



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

### **Sri Wahyu Widyaningsih**

Asst. Prof., corresponding author, Faculty of Teacher Training and Education, Universitas Papua, Indonesia, [s.widyaningsih@unipa.ac.id](mailto:s.widyaningsih@unipa.ac.id)

### **Irfan Yusuf**

Asst. Prof., Faculty of Teacher Training and Education, Universitas Papua, Indonesia, [i.yusuf@unipa.ac.id](mailto:i.yusuf@unipa.ac.id)

### **Zuhdan Kun Prasetyo**

Prof., Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Indonesia, [zuhdan@uny.ac.id](mailto:zuhdan@uny.ac.id)

### **Edi Istiyono**

Prof., Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Indonesia, [edi\\_istiyono@uny.ac.id](mailto:edi_istiyono@uny.ac.id)

This research aims to develop HOTS physics questions based on Modern Test Theory designed and presented with LMS Moodle on e-learning, which can be accessed online. This study also serves as one of the efforts to expand students' HOTS by applying a variety of HOTS-based learning sources. Further, this research employed the ADDIE model with analysis, design, development, implementation, and evaluation stages. The instrument consisted of 24 multiple-choice physics questions; the questions were designed by following the aspects and sub-aspects of HOTS and validated by the assessment of physics experts, physicists, and lecturers. Moreover, the validity analysis was based on Aiken's V formula, in which every aspect was confirmed to be valid. The instrument had been tested on 34 students of the Physics Education Department, Universitas Papua. 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, indicating that the items fitted the RM1-PL. In addition, the value of item reliability based on the item estimate summary arrived at 0.66; meanwhile, the case reliability under the summary of the case estimate accounted for 0.85.

**Keywords:** e-learning, HOTS test, modern test theory, physics, test

**Citation:** Widyaningsih, S. W., Yusuf, I., Prasetyo, Z. K., & Istiyono, E. (2021). The development of the HOTS test of physics based on modern test theory: Question Modeling through e-learning of moodle LMS. *International Journal of Instruction*, 14(4), 51-68. <https://doi.org/10.29333/iji.2021.1444a>

## **INTRODUCTION**

Assessment, especially 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 or measured. Students' cognitive skills in the learning process can be categorized into Lower-Order Thinking Skills (LOTS) and Higher-Order Thinking Skills (HOTS). The LOTS includes remembering, understanding, and applying; the HOTS, on the other hand, consists of analyzing, evaluating, and creating. HOTS is thinking skills that require not only the remembering skill but also other higher skills. Indicators to measure HOTS encompass analyzing (C4), evaluating (C5), and creating (C6) skills (Krathwohl & Anderson, 2010).

HOTS also refers 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, comprising 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 LOTS, but also HOTS. Teachers are also urged to deliver learning materials to students, including the HOTS, that can be improved by the HOTS instrument. A previous study has reported that the majority of teachers find it challenging to formulate an assessment instrument of learning outcomes, HOTS questions, in particular (Istiyono, 2018). For this reason, teachers' creativity is highly required to measure student learning outcomes. Today's development of 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 the evaluation of student learning outcomes.

Evaluation questions can be posed in an integrated manner through e-learning systems, such as Moodle Learning Management System (LMS) (Azevedo, 2015; Bogdanović, Barać, Jovanić, Popović, & Radenković, 2014). The Moodle provides different types of questions, namely multiple choices, true or false, and short answers; these are stored in the taught course database and can be reapplied (Limongelli, Sciarrone, & Vaste, 2011). Teachers are also able to offer 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 figure out their assessment results right away.

Teachers need to prepare a good test to measure student learning outcomes. There are two paradigms developed to assess student learning outcomes through the used 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 its limitations in measuring the item difficulty level and discrimination since both indicators' calculation is based on the test taker's total score. In contrast, the IRT frees up the dependence between the test item and the 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 (Raykov & Marcoulides, 2015). Therefore, the application answers the needs of modern measurement to date, i.e., comparing test taker's skills, question development, and even adaptive test development. It is considered able to overcome the limitations of the classical test theory.

On account of the simplicity of the analysis, most teachers have analyzed assessment tools using classical analysis techniques. The use of classical analytical techniques features some limitations, including the difficulty of defining individual learners' skills. The calculated error of measurement does not include persons but groups together. This is because each test taker's response to the questions cannot be clarified by classical test theory. Efforts are thereby required to free the measuring tool from attachment to the sample (sample-free) employing the IRT.

This is a preliminary 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. This research aims to develop HOTS physics questions based on IRT designed and presented with LMS Moodle on e-learning, which can be accessed online.

