking skills that do not only require the remembering skill but also require other higher skills. Indicators to measure HOTS consist of analyzing (C4), evaluating (C5), and creating (C6) skills (Krathwohl & Anderson, 2010).

HOTS also refer to thinking skills when one takes new information, connects it with initial information s/he has, and finally delivers the information to achieve goals or answer questions (Istiyono, Dwandaru, & Muthmainah, 2019). This is in line with skill characteristics in the 21st century published by Partnership of 21st Century Skill stating that 21st century learners should be able to develop competitive skills, such as critical thinking, problem-solving, communication, information and communication technology (ICT) literacy, ICT, information literacy, and media literacy (Brun & Hinostroza, 2014); these focus on HOTS development.

Physics serves as part of science consisting of abstract concepts that are difficult to be directly described. Learning physics is expected to help students develop their thinking skills, in which they are not only demanded to master LOT skills, but also HOTS. Teachers are also urged to deliver learning materials to students, including the HOTS that can be improved by HOTS instrument. A previous study has reported that the majority of teachers find it challenging to develop an assessment instrument of learning outcomes, HOTS questions, in particular (Istiyono, 2018). For this reason, teacher creativity is highly required to measure students' learning outcomes. Today's development of Information and Communication Technology (ICT) can be utilized to design and habituate students to learn anywhere at any time (Yusuf, Widyaningsih, & Sebayang, 2018). Relying on ICT during the learning process is one of the significant innovations, including in the evaluation of students' learning outcomes.

The presentation of evaluation questions can be done in an integrated manner through e-learning programs, one of which is Moodle learning management system (LMS) (Azevedo, 2015; Bogdanović, Barać, Jovanić, Popović, & Radenković, 2014).

The Moodle provides different types of questions, such as multiple choices, true or false, and short answers; these are stored in the taught course database and can be re-used (Limongelli, Sciarrone, & Vaste, 2011). Teachers are also able to give feedback directly to the students and give them correct answers to questions they have worked on (Pandey & Pandey, 2009). One of the advantages of an online evaluation through Moodle LMS is that students can directly figure out their assessment results.

Teachers need to prepare a good test to measure students' learning outcomes. There are two paradigms developed for students' learning outcome assessment through the applied test, i.e., classical and modern approaches. The classical paradigm being utilized is classical test theory or widely known as classical true-score theory, meanwhile, the modern paradigm is item response theory (IRT). The classical test theory is selected due to its ease in the application despite of its limitations in measuring the item difficulty level and discrimination since the calculation of both indicators is based on the test taker's total score. In contrast, the IRT frees up the independence between the test item and test taker (a concept of parameter invariance); the test taker's response to a test item does not affect another item (a concept of local independence), and; the test item does only measure one measurement dimension (unidimensional concept) (Raykov & Marcoulides, 2015). Therefore, the application answers the needs of modern measurement to date, i.e., a comparison between test taker's skills, question development, and even adaptive test development, so that it is considered able to overcome the classical test theory limitations.

This development study is an initial study with a long-term purpose of developing general physics questions with good quality at the Department of Physics Education, Universitas Papua. As the first stage, this study focuses on students at the department mentioned previously who enroll in General Physics subject taught by the researcher. This study also serves as one of the efforts to expand students' HOTS by applying a variety of HOTS-based learning sources.

METHOD

The ADDIE model, as employed by this study, refers to a general and systematic model of development study with a phased framework, allowing each element to connect with each other (Aldoobie, 2015). The stages of this model used in the development of HOTS instrument are presented in Figure 1.

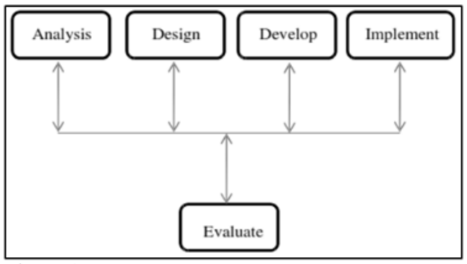

Figure 1
Stages of ADDIE Development Model in Designing Moodle LMS-based HOTS Test.

