Wahana Pendidikan Fisika (2023) 8 (1): 60-70
https://doi.org/10.17509/wapfi.v8i1.48272
2406-9205 (Print) | 2477-4820 (Online)
ORIGINAL RESEARCH

# Instrumental analysis of higher order thinking skills in linear motion topic using item response theory

**Pungkas Wijiasmoro, Taufik Ramlan Ramalis, Purwanto**

Received: 5 July 2022 · Accepted: 7 January 2023 · Published Online: 28 February 2023


**Abstract**
This study aims to analyze the characteristics of a higher-order thinking skills (HOTS) instrument on linear motion of physics subjects in high school using item response theory (IRT). The characteristics of the instrument analyzed were the suitable logistic model, validity, reliability, discrimination index, and difficulty. The method of this study was a descriptive study with quantitative data and the design used was a one-shot design. The participants of this study were 101 high school students who have already studied linear motion. The data obtained were analyzed using the item response theory 2-PL model. The instrument was 22 multiple-choice questions. The results of the study show that the instrument is valid. The reliability of the instrument shows that it is reliable to be given to students with low to medium ability with ability scores of -1.1 to 0.9. The overall discrimination index is considered to be good with a = 1.98. As for the discrimination index for each item, 11 items are considered to be having good discrimination indexes and the other 11 have bad discrimination indexes. The overall difficulty level belongs to the medium category with b = -0.1. As for the difficulty level for each item, one item belongs to the easy category, one item belongs to the hard difficulty, and the other 20 belong to the medium category

**Keywords**: *Higher Order Thinking Skills (HOTS) · Item Response Theory · Linear motion*

## INTRODUCTION

Helping students develop lifelong learning skills is an important educational outcome for 21st-century education (Ong et.al., 2016). These 21st-century skills include critical thinking and problem-solving, creativity and innovation, communication, and collaboration (Redhana & Wayan, 2019). To support this, the government through the implemented curriculum promotes skills that can encourage students' thinking abilities. Higher-order thinking ability or HOTS is a tool to facilitate the process of thinking with many variables under certain conditions (Suprapto et.al., 2020). Thinking skills can be divided into higher-order thinking skills (HOTS) and lower-order thinking skills (LOTS) (Iskandar et.al., 2015). Higher-order thinking skills basically mean thinking that is located at a higher level in the cognitive process hierarchy (Ramos et.al., 2013). In Indonesia, higher-order thinking skills are always associated with Bloom's revised taxonomy, especially at C4 to C6 (analyzing, evaluating, and creating), even

✉ Pungkas Wijiasmoro
pungkasw@student.upi.edu

Purwanto
purwanto@upi.edu

Taufik Ramlan Ramalis
taufik_lab.ipba@upi.edu

Departemen Pendidikan Fisika Universitas Pendidikan Indonesia. Bandung, Indonesia.

**How to Cite**: Wijiasmoro, P., Ramalis, T.R., & Purwanto. (2023). Instrumental analysis of higher order thinking skills in linear motion topic using item response theory. *Wahana Pendidikan Fisika*, *8*(1), 60-70. https://doi.org/10.17509/wapfi.v8i1.48272

the use of this taxonomy is also used in the curriculum used in Indonesia (Ramadhan et.al, 2019). Through HOTS students will be able to distinguish ideas or ideas clearly, argue well, be able to solve problems, be able to construct explanations, be able to hypothesize and understand complex things more clearly (Desilva et.al, 2020). Higher-order thinking skills (HOTS) are an important aspect of a learning activity (Heong et.al., 2011). The advantage of incorporating HOTS into learning in schools can increase student academic achievement as well as lifelong benefits for students (Conklin, 2011). One of the importance of higher-order thinking skills is to equip students to face the demands of modern times in the future (Mbayowo & Pasaribu, 2021).

Despite being a very important skill, several recent studies have shown that students' high-order thinking skills in Indonesia are still relatively low (Muhammad & Ratu, 2018)(Hidayati & Sinaga, 2019)(Pratama et.al., 2015). In a release from *the Program for International Student Assessment* or PISA which often conducts research in HOTS, Indonesia ranks 70th out of 78 countries in the field of science. This is clearly very unfortunate considering how important HOTS is for students. So that HOTS needs to be further promoted in learning at school. Apart from the daily learning activities, another thing that is needed to support this is the procurement of instruments to measure students' higher-order thinking skills. Besides being useful for teachers to know the characteristics of their student's abilities, this instrument also has benefits for students. Conducting an assessment of students' higher-order thinking skills can encourage students to learn to use higher-order thinking (Ramadhan et.al, 2019).