## METHOD

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

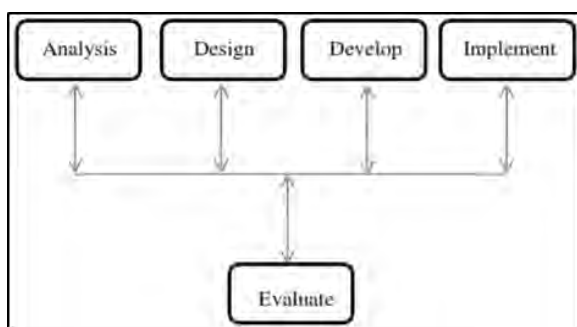


Figure 1  
Stages of ADDIE development model in designing moodle LMS-based HOTS test

**Analysis**

The analysis stage was a process of needs analysis to determine test objectives, identify problems, analyze tasks, and determine question formats to be applied. It was shown that the problems were related to the needs of HOTS instrument design for students at the Department of Physics Education, Universitas Papua.

**Design**

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

**Develop**

Every single thing required in the arrangement of HOTS questions has been prepared in the next stage. This stage also covered the process of making the questions regarding HOTS, as well as validating the questions that involved the experts of measurement, physics education, and practitioners. The validity analysis technique to assess the content validity of the developed questions relied on the Aiken's V formula (Aiken, 1980, 1985).

$$V = \frac{\sum s}{n(c-1)} \quad (1)$$

"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 Aiken's V index falls under the range of 0.37 to 1.00 (Kowsalya, Venkat Lakshmi, & Suresh, 2012). The value of Aiken's V of every test item was calculated based on the assessment items of every validator. In this stage, there was also an evaluation process, i.e., revising questions by following validators' corrections and suggestions.

**Implementation**

Another stage was applying HOTS questions that had been developed to 34 students in the site area who enrolled in general physics subject. This number followed 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, Yan, & Heene, 2020). Question item analysis was performed based on the raw score of the students by employing the Quest program.

**Evaluation**

The evaluation was a process of finding out whether HOTS's developed questions had met the expectation. The evaluation stage is carried out in every stage and is 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. Besides, the evaluation stage was undertaken after analyzing empirical questions mathematically by using the Quest software program by referring to the Rasch model. The Quest program can do the Rasch measurement, i.e., a comprehensive empirical test

of question items. There were three parameters being measured mathematically based on the empirical test of question items, as follows.

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 presented in Table 1 and Table 2 below.

Table 1  
Criteria of question item fitness with the rasch model

| MNSQ INFIT Value | Criteria                     |
|------------------|------------------------------|
| >1.33            | Does Not Fit the Rasch Model |
| 0.77 to 1.33     | Fits the Rasch Model         |
| <0.77            | Does Not Fit the Rasch Model |

Table 2  
The provision of outfit t for the rasch model.

| t OUTFIT Value       | Criteria                     |
|----------------------|------------------------------|
| OUTFIT $t \leq 2.00$ | Fits the Rasch Model         |
| OUTFIT $t \geq 2.00$ | Does Not Fit the Rasch Model |

2. The second parameter is reliability. The analysis result of the Quest program also showed the item and case reliability. The reliability value based on the item estimate is also called 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 the following table (Istiyono, 2017).

Table 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       | Fair      |
| < 0.67            | Poor      |

3. The third parameter is the 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 in Table 4.

Table 4  
Criteria of THRSHL value to categorize item difficulty level

| THRSHL Value          | Criteria       |
|-----------------------|----------------|
| $b > 2.00$            | Very Difficult |
| $1.00 < b \leq 2.00$  | Difficult      |
| $-1.00 < b \leq 1.00$ | Medium         |
| $-1.00 > b \geq 2.00$ | Easy           |
| $b < -2.00$           | Very Easy      |

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

Table 5  
Criteria of estimate value to categorize respondents' skills

| THRSHL Value          | Criteria       |
|-----------------------|----------------|
| $b > 2.00$            | Very Difficult |
| $1.00 < b \leq 2.00$  | Difficult      |
| $-1.00 < b \leq 1.00$ | Medium         |
| $-1.00 > b \geq 2.00$ | Easy           |
| $b < -2.00$           | Very Easy      |

The evaluation stage also included 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 was applied with the assumption that students' HOTS of physics were normally distributed. The ideal mean ( $I_m$ ) and ideal standard deviation ( $I_{sd}$ ) are based on the highest and lowest score of research variables. Table 6 shows the criteria of students' HOTS of physics.

Table 6  
Criteria of students' HOTS of physics

| Interval  | Criteria  |
|---|-----------|
| $I_m + 1.5 I_{sb} < \theta$                       | Very high |
| $I_m + 0.5 I_{sb} < \theta \leq I_m + 1.5 I_{sb}$ | High      |
| $I_m - 0.5 I_{sb} < \theta \leq I_m + 0.5 I_{sb}$ | Moderate  |
| $I_m - 1.5 I_{sb} < \theta \leq I_m - 0.5 I_{sb}$ | Low       |
| $0 < I_m - 1.5 I_{sb}$                            | Very Low  |

Meaning:

$I_m$  : ideal mean

$I_{sb}$  : ideal standard deviation

$X_{mak}$  : highest score

$X_{min}$  : lowest score

## FINDINGS

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

### Analysis

A needs analysis was the first stage being done by observation and interview to gather any information required in physics learning at the Department of Physics Education, Universitas Papua. The researchers' experience indicated that the lecturers had 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 designed HOTS questions in an online system through an e-learning program using the Moodle LMS.