Analysis

The analysis stage is a process of needs analysis in the form of determining test objectives, identifying problems, analyzing tasks, and determining question formats to be applied. It is revealed that the problems are related to the needs of HOTS instrument design for students at the Department of Physics Education, Universitas Papua.

Design

This stage comprises the process of designing HOTS questions to be used; the design process encompasses creating a question matrix and outline that covers question distribution in every aspect and sub-aspect of HOTS.

Develop

Moreover, every single thing required in the arrangement of HOT skill questions has been prepared in the next stage. This stage also comprises the process of making the questions regarding HOTS, as well as validating the questions that involve the experts of measurement, physics education, and practitioners. The technique of validity analysis to assess the content validity of the developed questions applies the V Aiken formula (Aiken, 1980, 1985).

$$V = \Sigma s / n(c-1) \quad (1)$$

"V" refers to the agreement index of validators in regards to item validity; "s" is the assessment score of validators subtracted by the assessment lowest score; "n" refers to the number of validators; "c" is the number of categories that can be chosen by validators. All test items are considered valid if the value of the V Aiken index falls into the range of 0.37 - 1 (Kowsalya, Venkat Lakshmi, & Suresh, 2012). The value of V Aiken of every test item is calculated based on the assessment items of every validator. In this stage, there is also an evaluation process, i.e., revising questions by following validators' corrections and suggestions.

Implementation

Another stage is applying HOTS questions that have been developed to 34 students in the site area who enroll in general physics subject. This number has been following the sample size for data stability in Rasch Model (RM) 1- PL, which is from 30 to 300, with the limit of INFIT t is from -2 to +2 (Bond & Fox, 2007). Question item analysis is performed based on the raw score of the students by employing the Quest program.

Evaluation

Evaluation is a process of finding out whether or not the developed questions of HOTS have met the expectation. The evaluation stage is carried out in every stage and called a formative evaluation intended for revisions (Lee & Zainal, 2017). For instance, in the design stage, the expert's review is necessary to provide input towards the design. Further, the evaluation stage is undertaken after analyzing empirical questions mathematically by using the Quest software program by referring to the Rasch model. The Quest program is able to do the Rasch measurement, i.e., a comprehensive empirical test of question items. There are three parameters being measured mathematically based on the empirical test of question items.

1. The first parameter is item fitness with the Rasch model by following the value of INFIT MNSQ or INFIT t of the item. The expected values of the unweighted mean square (Outfit MNSQ) in the Quest program and weighted mean square are 1; the variance is 0. On the contrary, the expected value of Mean INFIT t is equal to 0, with the variance equal to 1 (Adams & Khoo, 1996). The provision of INFIT MNSQ for the Rasch Model is shown in Table 1 and Table 2.

Tabel 1

Criteria of Question Item Fitness with the Rasch Model

MNSQ INFIT Value	Criteria
>1,33	Does Not Fit the Rasch Model
0,77 s.d. 1,33	Fits the Rasch Model
<0,77	Does Not Fit the Rasch Model

Tabel 2

The Provision of Outfit t for the Rasch Model.

t OUTFIT Value	Criteria
OUTFIT t ≤ 2,00	Fits the Rasch Model
OUTFIT t ≥ 2,00	Does Not Fit the Rasch Model

2. The second parameter is reliability. The analysis result of the Quest program also reveals the item and case reliability. The reliability value based on the item estimate is also called as sample reliability; the higher the value, the more the items that fit the tested model. Whereas, the lower the value, the less the items that fit the tested

model, so that it does not give the expected information. The reliability category is provided in Table 3 (Istiyono, 2017).

Tabel 3

Interpretation of Reliability Value

Reliability Value	Criteria
> 0,94	Excellent
0,91 – 0,94	Very Good
0,81 – 0,90	Good
0,67 – 0,80	Acceptable
< 0,67	Poor

3. The third parameter is item difficulty index and respondents' skills presented as difficulty index in the Quest output. Thresholds (THRSHL) show the item difficulty index in the logit scale along with its standard deviation (Hambleton & Rogers, 1989). The provision of the THRSHL value is given in Table 4.