In a preliminary study conducted by the author of high school teacher educators teaching physics in Bandung City, 80% of respondents said that only 25% of the HOTS questions were used in the test questions and the remaining 20% said they did not use HOTS questions at all. This was because according to the teacher, the HOTS questions were considered to take quite a long time for students to work on and were difficult for most students to understand. However, this proportion still adjusts to the conditions of the students who will be tested. If students are considered by the teacher to really understand the material being tested, then maybe more HOTS questions can be given. Regarding the obstacles in making HOTS questions, 60% of respondents felt that there were still not enough references to HOTS questions, 40% of respondents said that there was no guide in making HOTS questions, 20% of respondents said that students' ability to read and understand questions was an obstacle in making HOTS questions. , and the remaining 20% said that HOTS questions were almost similar to ordinary physics analysis questions. However, the teacher's ability to make HOTS questions and use HOTS in schools still needs to be studied further.

In an instrument preparation process, an analysis process is certainly carried out. Instrument analysis is carried out to investigate the characteristics of the instrument so that it can be seen whether the instrument is suitable for use as intended or not. Analysis of the test instrument performed can use classical test theory or CTT or other alternatives, namely item response theory or IRT. Item response theory is widely used in education to calibrate and evaluate items in tests, questionnaires, and other instruments and to score abilities, attitudes, or other hidden traits (Xinming & Yung, 2014). Test developers can analyze the parameters of the items and the abilities of students using IRT (Rakkapao et.al., 2016). The main advantage of IRT compared to classical theory is that in IRT analysis the item parameters do not depend on the latent nature (ability) of the respondent and the respondent's parameters do not depend on

https://doi.org/10.17509/wapfi.v8i1.48272 

the item being tested (Bortolotti et.al., 2013). This means that IRT analysis can be carried out on students with low to high abilities, regardless of whether they can do the test or not. Test takers with high abilities will have a greater probability of answering correctly compared to test takers who have low abilities (Novia et.al., 2018).

Popular IRT models that can be used are one-parameter, two-parameter, and three-parameter logistic models (Hambleton et.al., 1991). The one-parameter logistic model (1-PL), which is often referred to as the *Rasch model*, takes into account the item difficulty parameter (difficulty level). The two-parameter logistic model (2-PL) takes into account two item parameters, namely the level of difficulty and discriminating power. The three-parameter logistic model (3-PL) takes into account three item parameters, namely difficulty level, discriminating power, and pseudo-guessing level. The method of analysis using item response theory includes the use of the information function. The use of the information function is useful in describing the items and tests, selecting items, and comparing tests (Hambleton et.al., 1991). There are two types of information functions, namely the item information function and the test information function.

So based on the explanation above, the authors are interested in conducting research on students' higher-order thinking skills test instruments. Researchers are interested in seeing how the characteristics of the HOTS instrument are made, in terms of the logistic model, validity, reliability, level of difficulty, and the differentiability of the instrument. This instrument is only limited to one physics subject at school, namely rectilinear motion. The reference used by researchers in making the HOTS instrument is the revised Bloom's taxonomy.

Based on this, the formulation of the problem in this study is "What are the characteristics of the instrument for higher order thinking skills (HOTS) in straight motion material using Item Response Theory?". The purpose of this study is to analyze the characteristics of the instrument for higher-order thinking skills in straight motion topic using grain response theory.

## METHODS

This study uses descriptive research methods using quantitative data. The research design used is *a one-shot design* . *One-shot design* is a research design that uses one time data collection (Arikunto, 2019). The population of this study were all students of class X at SMA Bandung Barat Regency, and the sample was 101 students from class X majoring in Natural Sciences at SMA Kabupaten Bandung Barat. Sampling in this study using *purposive sampling technique*. *Purposive sampling* is a sample determination that is carried out with certain considerations (Sugiyono, 2013). The considerations for taking the sample in this study were high school students in class X who had studied straight motion material in physics lessons at their school and the number of classes that could be provided by the school for research.