### Design

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

Table 7  
The question matrix

| Aspects  | Sub Aspects     | Theories  |  |   |
|----------|-----------------|---|--|---|
|          |                 | 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 was based on the question matrix and outline that had been designed. In addition, the questions were formulated online through e-learning by utilizing the Moodle LMS. Figure 2 below shows all question items in the e-learning program.

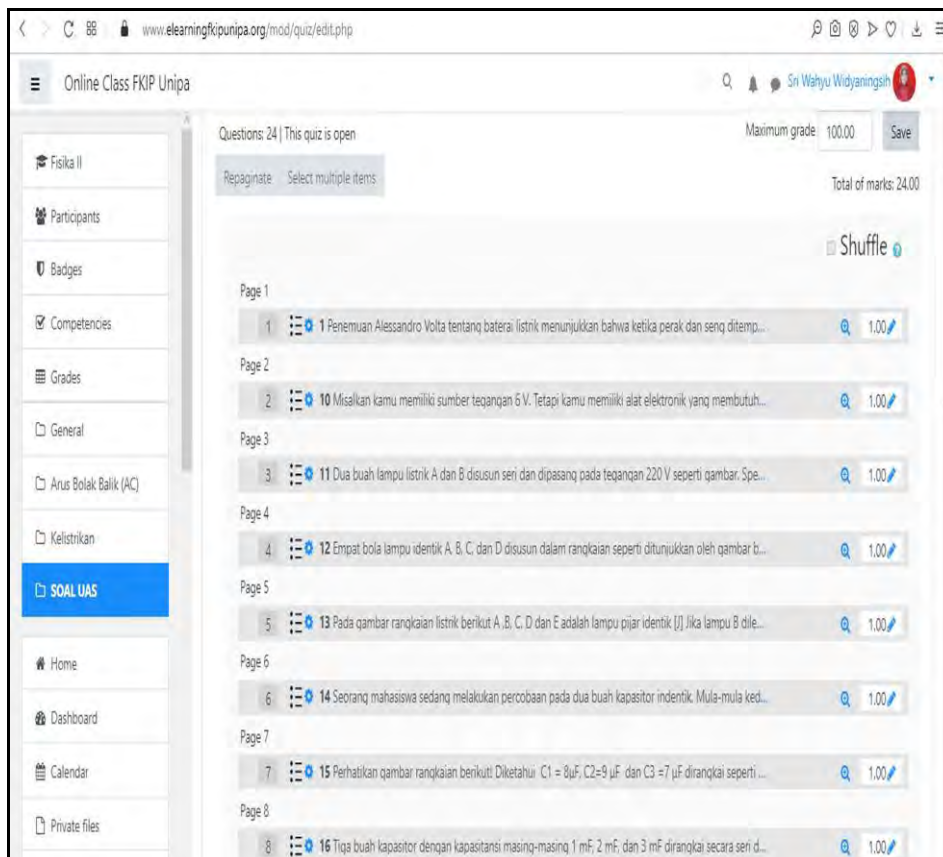


Figure 2  
All question items in the e-learning program

The questions are displayed interactively, and students can randomly work on the questions. Moodle LMS can present questions with a picture or other contents to make it easier for teachers to design the questions as expected. Figure 3 illustrates one of the HOTS questions displayed on the e-learning through the Moodle LMS.



Question 18  
Not yet answered  
Marked out of 1.00  
Flag question  
Edit question

Seorang anak merancang suatu percobaan baterai sederhana yaitu mencelupkan batang karbon dan seng pada suatu larutan asam seperti gambar berikut.

Terminal Terminal  
Elektroda Karbon Elektroda Seng  
Asam

Apabila kedua terminal batang tersebut dihubungkan dengan lampu LED maka lampu akan menyala dan lama kelamaan akan padam, hal ini disebabkan karena....

Select one:

- a. Semakin banyak ion seng yang terlarung menyebabkan elektroda atau yang lainnya terpakai habis menyebabkan sel mati
- b. Larutan asam melarutkan ion karbon lebih banyak
- c. Elektron-elektron terlepas dari elektroda seng sehingga elektroda seng bermuatan negatif dan akan habis
- d. Semakin banyak ion karbon yang memasuki larutan sehingga beda potensial sulit dipertahankan
- e. Elektron-elektron terlepas dari elektroda karbon menyebabkan elektroda seng bermuatan negatif dan lama kelamaan akan netral

Previous page Next page

Quiz navigation  
1 2 3 4 5  
6 7 8 9 10  
11 12 13 14 15  
16 17 18 19 20  
21 22 23 24  
Finish attempt...  
Start a new preview

Figure 3  
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 was carried out by involving seven validators, i.e., experts of measurement, physics education, physics, and practitioners. The validity test of the instrument included material, construction, and language. The analysis result of the question validity assessed by validators obtained the value of Aiken's  $V$  in the range of 0.76 to 1.00, showing a valid result. The questions validated by experts and practitioners were then revised following the provided corrections and suggestions.