Tabel 4

Criteria of THRSHL Value to Categorize Item Difficulty Level

THRSHL Value	Criteria
b > 2,00	Very Difficult
1,00 < b ≤ 2,00	Difficult
-1,00 < b ≤ 1,00	Medium
-1,00 > b ≥ 2,00	Easy
b < -2,00	Very Easy

Respondents' skills are shown by the value of the estimate error, in which the criteria of the estimate value of respondents' skills are presented in Table 5.

Tabel 5

Criteria of Estimate Value to Categorize Respondents' Skills

THRSHL Value	Criteria
b > 2,00	Very Difficult
1,00 < b ≤ 2,00	Difficult
-1,00 < b ≤ 1,00	Medium
-1,00 > b ≥ 2,00	Easy
b < -2,00	Very Easy

The evaluation stage also includes the process of analyzing the HOTS of students on the whole. The level of HOTS is categorized based on the ideal mean and standard deviation. This is applied with the assumption that students' HOTS of physics are normally distributed. The ideal mean (Im) and ideal standard deviation (Isd) are based on the highest and lowest score of research variables. Table 6 shows the criteria of students' HOTS of physics.

Tabel 6

Criteria of Students' HOTS of Physics

Interval	Criteria
$Im + 1,5\ Isb < \theta$	Very high
$Im + 0,5\ Isb < \theta \leq Im + 1,5\ Isb$	High
$Im - 0,5\ Isb < \theta \leq Im + 0,5\ Isb$	Medium
$Im - 1,5\ Isb < \theta \leq Im - 0,5\ Isb$	Low
$0 < Im - 1,5\ Isb$	Very Low

Meaning:
Im : ideal mean
Isb : ideal standard deviation
Xmak : highest score
Xmin : lowest score

RESULTS AND DISCUSSION

ADDIE development model can be used for different product developments in education, and one of which is the development of HOT skill questions. This model is simple and systematically structured in its implementation stages. The following is the description of each stage result.

Analysis

Needs analysis is the first stage being done by observation and interview to gather any information needed in the process of physics learning at the Department of Physics Education, Universitas Papua. The researcher's experience indicates that lecturers have applied HOTS learning in the classroom. However, a test to measure students' HOTS has not been conducted. The arrangement of HOTS instrument is required to train and develop students' HOTS. Accordingly, to facilitate the students in accessing other learning sources, this study designs HOT skill questions in an online system through an e-learning program using the Moodle LMS.

Design

In the design stage, the test instrument is designed based on the analysis result in the first stage. Test instrument design in this stage is in the form of question matrix and outline which are adjusted to students' needs and characteristics, and learning sources. The test is a multiple-choice test, in which 24 questions are adjusted to the formulation of a HOTS test that has been created in the test matrix and outline. The question matrix is provided in Table 7.

Tabel 7
The Question Matrix

Aspect	Sub Aspect	Theory		
		Electric current, Ohm's law, and electrical power	Series and parallel circuits of resistor and capacitor	Electric Force, Kirchoff's law, and RC circuit.
Analyze	Differentiating	8	12	21
	Organizing	3	15	20
	Attributing	2	9	23
Evaluate	Checking	4	11	22
	Critiquing	1	16	18
Create	Generating	5	13	19
	Planning	7	14	17
	Producing	6	10	24

Develop

The development of HOTS questions is based on the question matrix and outline that have been designed. Further, the questions are made online through e-learning by utilizing the Moodle LMS. Figure 2 shows all question items in the e-learning program.