To obtain the necessary data, the *Higher Order Thinking Skills* (HOTS) instrument is used to measure students' higher order thinking skills. This instrument is in the form of multiple choice questions with five answer choices (A, B, C, D, and E) totaling 22 questions given to students directly in class. This instrument is structured based on the cognitive dimensions of Bloom's revised taxonomy that denote HOTS, namely C4, C5, and C6. The analysis technique used is item response theory or IRT. The IRT model used is a two-parameter logistic model (2-PL) which takes into account two item parameters, namely difficulty level and discriminating

https://doi.org/10.17509/wapfi.v8i1.48272 

power. This model was chosen because in the analysis of research data, the 2-PL model showed the highest test information function value compared to the 1-PL and 2-PL models.

## RESULT AND DISCUSSION

The characteristics of the HOTS straight motion instrument tested can be obtained from the IRT analysis which was carried out using IRTPro for student and ministep software.

### Appropriate logistics model

The analytical model that can be used in this case is a one-parameter logistic model (1-PL), a two-parameter logistic model (2-PL), or a three-parameter logistic model (3-PL) as described in chapter 2. The most suitable model suitable for use in research data can be determined by comparing the value of the information function test of the three models. The value of the test information function (*I*) can be seen from the peak of the test information function curve based on the research data that has been obtained. The model that gives the highest test information function value is the most suitable model to be used in the analysis of the characteristics of this instrument. This analysis was carried out using the help of two *software*, namely *IRTPro for Student* and *ministep*.

Comparison of the three information function curves of the 1-PL, 2-PL, and 3-PL models is shown in one graph in the following figure with the red curve showing the information function curve of the 1-PL model test, the blue curve showing the information function curve of the 2-PL model test, and the green curve shows the information function curve of the 3-PL model test:

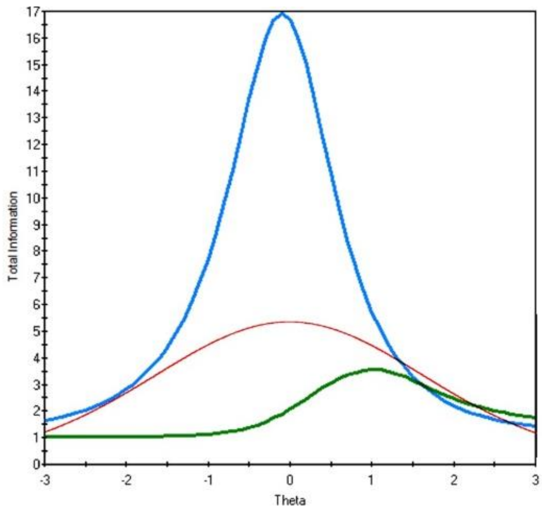


**Figure 1.** Comparison of the information function curves of the three logistics models

Based on the comparison of the three information function test curves above, it was found that the IRT 2-PL model had the highest test information function value, which was equal to *I* = 16.93. So the analytical model used to analyze research data is the IRT 2-PL model.

### Validity

The results of the validity analysis of the 22 items are shown in table 1. Items are said to be valid if they meet at least the following two criteria (Sumintono & Widhiarso, 2014):

https://doi.org/10.17509/wapfi.v8i1.48272 

- The value of the *fit fit mean square* (MnSq) is in the range 0.5 < MnSq < 1.5
- *Z-Standard outfit* value (ZStd) is in the range -2.0 < ZStd < 2.0
- The value of *point measure correlation* (PtMeaCorr) is in the range 0.4 < PtMeaCorr < 0.85