### Implementation

The implementation stage in this study was the product trial, in which HOTS questions were tried out to 34 students in the research site. The students worked on these questions online through e-learning by using their own Moodle account upon completing 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 was performed using 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.

Table 8

Values of item estimate and case estimate in the HOTS questions trial

| No. | Measurement                                    | Estimates for Items | Estimates for Testing |
|-----|--|---------------------|-----------------------|
| 1.  | Average values and standard deviations         | $0.00 \pm 0.57$     | $0.01 \pm 1.24$       |
| 2.  | Reliability Estimates                          | 0.66                | 0.85                  |
| 3.  | The mean and standard deviation of INFIT MNSQ  | $1.00 \pm 0.14$     | $0.99 \pm 0.15$       |
| 4.  | The mean and standard deviation of OUTFIT MNSQ | $1.09 \pm 0.52$     | $1.09 \pm 0.52$       |
| 5.  | The mean and standard deviation of INFIT t     | $-0.03 \pm 0.81$    | $0.00 \pm 0.72$       |
| 6.  | The mean and standard deviation of OUTFIT t    | $0.21 \pm 0.91$     | $0.17 \pm 0.81$       |

The analysis result suggested that the INFIT MNSQ got the range of 0.86 to 1.14, and INFIT t is -0.28 to 0.72. This signified that all 24 questions fit the model as they reached the range of INFIT MNSQ value from 0.77 to 1.30 and used INFIT t with the limit of -2.0 to 2.0. In addition to testing the fitness, the Quest program's output also presented the reliability estimate of the test instrument. The above table shows the value of item reliability based on the value of the item estimate summary, which is 0.66. On the other hand, the value of person reliability, as based on the case estimate summary, gets 0.85. These results were in line with the Rasch model, in which the reliability value fell under the range of 0.67 to 0.80 (quite reliable). On that ground, the instrument can be employed to measure students' HOTS in the General Physics subject.

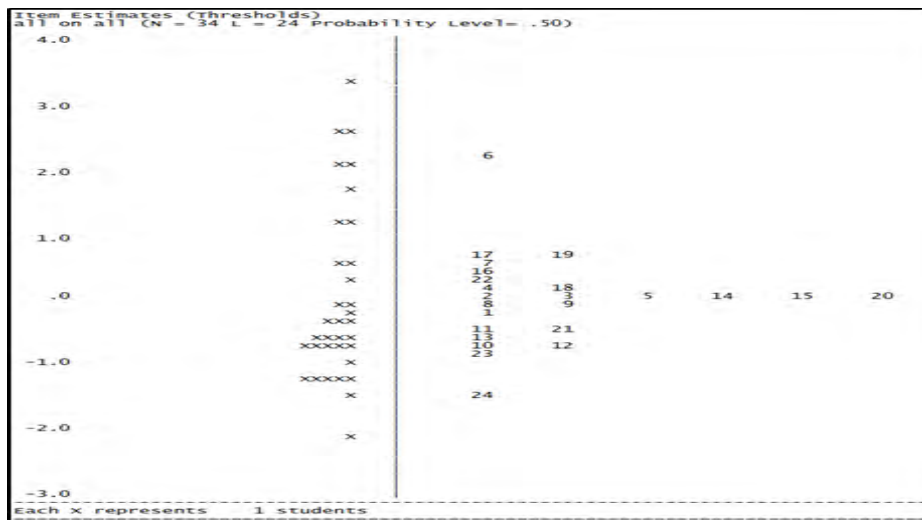


Figure 4

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 a log-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 many respondents with moderate skills. The level of item difficulty of threshold revealed that item 6 was the most difficult question, and item 24 was the easiest one.