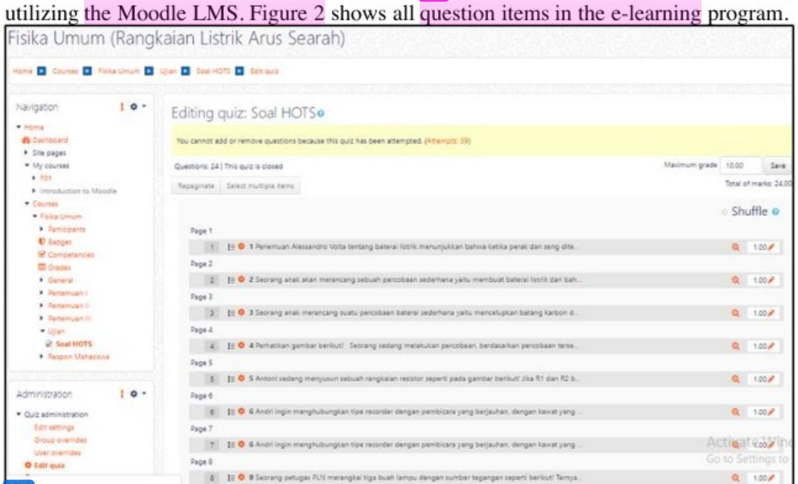

Figure 2
Shows All Question Items in the E-Learning Program

Moodle LMS program presents an interesting display and is easy to access by users (Martín-Blas & Serrano-Fernández, 2009). The questions are displayed interactively, and students can randomly work on the questions. Moodle LMS can present questions with a picture or other content to make it easier for teachers to design the questions as expected. Figure 3 shows one of the HOTS questions displayed on the e-learning through the Moodle LMS.

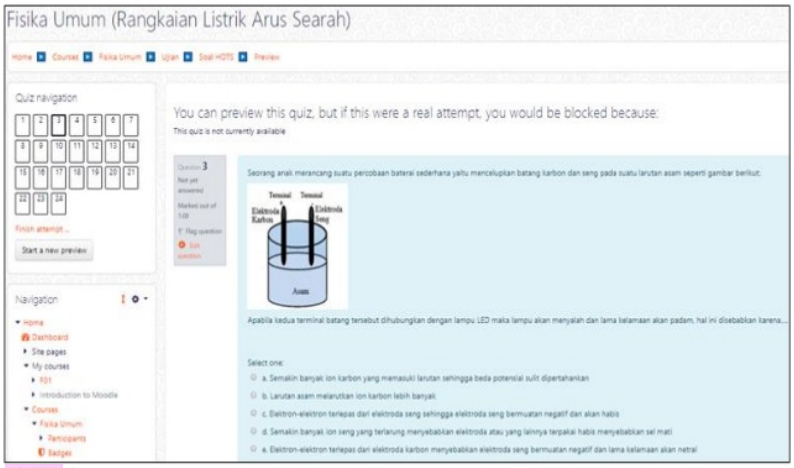

Figure 3
Shows of the HOTS Questions Displayed on the E-learning Through the Moodle LMS

The development stage aims to produce a HOTS test instrument that has been validated by experts and practitioners. Product validation is a process of assessing the designed product, or in this case, the test instrument of HOTS in general physics subject in the site area. Product validation is carried out by involving seven validators, i.e., experts of measurement, physics education, physics, and practitioners. The validity test of the instrument includes material, construction, and language. The analysis result of question validity that is assessed by validators obtains the value of V Aiken in the range of 0.76 - 1.00, showing a valid result. The questions validated by experts and practitioners are then revised based on provided corrections and suggestions.

Implementation

The implementation stage in this study is the product trial, in which HOTS questions are tried out to 34 students in the research site. The students work on these questions via online through e-learning by using their own Moodle account upon the completion of all learning stages. Results of the students' learning can be accessed after this process.

Evaluation

Before conducting the estimate analysis of respondents' skills and item difficulty level, the analysis of item fitness is performed by using parameters of INFIT and OUTFIT for mean square and t. The determination of the item fitness with the model is based on the value of INFIT MNSQ and the standard deviation or Infit t (Adams & Khoo, 1996). The fitness of each case is also based on the value of INFIT MNSQ or INFIT t of the item. Table 8 provides the testing result through the Quest program to obtain the values of item estimate and case estimate in the HOTS questions trial.

