

Developing and implementing an instrument for assessing critical thinking and visual representations in learning Physics materials of Optical Instruments

Dewi Nurulhasni*; Eka Ayu Nurbaiti; Rosa Safirotun Nabila; Jumadi

Universitas Negeri Yogyakarta, Indonesia

*Corresponding Author. E-mail: dewinurulhasni.2021@student.uny.ac.id

ARTICLE INFO

Article History

Submitted:

28 September 2022

Revised:

22 November 2022

Accepted:

30 May 2023

Keywords

assessment instrument;
critical thinking; visual
representation

Scan Me:



ABSTRACT

This study aims to develop an assessment instrument to measure students' critical thinking and visual representation abilities in optical instrument material. This research and development (R&D) employed the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) instructional model. This research developed a test instrument that has more specific stages so the ADDIE development stage is integrated with the development stage of the test instrument developed by Mardapi. The subjects in this study are grade XI students of State Madrasah Aliyah (MAN) 4 Bantul. This study produced a test instrument for measuring students' critical thinking and visual representation abilities, with high validity, and analyzed using Aiken's V. The items for assessing critical thinking and visual representation abilities were analyzed using the QUEST program with the following findings. (1) The analysis results of instrument validation using Aiken's V obtained an average Aiken's Value above 0.4 with high validity and moderate validity criteria so all items are valid. (2) The pre-test analysis on item estimates obtained MNSQ Infit in the range 0.77-1.30 and the value of outfit $t \leq 2$ and in the MNSQ Infit case estimates in the range 0.77-1.30, so overall, the items matched to Rasch models. (3) Post-test analysis on item estimates obtained MNSQ Infit in the range 0.77-1.30 and outfit t value ≤ 2 and in case estimates MNSQ Infit in the range 0.77-1.30 so, overall, items are in accordance with Rasch models. (4) The reliability of the pre-test items is 0.92 while the reliability of the post-test items is 0.76. (5) The difficulty level of the items in the pre-test results shows that Question 10 is the most difficult while number 3 is the easiest, and in the post-test results, Question 8 is the most difficult while Question 6 is the easiest.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



To cite this article (in APA style):

Nurulhasni, D., Nurbaiti, E., Nabila, R., & Jumadi, J. (2023). Developing and implementing an instrument for assessing critical thinking and visual representations in learning Physics materials of optical instruments. *Jurnal Penelitian dan Evaluasi Pendidikan*, 27(1), 52-62. doi:<https://doi.org/10.21831/jpep.v27i1.53550>

INTRODUCTION

Learning according to the [Regulation of the Minister of Education and Culture No. 23 of 2016](#), is a process of interaction between students, between students and educators, and learning resources within one scope of the learning environment. Current learning requires students to have knowledge, thinking abilities, and skills in the fields of media, technology, and information. The 21st-century skills focus on critical, analytical, and innovative thinking skills ([Ackerman, 2010](#)). This is in accordance with what is explained by the Ministry of Education and Culture that learning emphasizes the ability of students to formulate problems, think analytically, and cooperate in solving a problem. Therefore, learning physics requires students to have critical thinking skills to be applied in everyday life.

Critical thinking skills are one of the demands that must be met in current learning. Critical thinking skills have a big effect and role on people who follow the development of science and technology ([Luthvitasari et al., 2012](#)). The professional level of a person is heavily affected by

his critical thinking skills (Butler et al., 2008) and someone who has low critical thinking skills will find it difficult to compete in a global world (Frijters et al., 2006). Critical thinking skills in the learning process are needed, where when students receive knowledge transfer, they are not immediately taken for granted, but there is a process of classification, analysis, and evaluation carried out to filter this information. The average critical thinking ability of students in geometric optical material is still low with an average score of 27.20 out of 100.00, while the highest score is 71.05 and the lowest score is 20.63 (Pradana et al., 2017).