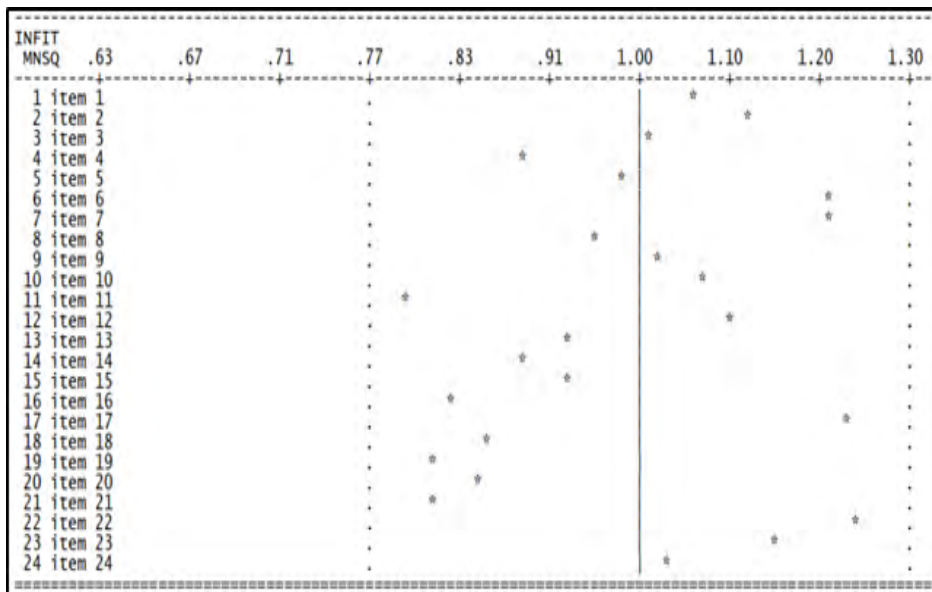


Figure 5  
Distribution of INFIT MNSQ values of each question item of HOTS

Question items that fit the Rasch model are in the range of 0.77 to 1.33. By referring to Figure 5, we can see that all 24 question items are in the line, implying that they fit the Rasch model.

| Item Estimates (Thresholds) In input Order<br>all on all (N = 34 L = 24 Probability Level = .50) |       |        |              |              |               |           |            |  |
|--|-------|--------|--------------|--------------|---------------|-----------|------------|--|
| ITEM NAME  | SCORE | MAXSCR | THRSH<br>1   | INFT<br>MNSQ | OUTFT<br>MNSQ | INFT<br>t | OUTFT<br>t |  |
| 1 item 1   | 18    | 34     | -.26<br>.39  | 1.06         | 1.15          | .4        | .5         |  |
| 2 item 2   | 16    | 34     | -.04<br>.40  | 1.12         | 1.17          | .7        | .6         |  |
| 3 item 3   | 16    | 34     | -.04<br>.40  | 1.01         | .91           | .1        | -.2        |  |
| 4 item 4   | 15    | 34     | -.19<br>.40  | .88          | .93           | -.6       | -.1        |  |
| 5 item 5   | 16    | 34     | -.04<br>.40  | .98          | .89           | .0        | -.2        |  |
| 6 item 6   | 5     | 34     | 2.27<br>.57  | 1.21         | 2.16          | .7        | 1.4        |  |
| 7 item 7   | 13    | 34     | -.52<br>.42  | 1.21         | 1.27          | 1.0       | .9         |  |
| 8 item 8   | 17    | 34     | -.11<br>.40  | .96          | 1.00          | -.2       | .1         |  |
| 9 item 9   | 17    | 34     | -.11<br>.40  | 1.02         | .91           | .2        | -.2        |  |
| 10 item 10   | 21    | 34     | -.70<br>.39  | 1.07         | 1.16          | .6        | .5         |  |
| 11 item 11   | 19    | 34     | -.41<br>.39  | .79          | .66           | -1.6      | -.9        |  |
| 12 item 12   | 21    | 34     | -.70<br>.39  | 1.10         | 1.14          | .8        | .5         |  |
| 13 item 13   | 20    | 34     | -.55<br>.39  | .93          | 1.09          | -.5       | .4         |  |
| 14 item 14   | 16    | 34     | -.04<br>.40  | .88          | .78           | -.7       | -.6        |  |
| 15 item 15   | 16    | 34     | -.04<br>.40  | .93          | .82           | -.4       | -.5        |  |
| 16 item 16   | 13    | 33     | -.47<br>.42  | .82          | .69           | -.8       | -.9        |  |
| 17 item 17   | 12    | 34     | -.69<br>.43  | 1.23         | 1.16          | 1.0       | .6         |  |
| 18 item 18   | 15    | 34     | -.19<br>.40  | .86          | .73           | -.8       | -.8        |  |
| 19 item 19   | 12    | 34     | -.69<br>.43  | .81          | .71           | -.8       | -.8        |  |
| 20 item 20   | 16    | 34     | -.04<br>.40  | .85          | .75           | -.9       | -.7        |  |
| 21 item 21   | 19    | 34     | -.41<br>.39  | .81          | .68           | -1.4      | -.8        |  |
| 22 item 22   | 14    | 34     | -.35<br>.41  | 1.24         | 1.23          | 1.2       | .8         |  |
| 23 item 23   | 22    | 34     | -.85<br>.40  | 1.15         | 3.04          | 1.1       | 3.1        |  |
| 24 item 24   | 26    | 34     | -1.50<br>.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 of HOTS questions

The previous figure presents the Item Estimate of HOTS questions based on the trial result. In this figure, there is SCORE-MAXSCR successively showing the respondents who answer correctly and the number of total respondents. Item 24 was the most correctly-answered, in which 26 out of 34 respondents could 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 got a THRSHL or difficulty index of 2.27 that was greater than 2.0, or in other words, this item was very difficult since only five students could give a correct answer. Also, the average value of THRSHL and its standard deviation accounted for  $0.00 \pm 0.71$  and fell under the range of -2 to 2 (Hambleton & Rogers, 1989). The average value of INFIT MNSQ was  $1.00 \pm 0.14$  and achieved the acceptance range of 0.77 to 1.33; the average value of OUTFIT t arrived at  $0.10 \pm 0.90$  and was included in the acceptance range of  $\leq 2.00$ . Accordingly, these results indicate that all question items being developed can be utilized 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 | OUTFIT 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  | 10    | 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         | -.30   | -.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 | 7     | 24     | -.57     | .44   | 1.08       | 1.06        | -.59   | -.28    |
| 17 17 | 9     | 24     | -.98     | .47   | 1.29       | 2.20        | 1.38   | 2.92    |
| 18 18 | 6     | 24     | -1.21    | .49   | 1.23       | 1.28        | .96    | 1.45    |
| 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   | -.50    |
| 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   | -.50    |
| 28 28 | 22    | 24     | -2.61    | .77   | .73        | .46         | -.30   | -.56    |
| 29 29 | 12    | 24     | -.02     | .43   | .92        | .88         | -.73   | -.30    |
| 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   | -.53    |
| 33 33 | 10    | 24     | -.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 estimates of every student