Tabel 8

Values of Item Estimate and Case Estimate in the HOTS Questions Trial

No.	Measurement	Estimates for Items	Estimates for Testi
1.	Average values and standard deviations	0,00 ± 0,57	0,01 ± 1,24
2.	Reliability Estimates	0,66	0,85
3.	The mean value and standard deviation of INFIT MNSQ	1,00 ± 0,14	0,99 ± 0,15
4.	The mean value and standard deviation of OUTFIT MNSQ	1,09 ± 0,52	1,09 ± 0,52
5.	The mean value and standard deviation of INFIT t	-0,03 ± 0,81	0,00 ± 0,72
6.	The mean value and standard deviation of OUTFIT t	0,21 ± 0,91	0,17 ± 0,81

The analysis result reveals that the INFIT MNSQ arrives at the range of 0.86 - 1.14, and INFIT t is -0.28 - 0.72. This result signifies that all 24 questions fit the model as they reach the range of INFIT MNSQ value from 0.77 to 1.30 and use INFIT t with the limit of -2.0 - 2.0 [16]. In addition to testing the fitness, the output of the Quest program also presents the reliability estimate of the test instrument. Table 8 provides the value of item reliability based on the value of summary of item estimate, which is 0.66. On the other hand, the value of person reliability, as based on the summary of case estimate, gets 0.85. These results are in line with the Rasch model, in which the reliability value falls under the range of 0.67 - 0.80 (quite reliable). On that ground, the instrument can be used to measure students' HOTS in the General Physics subject.

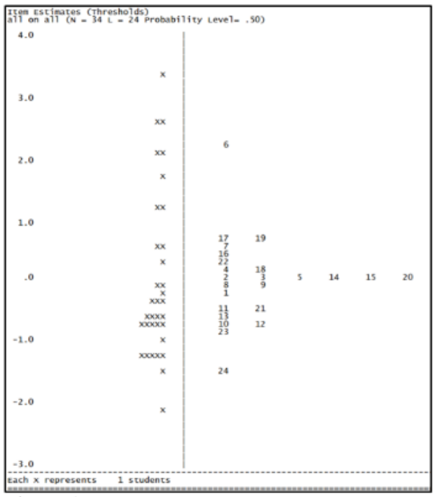

Figure 4
Distribution of Item Difficulty Level and Respondents' Skills

Figure 4 presents the distribution of the respondents according to the difficulty level in the logit scale from -4.0 to +4.0. This map displays the item difficulty level compared to the respondents' skills. Case and item difficulty levels in the Rasch model are expressed in one line in the form of abscissa in the graph with logg-odd unit. The graph of respondents' skills shows a normal curve, meaning that there are only a few respondents with low and high skills; and a lot of respondents with moderate skills. The level of item difficulty of threshold reveals that item 6 is the most difficult question, and item 24 is the easiest one.

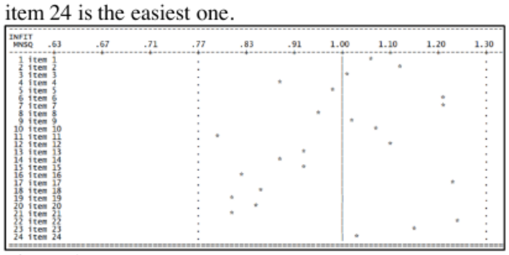

Figure 5
Distribution of INFIT MNSQ Values of Each Question Item of HOTS

Question items that fit the Rasch model are in the range of 0.77 - 1.33. Figure 5 shows that all 24 question items are in the line, implying that they fit the Rasch model.

Item Estimates (Thresholds) In input Order
all on all (N = 34 L = 24 Probability Level= .50)