**Table 1.** The results of the validity test of the instrument questions

| Question Number | *Outfit MnSq* | *Outfit ZStd* | *PtMeaCorr* | Category | Interpretation |
|---|---|---|---|---|---|
| 1 | 1.06 | 0.99 | 0.09 | fit 2 | valid |
| 2 | 0.9 | -1.96 | 0.41 | Outfits | valid |
| 3 | 1.03 | 0.28 | 0.13 | fit 2 | valid |
| 4 | 0.99 | 0.05 | 0.17 | fit 2 | valid |
| 5 | 0.99 | -0.24 | 0.22 | fit 2 | valid |
| 6 | 1.09 | 0.52 | 0.07 | fit 2 | valid |
| 7 | 1,1 | 0.83 | 0.06 | fit 2 | valid |
| 8 | 0.86 | -2.32 | 0.51 | fit 2 | valid |
| 9 | 1.03 | 0.45 | 0.12 | fit 2 | valid |
| 10 | 1.09 | 1.46 | 0 | fit 2 | valid |
| 11 | 0.95 | -0.6 | 0.3 | fit 2 | valid |
| 12 | 0.96 | -0.47 | 0.28 | fit 2 | valid |
| 13 | 0.95 | -0.06 | 0.2 | fit 2 | valid |
| 14 | 0.95 | -0.95 | 0.31 | fit 2 | valid |
| 15 | 0.87 | -1.26 | 0.44 | Outfits | valid |
| 16 | 1.29 | 0.99 | -0.09 | fit 2 | valid |
| 17 | 1,12 | 0.51 | 0.03 | fit 2 | valid |
| 18 | 1.07 | 0.38 | 0.04 | fit 2 | valid |
| 19 | 1.05 | 0.38 | 0.14 | fit 2 | valid |
| 20 | 0.91 | -1.59 | 0.4 | fit 2 | valid |
| 21 | 1,1 | 0.74 | 0.01 | fit 2 | valid |
| 22 | 1.04 | 0.44 | 0.12 | fit 2 | valid |

From the table, it can be seen that of the 22 items in the instrument tested, 20 questions fall into the *fit* 2 category and the remaining two questions, namely questions 2 and 15, fall into the *outfit category* . So that each item in the HOTS straight motion instrument tested in this study can be said to meet the requirements to be said to be valid. Validity is a measure that shows the levels of validity or validity of an instrument (Arikunto, 2019). A valid instrument means that the instrument can be used to make measurements according to its purpose, in this case students' higher-order thinking skills in straight motion material.

## Reliability

Instrument reliability can be determined by looking at the intersection of the total information function curve and the standard error measurement curve (SEM). The curve is shown in the figure as follows:

https://doi.org/10.17509/wapfi.v8i1.48272

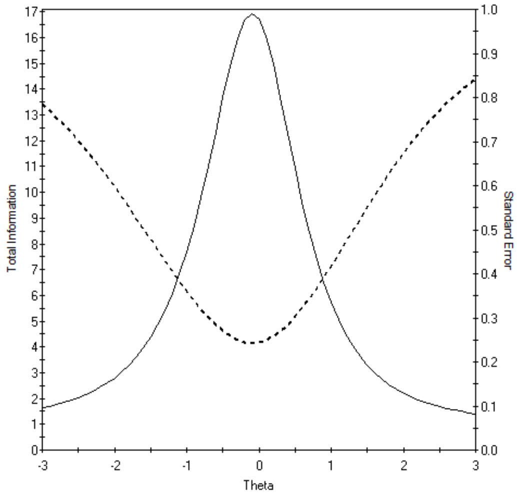


**Figure 2.** Information function curves test the 2-PL model of research data

In the figure above, it is known that the intersection of the information curve with the standard error is on a capability scale (θ) -1.1 with a total information value ( *I* ) of 6.82 and an SEM value of 0.38 and on a capability scale of 0.9 with information value of 6.50 and SEM value of 0.39. These data indicate that the HOTS rectilinear instrument in this study will be reliable or more effective to be administered to students with a range of ability scales -1.1 to 0.9, which means that this instrument will be effectively administered to students with low to moderate abilities. Reliability indicates an understanding that an instrument can be trusted to be used as a data collection tool because the instrument is good (Arikunto, 2019). Reliability can be said as the level of reliability of an instrument. So that it can be said that the straight motion HOTS instrument in this study is quite reliable or effective if it is given to students with low to moderate abilities. On the other hand, this instrument may not be sufficiently reliable or effective if administered to high ability students.

## Discriminating Power

The overall differentiating power of the instrument in the analysis using the grain response theory can be seen from the test characteristic curve *or* TCC. The TCC curve obtained from this research data can be seen in the following figure:

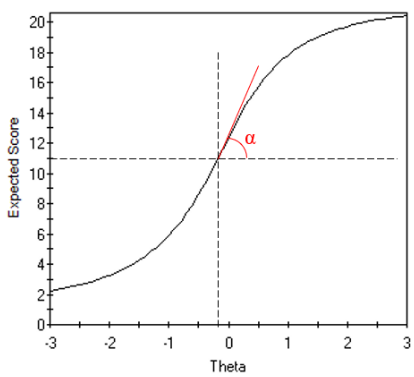


**Figure 3.** The TCC curve of the HOTS instrument research data is straight motion

https://doi.org/10.17509/wapfi.v8i1.48272

From the curve above, *the theta* shows the ability of students and *the expected score* shows the expected score when a student takes a test with this instrument with one question worth 1 so that the range of values that can be obtained from 22 questions is from 0 to 22. The curve shows that when a participant with ability -3 has an expected score of 2 and when a participant with ability 3 has an expected score of 20. The overall differentiability of the instrument is indicated by the magnitude of the slope of the curve in the middle of the expected score *range*.