Figure 7 serves as the case estimate or the skill level of each student. Information obtained from the case estimate is that the SCORE-MAXSCR shows each respondent's score from the maximum score sequentially. Respondent 31 answered the majority of the questions (23 out of 24 questions) correctly compared to other respondents. The average estimate value and its standard deviation got  $0.01 \pm 1.35$  and were in a moderate category. The analysis result of the case estimate revealed that students' skills were in the moderate category.

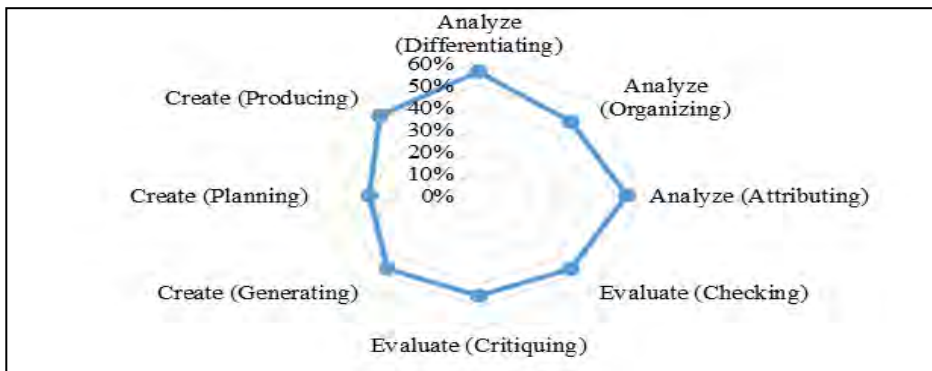


Figure 8  
Distribution of students' answer percentage HOTS

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

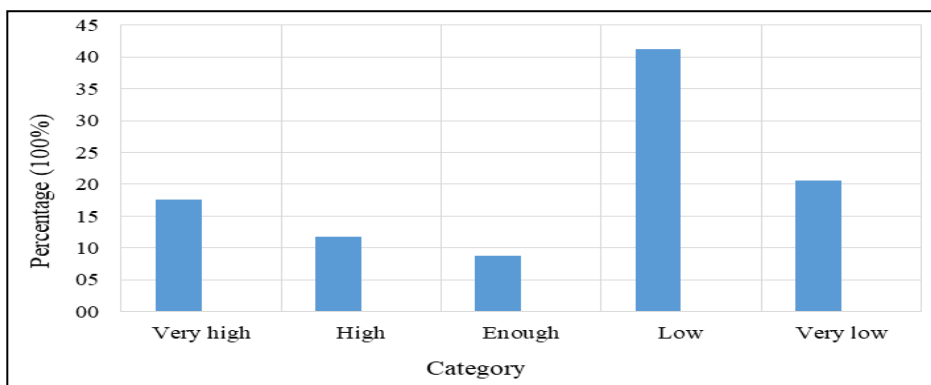


Figure 9  
Percentage of students' HOTS

The above figure 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%).

## DISCUSSION

This study aims to produce the HOTS instrument presented in e-learning using Moodle LMS and determine the number of HOTS after using the instrument. The findings were valid and useable. The HOTS instrument validity was seen from the construct validity and face validity. Construct validity intends to investigate the HOTS instrument's accuracy and collect responses from experts and practitioners. Based on validator evaluation, the Aiken's V value was obtained from 0.76 to 1.00, suggesting a valid result. This result indicated that the HOTS instrument featured good material, design, and language aspects. The material aspect relates to the question items according to the indicators; has only one correct answer key; contents follow the calculation goal and the education level; the item distractors work properly. The construction feature of the HOTS instrument associates with the subject matter; has clearly-formulated answer choices; the subject matter does not lead to a correct answer; no multiple negative shapes; has homogeneous answer choices; has a similar length of answer choices; the items do not depend on each other; and the options are type. Next, it relates to the formulation of communicative language, grammatical sentences, non-multi-significant sentences, and standard/general/neutral vocabulary in the language aspect. Using Moodle LMS as a medium to serve HOTS instruments will promote the access of the

students to online questions. E-learning using LMS Moodle is equipped with various facilities supporting online learning implementation that allows students to learn independently (Martín-Blas & Serrano-Fernández, 2009; Yildiz, Tezer, & Uzunboylu, 2018). Moodle LMS program presents an interesting display and is user-friendly (Martín-Blas & Serrano-Fernández, 2009). Students can work on the questions interactively and see the results directly.