Physics is a science that studies natural phenomena. These natural phenomena are studied mathematically using various symbols or equations, so the ability of students to represent concepts visually greatly affects their understanding of physics concepts. Most humans are visual learners, and if teaching materials are equipped with lots of visualizations, the information will last longer (Felder & Soloman in Hikmat & Efendi, 2011). Representation is something that represents, describes, or symbolizes an object and/or process (Rosengrant et al., 2007). Visual representation ability is the ability to explain a concept using a model that makes it easier for students to solve problems and find solutions, such as using pictures and graphics. Visual representation is another alternative that is used to correct communication errors when conventional methods fail to convey the concept completely (Sankey, 2005).

The results of the interview with the physics teacher at MAN 4 Bantul showed that in the material for optical instruments, a lot of optical devices were discussed, but the media in schools was inadequate, making it difficult for teachers to carry out demonstrations. The laboratory space available in schools is incomplete and rarely used. The material for this optical instrument has many similarities and images that require students' critical thinking and visual representation abilities to understand it. In previous studies, it was known that there were some difficulties experienced by students in the material of optical instruments, namely difficulties in experimenting with optical devices of 38.04%, difficulties in making diagrams of the path of light in reflection and refraction 77.17%, difficulties in solving mathematical problems of 60.87%, and difficulties learning optical instruments in class and outside the classroom (Ainiyah et al., 2020).

Students' critical thinking and visual representation abilities need to be measured with appropriate measuring instruments, one of which is a test. In the development of a critical thinking test on optical device material, after it went through qualitative and quantitative analysis, 38.9% of the test items were accepted, 61.1% were revised, and no items were rejected (Nur'asiah et al., 2015). The use of assessment instruments is effective in improving students' critical thinking skills (Asmawati et al., 2018). The level of students' understanding regarding the optical instruments material can be known by conducting an assessment. Assessments usually carried out by physics teachers to find out students' understanding are using formative and summative tests.

Previous researchers have developed critical thinking instruments on optical material (Pradana et al., 2017), and visual representations of critical thinking skills on optical material (Nur'asiah et al., 2015). In this study, in addition to developing instruments, visualization of students' critical thinking was also carried out on optical material. Critical thinking indicators used in this study include interpreting, formulating problems, analyzing, concluding, evaluating, and developing strategies and tactics. In addition to critical thinking skills, to support students in understanding physics material, visual representation abilities are needed as a medium to help students visualize a concept. The indicators used in visual representation are to represent the equations in the concept of the eye and the loop in the form of an image.

RESEARCH METHOD

This research was conducted to develop an assessment instrument so the method used was research and development (R&D), with the aim of developing an instrument for assessing students' critical thinking and visual representation abilities. The subjects of this study are 61 grade XI students of MAN 4 Bantul. The development model used is ADDIE (analyze, design,

develop, implement, and evaluate) model. This ADDIE model is basically for the development of learning instructional products. Meanwhile, in this research, the product being developed is a test that has a specific flow and process, so the ADDIE development stage is integrated with the test development stage developed by Mardapi which includes: (1) determining instrument specifications, (2) writing the instruments, (3) determining the scale of the instrument, (4) determining the scoring system, (5) examining the instrument, (6) conducting trials, (7) analyzing the instrument, (8) assembling the instrument, (9) carrying out the measurement, and (10) interpreting the measurement results (Mardapi, 2012). The data collection in this study was conducted through observation, interviews, and documents in the form of photos and videos. The research flow design is shown in Figure 1.