*The expected score* range is from 2 to 20 so that the mean is at the *expected score* is 11. From the figure above, the overall discriminating power of the instrument is indicated by the magnitude of the slope of the curve or tan α. In the case above, the value of tan α is 1.98. So that the overall discriminating power of the straight motion HOTS instruments tested in this study was 1.98. This value is still in the range of criteria for good discriminating power, that is, if the value is 0 to 2. As for the discriminating power of individual items in the analysis of item response theory using the 2-PL model in IRTPro for Student software, it can be seen from the parameter values indicating discriminating *power*. ( *a* ). This *a* value can then be interpreted to be a good discriminatory power if the value is in the range of 0 to 2. From the analysis carried out, the obtained a values are *displayed* along with their interpretations in the following table:

**Table 2.** Value of discriminating power (a) items

| Question Items | Distinguishing power value (a) | Interpretation |
|---|---|---|
| 1 | 0.2 | Good |
| 2 | 0.47 | Good |
| 3 | 1.34 | Good |
| 4 | 3,4 | Not good |
| 5 | 2,31 | Not good |
| 6 | 2.61 | Not good |
| 7 | 1.01 | Good |
| 8 | 0.91 | Good |
| 9 | 1.14 | Good |
| 10 | 0.52 | Good |
| 11 | 1.34 | Good |
| 12 | -0.14 | Not good |
| 13 | 2.02 | Not good |
| 14 | 0.65 | Good |
| 15 | 1.67 | Good |
| 16 | 3.07 | Not good |
| 17 | 2.57 | Not good |
| 18 | 1.61 | Good |
| 19 | 1.86 | Good |
| 20 | 0.53 | Good |
| 21 | 1.34 | Good |
| 22 | 1.18 | Good |

*a* values obtained, of the 22 questions in the HOTS straight motion instrument tested, 15 items were included in the good discriminating power category and the remaining 7 questions were in the bad discriminating power category. Items that can be answered correctly by students in the upper and lower groups have poor discriminating power, similarly if students in the upper and lower groups cannot answer the items correctly then the item has poor discriminating power, good items. is a question that can be answered correctly by students in the upper group

https://doi.org/10.17509/wapfi.v8i1.48272

only (Arikunto, 2019). So that the discriminating power in this instrument can be said to be good in distinguishing students from the upper and lower groups. The discriminating power value of $a$ = 1.98 is still in the range of 0 to 2 which can be considered good but is very close to the upper limit (2). If we look at each question individually, half the questions cannot be said to have good discriminating power so that even though they are considered good, the discriminating power cannot be said to be very good in distinguishing students from the upper and lower groups.

### Difficulty Level

The overall difficulty level of the instrument in the analysis using item response theory can be seen from the test characteristic curve or TCC as shown in Figure 2 above. The overall difficulty level of the instrument is indicated by the theta value at the median expected score. It is known that the mean value is at the expected score of 11. At that point, theta is -0.1 and this value is included in the category of moderate difficulty level. So that the difficulty level of the HOTS straight motion instrument in this study can be said to be moderate.

As for the difficulty level of individual items in the analysis of item response theory using the 2-PL model in the IRTPro for Student software, it can be seen from the parameter values that indicate the level of difficulty, namely the threshold ( b ). From the analysis performed, the b values are displayed along with their interpretation in the following table:

**Table 3.** Difficulty level value (b) item questions

| Question Items | Difficulty level value (b) | Interpretation |
|---|---|---|
| 1 | -1.25 | easy |
| 2 | -0.53 | moderate |
| 3 | -0.19 | moderate |
| 4 | -0.07 | moderate |
| 5 | -0.11 | moderate |
| 6 | -0.1 | moderate |
| 7 | -0.25 | moderate |
| 8 | -0.27 | moderate |
| 9 | -0.22 | moderate |
| 10 | -0.48 | moderate |
| 11 | -0.19 | moderate |
| 12 | 1.79 | hard |
| 13 | -0.12 | moderate |
| 14 | -0.39 | moderate |
| 15 | -0.15 | moderate |
| 16 | -0.08 | moderate |
| 17 | -0.1 | moderate |
| 18 | -0.16 | moderate |
| 19 | -0.13 | moderate |
| 20 | -0.47 | moderate |
| 21 | -0.19 | moderate |
| 22 | -0.21 | moderate |