	ITEM NAME	SCORE	MAXSCR	THRSH 1	INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
1	item 1	18	34	-.26 .39	1.06	1.15	.4	.5
2	item 2	16	34	.04 .40	1.12	1.17	.7	.6
3	item 3	16	34	.04 .40	1.01	.91	.1	-.2
4	item 4	15	34	.19 .40	.88	.93	-.6	-.1
5	item 5	16	34	.04 .40	.98	.89	.0	-.2
6	item 6	5	34	2.27 .57	1.21	2.16	.7	1.4
7	item 7	13	34	.52 .42	1.21	1.27	1.0	.9
8	item 8	17	34	-.11 .40	.96	1.00	-.2	.1
9	item 9	17	34	-.11 .40	1.02	.91	.2	-.2
10	item 10	21	34	-.70 .39	1.07	1.16	.6	.5
11	item 11	19	34	-.41 .39	.79	.66	-1.6	-.9
12	item 12	21	34	-.70 .39	1.10	1.14	.8	.5
13	item 13	20	34	-.55 .39	.93	1.09	-.5	.4
14	item 14	16	34	.04 .40	.88	.78	-.7	-.6
15	item 15	16	34	.04 .40	.93	.82	-.4	-.5
16	item 16	13	33	.47 .42	.82	.69	-.8	-.9
17	item 17	12	34	.69 .43	1.23	1.16	1.0	.6
18	item 18	15	34	.19 .40	.86	.73	-.8	-.8
19	item 19	12	34	.69 .43	.81	.71	-.8	-.8
20	item 20	16	34	.04 .40	.85	.75	-.9	-.7
21	item 21	19	34	-.41 .39	.81	.68	-1.4	-.8
22	item 22	14	34	.35 .41	1.24	1.23	1.2	.8
23	item 23	22	34	-.85 .40	1.15	3.04	1.1	3.1
24	item 24	26	34	-1.50 .43	1.03	1.02	.2	.2
Mean				.00	1.00	1.09	.0	.1
SD				.71	.14	.52	.8	.9

Figure 6
Item Estimates from HOTS Questions

Figure 6 presents the Item Estimate of HOT skill questions based on the trial result. In this figure, there is SCORE-MAXSCR successively showing the score of the respondents who answer correctly, and the number of total respondents. Item 24 is the most correctly-answered, in which 26 out of 34 respondents are able to work on this item. Figure 6 also provides the value of THRSHL that shows the item difficulty index in the logit scale along with its standard deviation. Item 6 has THRSHL or difficulty index of 2.27 that is greater than 2.0, or in other words, this item is very difficult since only five students can give a correct answer. The average value of THRSHL and its standard deviation accounts for 0.00 ± 0.71 and falls into the range of -2 - 2 (Hambleton & Rogers, 1989). The average value of INFIT MNSQ is 1.00 ± 0.14 and falls under the acceptance range of 0.77 - 1.33; the average value of OUTFIT t arrives at 0.10 ± 0.90 and falls into the acceptance range of ⩽ 2.00. All of these results indicate that all question items that have been developed can be employed to measure students' HOTS.

Case Estimates In input Order
all on all (N = 34 L = 24 Probability Level= .50)

NAME	SCORE	MAXSCR	ESTIMATE	ERROR	INFIT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
1 01	12	24	-.02	.43	1.06	1.02	.58	.18
2 02	6	24	-1.21	.49	1.17	1.09	.75	.36
3 03	8	24	-.77	.45	.98	1.01	-.06	.13
4 04	8	24	-.77	.45	.83	.81	-1.04	-.48
5 05	8	24	-.77	.45	.89	.83	-.59	-.41
6 06	6	24	-1.21	.49	.79	.70	-.84	-.67
7 07	10	24	-.38	.43	.99	.95	-.01	-.09
8 08	6	24	-1.21	.49	1.07	2.30	.36	2.44
9 09	3	24	-2.10	.63	.98	.85	.11	.00
10 10	9	24	-.57	.44	.88	.83	-.80	-.48
11 11	22	24	2.61	.77	.73	.46	-.30	-.56
12 12	5	24	-1.46	.52	.89	.85	-.29	-.18
13 13	20	24	1.75	.57	1.21	1.45	.64	.93
14 14	11	24	-.20	.43	.86	.83	-1.33	-.59
15 15	21	24	2.12	.64	1.18	1.05	.52	.29
16 16	9	24	-.57	.44	1.08	1.06	.59	.28
17 17	7	24	-.98	.47	1.29	2.20	1.38	2.55
18 18	6	24	-1.21	.49	1.23	1.28	.96	.75
19 19	14	24	.35	.43	.92	.87	-.56	-.40
20 20	15	24	.54	.44	.97	1.09	-.13	.40
21 21	18	24	1.19	.49	.94	.86	-.16	-.25
22 22	21	24	2.12	.64	.93	1.23	-.01	.55
23 23	9	24	-.57	.44	1.07	1.01	.54	.15
24 24	8	24	-.77	.45	1.01	.95	.13	-.03
25 25	10	24	-.38	.43	.87	.82	-1.06	-.57
26 26	15	24	.54	.44	1.05	1.22	.36	.80
27 27	6	24	-1.21	.49	.82	.74	-.69	-.56
28 28	22	24	2.61	.77	.73	.46	-.30	-.56
29 29	12	24	-.02	.43	.92	.88	-.73	-.39
30 30	9	24	-.57	.44	.90	.90	-.64	-.23
31 31	23	24	3.40	1.05	1.18	3.14	.49	1.53
32 32	18	24	1.19	.49	.85	.75	-.53	-.58
33 33	10	23	-.32	.44	1.11	1.11	.97	.45
34 34	8	24	-.77	.45	1.29	1.34	1.61	1.03
Mean			.01		.99	1.09	.00	.17
SD			1.35		.15	.52	.72	.81