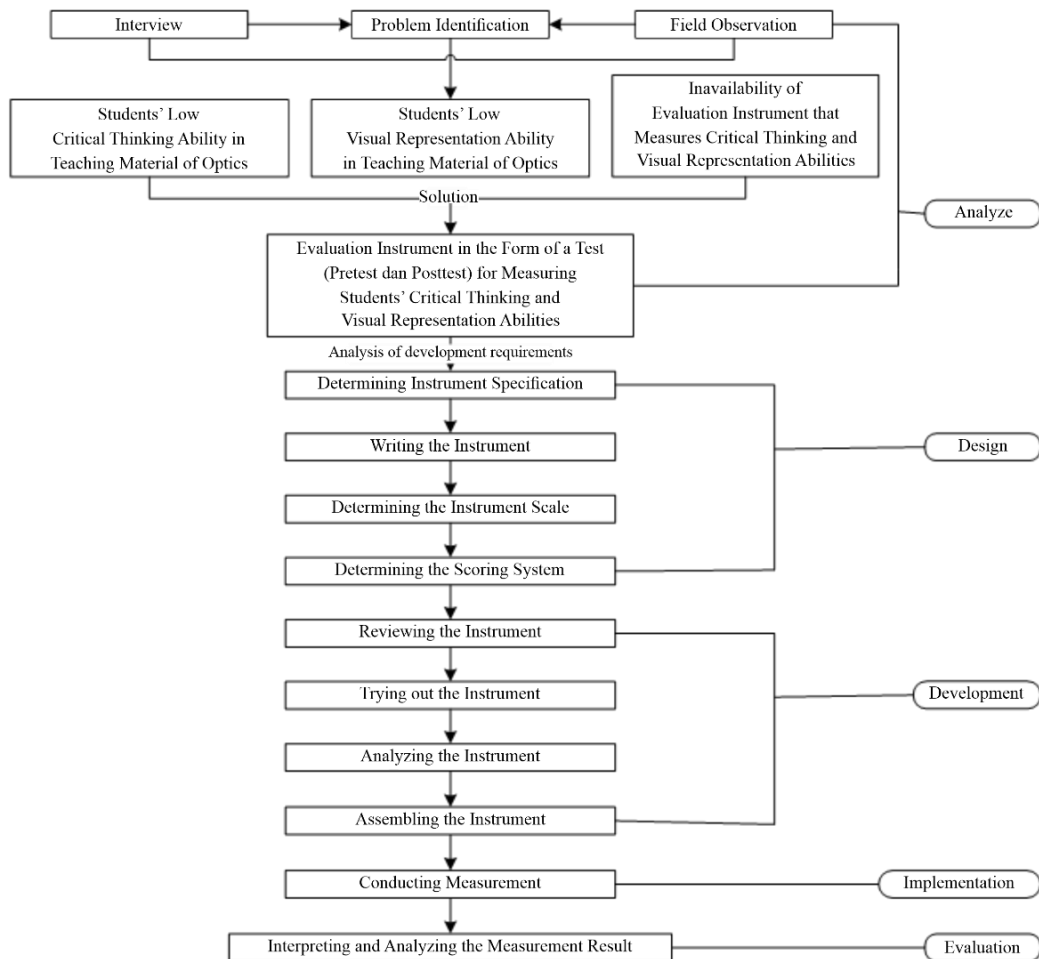


Figure 1. Research Design

The instrument validity analysis uses Aiken's V. Aiken formulates Aiken's V to calculate the content-validity coefficient, which is based on the results of an expert panel's assessment of an item in terms of the extent to which the item represents the construct being measured (Hendryadi, 2014). The formula proposed by Aiken is presented in Formula (1).

$$V = \frac{\sum s}{[n(c-1)]} \dots\dots\dots (1)$$

Notes:

- n_r = number of raters
- s = rater scale
- l_o = the lowest validity rating score (i.e. 1)
- c = the highest validity rating score (i.e. 5)
- r = the score given by the rater
- $s = r - l_o$

Aiken's V coefficient values range from 0 to 1. If half of the raters say an item can be used, it means the item is valid. The criteria for the level of validity are shown in Table 1 (Retnawati, 2016).

Table 1. Criteria for the Level of Validity

No.	Score Range (V)	Level of Validity
1.	$V < 0.4$	Low Validity
2.	$0.4 - 0.8$	Moderate Validity
3.	$V > 0.8$	High Validity

Source: Retnawati (2016)

The items of the instrument for assessing critical thinking and visual representation abilities were analyzed using the QUEST program to determine the validity, reliability, and level of difficulty of the items. The data analyzed were the results of the pre-test and post-test. The items analyzed using the QUEST program are declared valid if the Infit Mean Square (MNSQ INFIT) value ranges from 0.77 to 1.30. The results of the analysis are shown in Table 2 and Table 3. Meanwhile, the critical thinking indicators and visual representations used in this study are described in Table 4 and Table 5.