*b* values obtained, of the 22 questions in the straight-line HOTS instrument tested, there were 20 items included in the moderate category, one item included in the easy category, and one item included in the difficult category. Questions that fall into the easy category are

https://doi.org/10.17509/wapfi.v8i1.48272

question number 1 with a *b value* of -1.25. Questions included in the difficult category are question number 12 with a *b value* of 1.79. While the remaining 20 items are included in the moderate category with a value of *b* from -0.53 to -0.07. From the value of *b* above, it was found that most of the questions were in the medium category. This finding is in line with the results of the TCC curve analysis which states that the difficulty level of the HOTS straight motion instrument in this study is moderate. Items belonging to the moderate category are items that are neither too difficult nor too easy (Arikunto, 2019). A good question is one that is neither too easy nor too difficult. So based on this statement, most of the items in this instrument can be said to be good in terms of difficulty level because they are neither too easy nor too difficult.

## CONCLUSION

From the research and analysis using item response theory (IRT) that has been carried out, it can be concluded that the characteristics of the higher order thinking skill (HOTS) instrument for high school physics material for straight motion are made as follows: 1) The best logistic parameter model to be used in analyzing the Higher-Order Thinking Skills instrument for straight motion material in this study is the 2-PL model with an information function test value of I = 16.93.; 2) This instrument passed the validity test so that it can be said that this instrument is valid with 2 items included in the *outfit category* and 20 questions in the *fit* 2 category; 3) Regarding reliability, this instrument can be relied upon when tested on students with low to moderate abilities on a test taker's ability scale of -1.1 to 0.9; 4) The discriminating power of this instrument as a whole is good so that it can distinguish students from the upper group and the lower group with the overall discriminating power having a value of a = 1.98; and 5) The difficulty level of this instrument as a whole is in the medium category with the overall instrument difficulty level value of b = -0.1.

Based on the research results that have been obtained and described, the HOTS physics instrument for straight motion material that has been made can be used as a reference by educators and other parties in evaluating or training students' higher order thinking skills, especially in straight motion material. In addition, the results of this study can also add insight into the application of item response theory (IRT) in analyzing the characteristics of an instrument, especially similar instruments, as well as add insight into instruments that can be used in measuring higher order thinking skills in physics subjects. While the recommendations that the author can give after conducting this research are as follows: 1) For further research, the research was carried out in the same semester as the semester in which the students studied the rectilinear motion material so that the rectilinear motion material tested with the HOTS instrument under study was still fresh in the students' minds; 2) For further research, it is also advisable to work closely with the school physics teacher who is conducting the research so that the results of the students' work in the research are taken as additional value for students so that students want to take the test seriously; 3) For further research, in preparing the test it is better to use questions with physical phenomena that are closer to everyday life in accordance with the material being tested; and 4) For further research, the HOTS instrument that has been made can be disseminated more widely to students who have studied rectilinear motion material so that more data can be obtained. For teachers, it is necessary to train students with higher-order thinking questions so that students are used to it and can improve students' high-order thinking skills.

https://doi.org/10.17509/wapfi.v8i1.48272

## REFERENCES

An, X., & Yung, Y. F. (2014). Item response theory: What it is and how you can use the IRT procedure to apply it. *SAS Institute Inc. SAS364-2014*, *10*(4), 1-14.

Bortolotti, S. L. V., Tezza, R., de Andrade, D. F., Bornia, A. C., & de Sousa Júnior, A. F. (2013). Relevance and advantages of using the item response theory. Quality & Quantity, 47, 2341-2360. https://doi.org/10.1007/s11135-012-9684-5

Conklin, W. (2011). *Higher-order thinking skills to develop 21st century learners.* Teacher Created Materials.