Figure 7
Case Etimates from Every Student

Figure 7 serves as the case estimate or the skill level of each student. Information obtained from the case etimate is that the SCORE-MAXSCR shows the score of each respondent from the maximum score sequentially. Respondent 31 answers the most questions (23 out of 24 questions) correctly compared to other respondents. The average estimate value and its standard deviation gets 0.01 ± 1.35 and falls under a moderate category. The analysis result of the case estimate reveals that students' skills are in the moderate category.

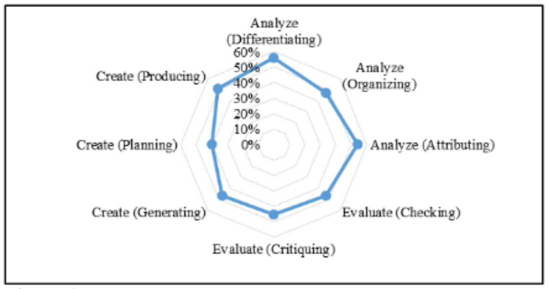

Figure 8
Distribution of Student Answer Percentage HOTS

Figure 8 gives the percentage of students' answers based on the aspects and sub-aspects of HOTS. The analysis result brings out the fact that students tend to find it difficult to answer questions regarding the creating aspect, especially the planning sub-aspect. Creating is the highest level HOTS in Bloom's taxonomy, which therefore, students need to practice developing their creating skills. This figure also signifies that the majority of the students find it easy to answer HOTS questions related to the analysis aspect, differentiating sub-aspect in particular.

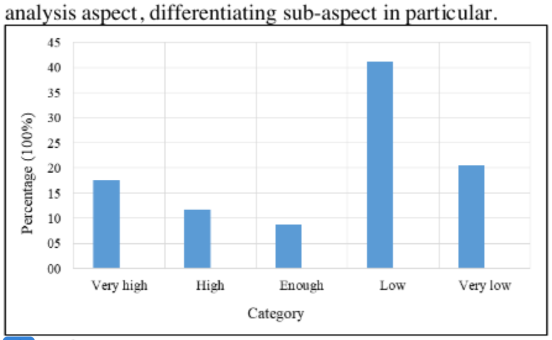

Figure 9
Percentage of Students' HOTS

Figure 9 shows the percentage of students' HOTS. It is seen that most students (41.2%) still have low HOTS; the categories consist of very low (20.6%), moderate (8.8%), high (11.8%), and very high (17.6%). The low category of students' HOTS is influenced by several factors, one of which is that the students are not used to working on HOTS questions (Tanujaya, Mumu, & Margono, 2017; Yusuf & Widyaningsih, 2019). They need to practice developing their HOTS by being exposed to HOTS-based learning sources.