Table 2. MNSQ Infit Determination

INTIF MNSQ Score	Notes
> 1.30	Incompatible with Rasch Model
$0.77 - 1.30$	Compatible with Rasch Model
< 0.77	Incompatible with Rasch Model

Table 3. T-Outfit Determination

OUTFIT T Score	Category
$OUTFIT T \leq 2.00$	Compatible with Rasch Model
$OUTFIT T > 2.00$	Incompatible with Rasch Model

Table 4. Critical Thinking Indicator

No.	Indicator	Sub-Skills
1.	Interpreting	Categorizing, explaining the significance, and explaining the meaning.
2.	Formulating the problem	Defining terms and identifying assumptions
3.	Analyzing	Examining ideas, identifying arguments, and identifying reasons and claims
4.	Concluding	Questioning the evidence, surmising several alternatives, and drawing conclusions deductively or inductively
5.	Evaluating	Asking for results, justifying procedures, and giving reasons
6.	Developing strategies and tactics	Self-monitoring, and self-correcting

Table 5. Visual Representation Indicator

No.	Indicator	Sub-Skills
1.	Students can represent equations in the eye concept in the form of images.	Categorizing, explaining the significance, and explaining the meaning.
2.	Students can represent the equations in the loop concept in the form of pictures.	Categorizing, explaining the significance, and explaining the meaning.

FINDINGS AND DISCUSSION

This research was conducted at MAN 4 Bantul regarding the development of an instrument for assessing critical thinking and visual representation abilities. It produced some data in

the form of the results of test instrument validation conducted by four experts as well as pre-test and post-test outputs which were analyzed using the QUEST program.

Findings

Based on the research that has been done, the results of instrument validation using Aiken's V were obtained. The detailed results are presented in Table 6.

Table 6. Results of the Validation of Critical Thinking and Visual Representation Assessment Instruments

Item	Score				Analysis Process						Aiken's Value	Criteria
	Val. 1	Val. 2	Val. 3	Val. 4	S1	S2	S3	S4	ΣS	n(c-1)		
Item 1	2	3	3	3	1	2	2	2	7	8	0.875	Highly Valid
Item 2	2	3	2	2	1	2	1	1	5	8	0.625	Moderately Valid
Item 3	2	3	2	3	1	2	1	2	6	8	0.750	Moderately Valid
Item 4	2	3	2	2	1	2	1	1	5	8	0.625	Moderately Valid
Item 5	2	3	2	2	1	2	1	1	5	8	0.625	Moderately Valid
Item 6	2	3	2	3	1	2	1	2	6	8	0.750	Moderately Valid
Item 7	2	2	2	3	1	1	1	2	5	8	0.625	Moderately Valid
Item 8	2	3	2	3	1	2	1	2	6	8	0.750	Moderately Valid
Item 9	2	3	2	3	1	2	1	2	6	8	0.750	Moderately Valid
Item 10	2	2	2	3	1	2	1	2	6	8	0.750	Moderately Valid

Note: Val. = Validator

The Output of the Pre-test and Post-test Using the QUEST Program

The analysis of the instrument for assessing critical thinking and visual representations based on pre-test and post-test scores using the QUEST program was carried out to determine the validity of the items, the reliability of the items, and the level of difficulty of the items. The results obtained are shown in Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, and Figure 7. In addition, the results of the pre-test and post-test scores of students in two different classes are shown in Table 7 and Table 8.

Case Estimates
all on all (N = 56 L = 15 Probability Level= .50)

Summary of case Estimates

Mean	-.57
SD	1.03
SD (adjusted)	.59
Reliability of estimate	.32

Fit Statistics

Infit Mean Square		Outfit Mean Square	
-------------------	--	--------------------	--

Mean	.98	Mean	1.18
SD	.38	SD	1.72

Infit t		Outfit t	
---------	--	----------	--

Mean	.02	Mean	.18
SD	1.03	SD	.85

1 cases with zero scores
0 cases with perfect scores

Case Estimates
all on all (N = 38 L = 15 Probability Level= .