Desilva, D., Sakti, I., & Medriati, R. (2020). Pengembangan Instrumen Penilaian Hasil Belajar Fisika Berorientasi Hots (Higher Order Thinking Skills) Pada Materi Elastisitas Dan Hukum Hooke. *Jurnal Kumparan Fisika*, *3*(1), 41-50. https://doi.org/10.33369/jkf.3.1.41-50

Hambleton, Ronald, K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Publications

Heong, Y. M., Othman, W. B., Yunos, J. B. M., Kiong, T. T., Hassan, R. B., & Mohamad, M. M. B. (2011). The level of marzano higher order thinking skills among technical education students. *International Journal of Social Science and Humanity*, *1*(2), 121-125.

Hidayati Y. & Sinaga, P. (2019). The profile of critical thinking skills students on science learning. *Journal of Physics: Conference Series*, *1402*(4), https://doi.org/10.1088/1742-6596/1402/4/044075

Iskandar, D., & Senam, S. (2015). Studi kemampuan guru kimia sma lulusan UNY dalam mengembangkan soal UAS berbasis HOTS. *Jurnal Inovasi Pendidikan IPA*, *1*(1), 65-72. https://doi.10.21831/jipi.v1i1.4533

Mbayowo, R. & Pasaribu, M. (2021). Analisis kemampuan kemampuan berpikir tingkat tinggi siswa dalam menyelesaikan soal fisika bentuk representasi gambar di SMA negeri se-kabupaten morowali utara. *Wahana Pendidikan Fisika*, *6*(1), 96-103. https://doi.org/10.17509/wapfi.v6i1.32458

Muhammad, E. & Ratu, T. (2018). Pencapaian HOTS (higher order thinking skills) mahasiswa program studi pendidikan fisika FKIP Universitas Samawa. *Jurnal Pendidikan Fisika dan Teknologi*, *4*(2), 208-212. https://doi.org/10.29303/jpft.v4i2.831

Novia, R., Ramalis, T. R., & Efendi, R. (2018). Pengembangan dan Karakterisasi Tes Keterampilan Berpikir Kritis Materi Tekanan berdasarkan Teori Respon Butir. *Wahana Pendidikan Fisika*, *4*(2), 155-162. https://doi.org/10.17509/wapfi.v4i2.20181

Ong, K. K. A., Hart, C. E., & Chen, P. K. (2016). Promoting higher-order thinking through teacher questioning: a case study of a Singapore science classroom. *New Waves-Educational Research and Development Journal*, *19*(1), 1-19.

Pratama, N. S. & E. Istiyono. (2015). Studi pelaksanaan pembelajaran fisika berbasis higher order thinking (HOTS) pada kelas X di SMA Negeri Kota Yogyakarta. *Prosiding: Seminar Nasional Fisika Dan Pendidikan Fisika*, *6*(2). 104-112.

Rakkapao, S., Prasitpong, S., & Arayathanitkul, K. (2016). Analysis test of understanding of vectors with the three-parameter logistic model of item response theory and item response curves technique. *Physical review physics education research*, *12*(2), 020135.. https://doi.org/10.1103/PhysRevPhysEducRes.12.020135

Ramos, J. L. S., Dolipas, B. B., & Villamor, B. B. (2013). Higher order thinking skills and academic performance in physics of college students: A regression analysis. *International Journal of Innovative Interdisciplinary Research*, *4*(1), 48-60. http://dx.doi.org/10.18415/ijmmu.v7i9.1884

Ramadhan, S., Mardapi, D., Prasetyo, Z. K., & Utomo, H. B. (2019). The development of an instrument to measure the higher order thinking skill in physics. *European Journal of Educational Research*, *8*(3), 743-751. https://doi.org/10.12973/eu-jer.8.3.743

Redhana, I. W. (2019). Mengembangkan keterampilan abad ke-21 dalam pembelajaran kimia. Jurnal *Inovasi Pendidikan Kimia*, *13*(1). 2239-2253. https://doi.org/10.15294/jipk.v13i1.17824

 

Sugiyono. (2013). *Metode penelitian pendidikan pendekatan kuantitatif, kualitatif dan R&D.* Bandung, Alfabeta

Suharsimi, A. (2006). *Prosedur penelitian suatu pendekatan praktik.* Jakarta: Rineka Cipta, 134.

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi).* Trim Komunikata Publishing House.

Suprapto, E., Sumiharsono, R., & Ramadhan, S. (2020). The Analysis of Instrument Quality to Measure the Students' Higher Order Thinking Skill in Physics Learning. *Journal of Turkish Science Education*, *17*(4), 520-527.