Face validity in this analysis was obtained and evaluated based on students' HOTS instrument tests. Analyzing the HOTS instrument used IRT analysis methodology. It was suggested that all 24 items were fit as they reached the range of 0.77 to 1.30 in the MNSQ INFIT value, and -2.0 to 2.0 in the INFIT t. The item reliability value following the item estimate value summary measured at 0.66; meanwhile, the person's reliability based on the case estimate summary was 0.85 or very accurate (0.67 to 0.80). Thus, the instrument produced is appropriate for measuring students' HOTS as it has met the requirements according to the IRT analysis result.

The analysis result of students' HOTS obtained the average approximate value or skill level of each student, along with the standard deviation of  $0.01 \pm 1.35$  (moderate category). The case estimate result indicated that the HOTS skills of the students were in the moderate category. The low category of students' HOTS was influenced by several factors, one of which was that the students were not used to working on HOTS questions (Tanujaya, Mumu, & Margono, 2017; Yusuf & Widyaningsih, 2019). They needed to practice developing their HOTS by being exposed to HOTS-based learning sources. To realize HOTS, students are required to be more active in learning (Winarti, Cari, Widha, & Istiyono, 2015; Yusuf & Widyaningsih, 2019). Lecturers are also expected to act as facilitators who provide various learning resources and provide feedback on the students' tasks (Masruroh & Prasetyo, 2018). The use of e-learning allows students to access different learning resources in the form of texts, animations, simulations, multimedia, or virtual laboratories that can be accessed directly (Skultety, Gonzalez, & Vargas, 2017; Tee, Siti, Tengku, & Zainudin, 2013). It is expected that these e-learning facilities can facilitate students in learning so that their HOTS can be developed. Students' HOTS can also be improved through assignments and exercises in the learning process (Istiyono, Dwandaru, Megawati, & Ermansah, 2018; Yusuf & Widyaningsih, 2018). On this ground, it is of major importance to train the students' HOTS by applying learning technologies and quality instrument presentations through the IRT analysis.

## CONCLUSION

The HOTS instrument presented by Moodle LMS in e-learning obtains a good performance. The IRT analysis, including item fit, reliability, and difficulty, acquires the mean and standard deviation parameters for INFIT MNSQ of 1.0 and 0.0; the items have proven to fit RM 1-PL. Additionally, test characteristics comprised item fitness, reliability, and difficulty. The trial result obtains the criteria of INFIT MNSQ mean and standard deviation of 1.0 and 0.0, respectively, implying that the items fit the RM1-PL. In addition, the value of item reliability based on the value of item estimate summary arrives at 0.66; meanwhile, the person reliability under the case estimate summary 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 Quest output result also reveals that the average values of THRSHL and its standard deviation are  $0.00 \pm 0.71$ , 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 used in the HOTS measurement. Every student's average estimate or skill level along with the standard deviation is  $0.01 \pm 1.35$  or in the moderate category. Students' HOTS must be practiced by providing HOTS-based learning resources.

#### ACKNOWLEDGMENT

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

#### REFERENCES

- Adams, R. J., & Khoo, S.-T. (1996). *Quest: the interactive test analysis system*. Camberwell, Vic.: Australian Council for Educational Research.
- Aiken, L. R. (1980). Content Validity and Reliability of Single Items or Questionnaires. *Educational and Psychological Measurement*, 40(4), 955–959.
- Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45(1), 131–142.
- Aldoobie, N. (2015). ADDIE Model. *American International Journal of Contemporary Research*, 5(6), 72.
- Azevedo, J. M. (2015). e-Assessment in mathematics courses with multiple-choice questions tests. *CSEDU 2015 - 7th International Conference on Computer Supported Education, Proceedings*, 2, 260–266. <https://doi.org/10.5220/0005452702600266>
- Bogdanović, Z., Barać, D., Jovanić, B., Popović, S., & Radenković, B. (2014). Evaluation of Mobile Assessment in A Learning Management System. *British Journal of Educational Technology*, 45(2), 231–244. <https://doi.org/10.1111/bjet.12015>
- Bond, T., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences*. New York: Routledge.
- Brun, M., & Hinostroza, J. E. (2014). Learning to become a teacher in the 21st century: ICT integration in Initial Teacher Education in Chile. *Journal of Educational Technology & Society*, 17(3), 222–238.
- Hambleton, R. K., & Rogers, H. J. (1989). Detecting Potentially Biased Test Items: Comparison of IRT Area and Mantel-Haenszel Methods. *Applied Measurement in Education*, 2(4), 313–334. [https://doi.org/10.1207/s15324818ame0204\\_4](https://doi.org/10.1207/s15324818ame0204_4)
- Istiyono, E. (2017). The Analysis of Senior High School Students' Physics HOTS in Bantul District Measured using PhysRemChoTHOTS. *AIP Conference Proceedings*, 1868(August), 1–7. <https://doi.org/10.1063/1.4995184>