Summary of case Estimates

Mean	-.03
SD	.87
SD (adjusted)	.49
Reliability of estimate	.32

Fit Statistics

Infit Mean Square		Outfit Mean Square	
-------------------	--	--------------------	--

Mean	1.00	Mean	1.03
SD	.22	SD	.36

Infit t		Outfit t	
---------	--	----------	--

Mean	.00	Mean	.13
SD	.86	SD	.67

0 cases with zero scores
0 cases with perfect scores

Figure 2. Validity and Reliability of Test Instruments (Pre-test)

Figure 3. Validity and Reliability of Test Instruments (Post-test)

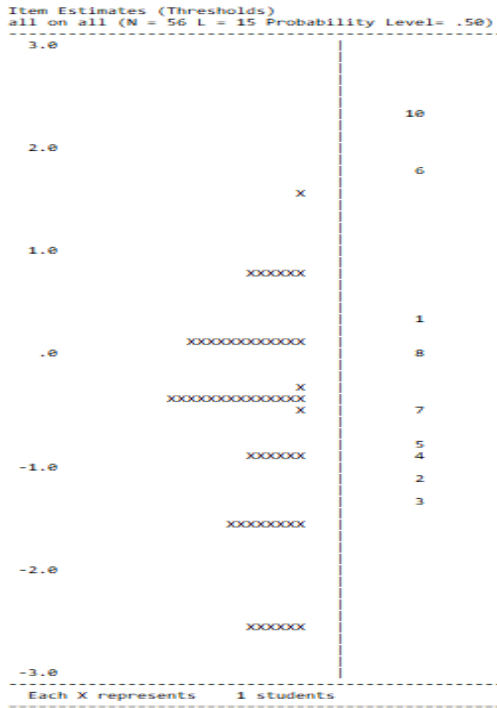


Figure 4. Difficulty Level of Critical Thinking and Visual Representation Questions of Test Instruments (Pre-test)

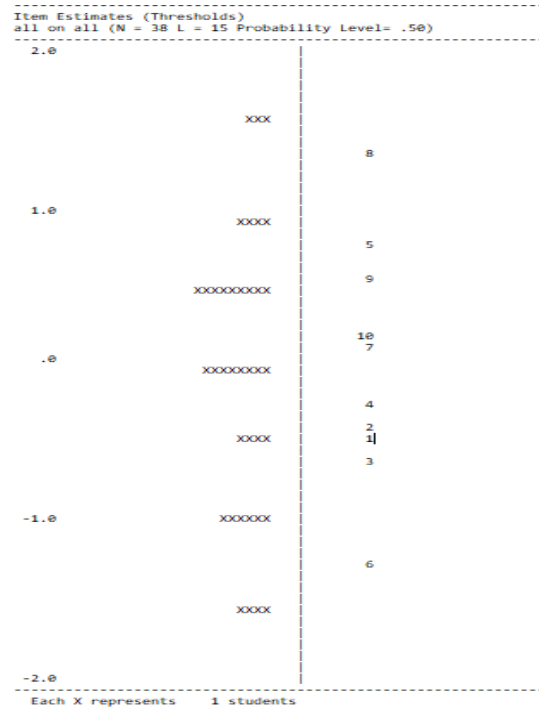


Figure 5. Difficulty Level of Critical Thinking and Visual Representation Questions of Test Instruments (Post-test)

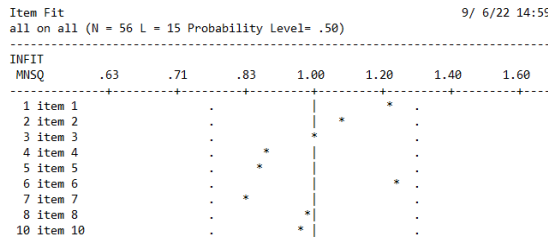


Figure 6. Compatibility of Partial Credit Model (1-PL) Questions of Critical Thinking and Visual Representation (Pre-test)

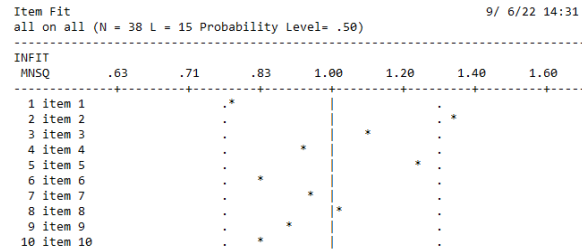


Figure 7. Compatibility of Partial Credit Model (1-PL) Questions of Critical Thinking and Visual Representation (Post-test)