CONCLUSION

Test characteristics comprise item fitness, reliability, and difficulty. Dichotomy data analysis used the Rasch Model through the Quest program. The trial result obtained the criteria of INFIT MNSQ mean and standard deviation of 1.0 and 0.0, respectively, showing that the items fit the RM1-PL. In addition, the value of item reliability based on the value of summary of item estimate arrives at 0.66; meanwhile, the person reliability under the summary of case estimate reaches 0.85, i.e., the reliability value is in the range of 0.67 - 0.80 (quite reliable). As based on the criteria of minimum and maximum INFIT MNSQ of 0.77 and 1.30, 24 question items fit the RM 1-PL model. The result of the Quest output also reveals that the average value of THRSHL and its standard deviation is 0.00 ± 0,71, or in the acceptance range of -2 to 2. To sum up, all 24 question items that had been tried out have fit the model with a good category, so that they can be utilized in HOTS measurement.

ACKNOWLEDGMENT

We would like to acknowledge the contribution of the Ministry of Research and Higher Education in funding this study through Inter Higher Education Institution Cooperation scheme with the contract number: 198/SP2H/AMD/LT/DRPM/2020.

The Development of HOTS Test of Physics Based on the Modern Test Theory: Question Modeling through E-learning of Moodle LMS

ORIGINALITY REPORT

13%	4%	5%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Syiah Kuala University Student Paper	2%
2	Submitted to Universitas Pendidikan Ganesha Student Paper	2%
3	www.e-iji.net Internet Source	2%
4	aip.scitation.org Internet Source	1%
5	Submitted to Universitas Pendidikan Indonesia Student Paper	1%
6	Newman C Enyioko, Gabriel A Okwandu. "Effect of Social Media Marketing on The Conversion Rate of Deposit Money Banks in Nigeria", JEMA: Jurnal Ilmiah Bidang Akuntansi dan Manajemen, 2019 Publication	1%
7	Irfan Yusuf, Sri Wahyu Widyaningsih.	

"Implementing E-Learning-Based Virtual Laboratory Media to Students' Metacognitive Skills", International Journal of Emerging Technologies in Learning (iJET), 2020 <1%
Publication

8 Submitted to Program Pascasarjana Universitas Negeri Yogyakarta <1%
Student Paper

9 Submitted to University of Lancaster <1%
Student Paper

10 Submitted to Universiti Tenaga Nasional <1%
Student Paper

11 www.researchgate.net <1%
Internet Source

12 Submitted to Widener University <1%
Student Paper

13 Submitted to IAIN Syaikh Abdurrahman Siddik Bangka Belitung <1%
Student Paper

14 Fajar Nur Cahyani, Bambang Subali. "The Validity of Item Tests of Scientific Method of Divergent Patterns in Biology Subject at Senior High Schools Viewed From The Number of Traits", Journal of Physics: Conference Series, 2019 <1%
Publication

15	Submitted to Universiti Sultan Zainal Abidin Student Paper	<1%
16	www.stran.ac.uk Internet Source	<1%
17	S W Widyaningsih, Mujasam, I Yusuf. "Effect of RMS models (reading, mind mapping and sharing) based on e-learning to the learning outcomes of students in the teaching and learning subject", Journal of Physics: Conference Series, 2019 Publication	<1%
18	Sri Wahyu Widyaningsih, Irfan Yusuf, Insar Damopolii. "PELATIHAN PENELITIAN TINDAKAN KELAS BAGI GURU IPA KABUPATEN MANOKWARI PAPUA BARAT", JMM (Jurnal Masyarakat Mandiri), 2019 Publication	<1%
19	digital.library.adelaide.edu.au Internet Source	<1%
20	Submitted to Leiden University Student Paper	<1%
21	www.iarjournal.com Internet Source	<1%
22	Submitted to Universitas Muhammadiyah Surakarta Student Paper	<1%

23	Submitted to RMIT University Student Paper	<1%
24	Submitted to Surabaya University Student Paper	<1%
25	H U Amrina, H Sutrisno. "Application of guided inquiry learning model based on inter-intrapersonal intelligence in chemistry classroom", Journal of Physics: Conference Series, 2020 Publication	<1%
26	"Innovations in Intelligent Machines – 2", Springer Science and Business Media LLC, 2012 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		