- Istiyono, E. (2018). IT-based HOTS assessment on physics st learning as the 21 century demand at senior high schools: Expectation and reality IT-Based HOTS Assessment on Physics Learning as the 21 st Century Demand at Senior High Schools : Expectation and Reality. *AIP Conference Proceedings*, 2014(020014), 1–6.
- Istiyono, E., Dwandaru, W. S. B., Megawati, I., & Ermansah. (2018). Application of Bloomian and Marzanoian Higher Order Thinking Skills in the Physics Learning Assessment: an Inevitability. *Advances in Social Science, Education and Humanities Research*, 164(ICLI 2017), 136–142. <https://doi.org/10.2991/icli-17.2018.26>
- Istiyono, E., Dwandaru, W. S. B., & Muthmainah. (2019). Developing of Bloomian HOTS Physics Test: Content and Construct Validation of The PhysTeBloHOTS. *Journal of Physics: Conference Series*, 1397(012017), 1–9.
- Kowsalya, D. N., Venkat Lakshmi, H., & Suresh, K. P. (2012). Development and Validation of a Scale to assess Self-Concept in Mild Intellectually Disabled Children. *International Journal of Social Sciences & Education*, 2(4).
- Krathwohl, D. R., & Anderson, L. W. (2010). Merlin C. Wittrock and the Revision of Bloom's Taxonomy. *Educational Psychologist*, 45(1), 64–65. <https://doi.org/10.1080/00461520903433562>
- Lee, M. F., & Zainal, N. A. (2017). Development of needham model based E-module for electromagnetic field & wave. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 120–124). <https://doi.org/10.1109/IEEM.2017.8289863>
- Limongelli, C., Sciarrone, F., & Vaste, G. (2011). Personalized e-learning in Moodle: the Moodle\_LS System. *Journal of E-Learning and Knowledge Society*, 7(1), 49–58. Retrieved from <https://www.learntechlib.org/p/43340>
- Martin-Blas, T., & Serrano-Fernández, A. (2009). The role of new technologies in the learning process: Moodle as a teaching tool in Physics. *Computers & Education*, 52(1), 35–44. <https://doi.org/10.1016/J.COMPEDU.2008.06.005>
- Masruroh, A. N., & Prasetyo, Z. K. (2018). Effect of E-Module with Guided Inquiry Approach Containing Nature of Science to Student's Science Literacy. *E-Journal Pend. IPA*, 7(3), 165–171.
- Pandey, S. R., & Pandey, S. (2009). Developing a More Effective and Flexible Learning Management System (LMS) for the Academic Institutions using Moodle. *ICAL 2009 - Technology, Policy and Innovation*, 249–254.
- Raykov, T., & Marcoulides, G. A. (2015). On the Relationship Between Classical Test Theory and Item Response Theory: From One to the Other and Back. *Educational and Psychological Measurement*, 76(2), 325–338. <https://doi.org/10.1177/0013164415576958>

- Skultety, L., Gonzalez, G., & Vargas, G. (2017). Using Technology to Support Teachers' Lesson Adaptations during Lesson Study. *Journal of Technology and Teacher Education*, 25(2), 185–213. Retrieved from <https://www.learntechlib.org/p/172139>
- Tanujaya, B., Mumu, J., & Margono, G. (2017). The Relationship between Higher Order Thinking Skills and Academic Performance of Student in Mathematics Instruction. *International Education Studies*, 10(11), 78–85.
- Tee, S. S., Siti, T., Tengku, M., & Zainudin, S. (2013). User Testing for Moodle Application. *International Journal of Software Engineering and Its Applications*, 7(5), 243–252.
- Winarti, Cari, Widha, S., & Istiyono, E. (2015). Analysis of Higher Order Thinking Skills Content of Physics Examinations In Madrasah Aliyah. In *International Conference on Mathematics, Science, and Education 2015 (ICMSE 2015)* (Vol. 2015, pp. 32–38).
- Yildiz, E. P., Tezer, M., & Uzunboylu, H. (2018). Student Opinion Scale Related to Moodle LMS in an Online Learning Environment: Validity and Reliability Study. *International Journal of Interactive Mobile Technologies (IJIM)*, 12(4), 97–108.
- Yusuf, I., & Widyaningsih, S. W. (2018). Profil Kemampuan Mahasiswa dalam Menyelesaikan Soal HOTS di Jurusan Pendidikan Fisika Universitas Papua. *Jurnal Komunikasi Pendidikan*, 2(14), 42–49.
- Yusuf, I., & Widyaningsih, S. W. (2019). HOTS profile of physics education students in STEM-based classes using PhET media. *Journal of Physics: Conference Series*, 1157(032021), 1–5.
- Yusuf, I., Widyaningsih, S. W., & Sebayang, S. R. B. (2018). Implementation of E-learning based-STEM on Quantum Physics Subject to Student HOTS Ability. *Turkish Science Education*, 15(December), 67–75.