Table 7. Average Scores of Students on the Pre-test and Post-test

Class	Pre-test Result	Post-test Result
XI MIPA 1	32.08	44.38
XI MIPA 2	37.94	59.17

Table 8. Pre-test and Post-test Output Results Using the QUEST Program

Class	MNSQ Infit	Outfit MNSQ	Infit T	Outfit T
Pre-Test	0.85 – 1.15	0.45 – 1.89	-1.04 – 0.88	-1.02 – 1.08
Post-Test	0.81 – 1.17	0.72 – 1.32	-1.35 – 1.19	-1.01 – 1.13

Based on the results of the pre-test analysis, the mean square infit was in the range 0.77-1.30 and the outfit t value ≤ 2 , so all items fit the Rasch Model. Based on the results of the post-test analysis, the mean square infit was in the range 0.77-1.30 and the outfit t value ≤ 2 , so all items fit the Rasch Model.

Students' Pre-test Answers

The answers of students' pre-test are presented in Figure 8 and Figure 9.

Sebutkan dan jelaskan secara singkat macam-macam cacat mata!

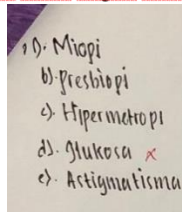


Figure 8. Student Answers about Eye Defects

Gambarkan pembentukan bayangan pada mata normal, mata rabun jauh, dan mata rabun dekat!

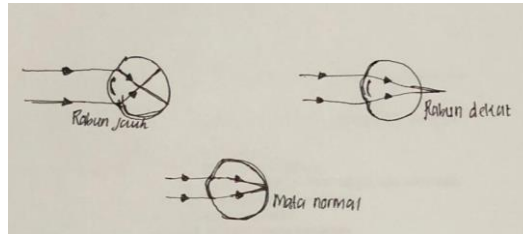


Figure 9. Students' Answers about Shadow Formation

Students' Answers to Post-test Questions

Students' answers to post-test questions are presented in Figure 10, and Figure 11.

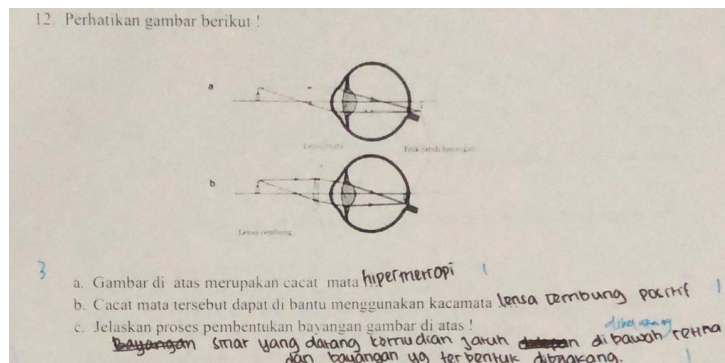


Figure 10. Students' Answers about Eye Defects and Shadow Formation

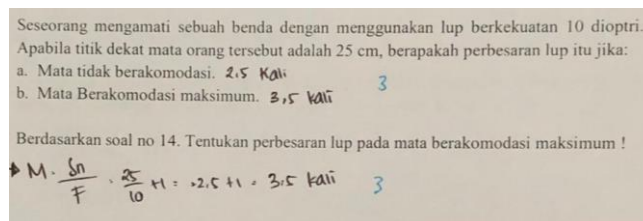


Figure 11. Students' Answers about the Magnification of the Loop

Discussion

The assessment instrument that was developed used the Rasch Model by utilizing the valid and reliable Quest application (Sari, 2020). In this study, the validation of the assessment instrument was carried out by four experts as presented in Table 6, which shows that the average rating instrument obtained moderately valid and highly valid criteria. With reference to Table 1, moderate validity has Aiken's Value which is in the range of 0.4-0.8, and high validity has Aiken's Value > 0.8 . Therefore, this indicates that the assessment instrument is ready to be tested on students.

The process of testing the instrument was carried out in class XI involving a total of 61 students. The pre-test was administered at the beginning of a lesson to find out the extent to which students' abilities were. Then at the last meeting on optical instrument material, students were again given post-test questions which were almost similar to the pre-test questions given earlier. After that, an analysis was carried out using the QUEST program to measure the validity, reliability, and difficulty level of the questions.

The way to use the Quest application is by typing all the students' answers into a notepad and making syntax. Student syntax and answers were stored in notepad form and stored in the same folder as the Quest application. Failure in analysis using the QUEST program often occurs, this is because there is something wrong in the syntax or the student's answer file so that the output results do not come out. Therefore, making syntax and writing students' answers must be done very carefully so that no failure occurs in the analysis process.

One of the Quest outputs is the Reliability of Item Estimate. Reliability of Item Estimate or reliability value based on item estimation, also known as sample reliability. The higher the reliability value, the more the items that fit or match the model being tested. Conversely, the lower the reliability value, the more the items that do not fit or match the model being tested. The developed critical thinking instrument has high reliability, which is equal to 0.86, and is suitable for use as a good measuring instrument (Mukti & Istiyono, 2018). Reliability is the level of consistency or constancy of an item, so when the reliability value is low, it cannot provide the expected information. Determination of fit items in the QUEST program as a whole is based on the average value of the Infit Mean of Square (MNSQ Infit) and its standard deviation. Determination of the fit of each item with the model in the QUEST program is based on the magnitude of the MNSQ Infit value or the Outfit t-value of the item concerned. The determination of MNSQ Infit and Outfit is for the Rasch model.

Figure 2 shows a Reliability of the Case Estimate or a summary of students with a value of 0.32. The reliability value shows that the higher the value, the more convincing that the measurement gives consistent results. However, if the reliability value is low, it means that students are inconsistent. The mean of the case estimate shows that the ability of students is lower than the item difficulty level. The determination of the overall fit case with the model in the QUEST program is based on the MNSQ Infit value and its standard deviation. Therefore, the output of the QUEST program analysis related to the estimated reliability of the case obtained the average MNSQ Infit value of 0.98 with a standard deviation of 0.38. When viewed more closely by relating the standard deviation, the MNSQ Infit value is 0.98 ± 0.38 or $0.98-0.38 = 0.60$ to $0.98+0.38 = 1.36$. Based on the results of the analysis, it was found that the MNSQ Infit was 0.60-1.36 so some were not in the range 0.77-1.33 so that, overall, the items were in accordance with the Rasch Model.

Figure 3 shows that the Reliability of the Case Estimate is a value of 0.32. The Mean of Case Estimate shows that the ability of students is lower than the item difficulty level. The determination of the overall fit case with the model in the QUEST program is based on the MNSQ Infit value and its standard deviation. Therefore, the output of the QUEST program analysis related to the estimated reliability of the case obtained the average MNSQ Infit value of 1.00 with a standard deviation of 0.22. When viewed more closely by relating the standard deviation,

the MNSQ Infit value is 1.00 ± 0.22 or $1.00 - 0.22 = 0.78$ to $1.00 + 0.22 = 1.22$. Based on the results of the analysis, the MNSQ Infit is 0.78-1.22, which means it is in the range 0.77-1.30 so overall the items are in accordance with the Rasch Model.

Figure 4 shows that x is the identity of the subject, while the number opposite it is the item number. The figure shows that overall, the ability of the subject is lower than the difficulty level of the questions. Question 10 shows the highest item difficulty level because there are no students at all who can answer the item correctly. Meanwhile, Question 3 shows the lowest item difficulty level (easy item), although there are still some students who cannot answer it correctly.

The difficulty level is in the ability range of -2 to +2 (Hambleton & Swaminathan, 1985) and the developed test instrument has a good level of difficulty with a range of -2.00 and 2.00 (Mukti & Istiyono, 2018). As is the case in Figure 5, x is the identity of the subject, while the number opposite x is the item number. The figure shows that overall, the ability of the subject is lower than the level of difficulty of the item, but there is also something higher than the level of difficulty of the item. Question 8 shows the highest item difficulty level, but there are some students who can answer the item. Meanwhile, number 6 shows the lowest item difficulty level (easy item), there are still some students who cannot answer it.

Each item is declared fit or appropriate if the MNSQ Infit value is between ≥ 0.77 to ≤ 1.30 (Adams & Kho: 1996; Subali & Suyata: 2012). Figure 6 shows that all of the ten items are within the line in the range of 0.77-1.30. Meanwhile, Item 3 is right on the midline, which indicates a very good item. Overall, these items are in accordance with the Rasch model. Figure 7 shows that all of the ten items are within the line in the range of 0.77-1.30. Overall, the items are in accordance with the Rasch model, except for Item 3, which is outside the line.

Figure 8 shows the students' answers during the pre-test, which is about eye defects. The answers are still incomplete because they mention only the types of eye defects, even though the instructions contained in the questions are to mention and explain eye defects. The problem is a matter of critical thinking, where there is a process of representing and identifying or formulating a problem.

Figure 9 shows the students' answers during the pre-test about shadow formation. The student's answers are correct because they are in accordance with the instructions, namely to describe the shape of normal eye shadow, nearsightedness, and farsightedness. This question is a matter of visual representation, where students must represent the concept in the form of an image.

Figure 10 shows the students' answers during the post-test about eye defects and shadow formation. The answer is correct because the students answered all the questions and filled them in with the correct answers. This question is a matter of visual representation, where students must analyze the images presented, and then answer questions based on these images.

Figure 11 shows the students' answers during the post-test about the magnification of the loop. Their answers to the questions are correct but incomplete, because in one of the questions students only give answers directly, meaning they do not write down the steps for solving them. The questions are concerned with critical thinking, where there is a process of analysis, evaluation, and preparation of strategies and tactics to solve these questions in order to get the right answers.

Based on the results of these answers, students on the pretest and posttest still have low critical thinking skills. This is in line with the findings of the research by Makhrus et al. (2020) which reported that critical thinking skills are still low, and then given teaching treatments, the students show higher critical thinking skills. Other research also provides learning treatments through animation (Disman et al., 2020) and the application of learning models (Murniati et al., 2020). The importance of implementing critical thinking learning needs special attention, to prepare all components that support critical thinking skills, instruments, media, and appropriate teaching materials.

CONCLUSION

Based on the results of this research, which was conducted at MAN 4 Bantul, the following conclusions can be drawn. (1) The results of the analysis of instrument validation using Aiken's V show an average Aiken's Value ≥ 0.4 with highly valid criteria and 0.4-0.8 with moderately valid criteria, so it can be said that all items in the instrument are valid. (2) The pre-test analysis on item estimates obtained by MNSQ Infit was in the range of 0.77-1.30 and the value of outfit $t \leq 2$, so all of the items matched the Rasch Model, and in the MNSQ Infit case estimates were in the range 0.77-1.30, so all of the items are in accordance with the Rasch Model. (3) Post-test analysis on item estimates obtained by MNSQ Infit is in the range of 0.77-1.30 and the value of outfit $t \leq 2$, so all of the items are in accordance with the Rasch Model, and in the MNSQ Infit case estimates are in the range of 0.77-1.30 so all of the items are in accordance with the Rasch Model. (4) The reliability of the pre-test items is 0.92 while the reliability of the post-test items is 0.76. (5) In terms of the item difficulty level, on the pre-test, Question 10 is the item with the highest item difficulty level, while Question 3 is the item with the lowest item difficulty level (easy item), and on the post-test, Question 8 is the item with a high difficulty level, while Question 6 is the item with the lowest item difficulty level (easy item).

REFERENCES

- Ackerman, G. L. (2010). *Brief review of Fadel, C., & Trilling, B. (2009) 21st century skills: Learning for life in our times*. Jossey-Bass.
- Adams, R. J., & Kho, S.-T. (1996). *Acer Quest version 2.1*. The Australian Council for Educational Research.
- Ainiyah, Q., Yuliati, L., & Parno, P. (2020). Analisis penguasaan konsep dan kesulitan belajar materi Alat-Alat Optik pada siswa kelas XI MAN Tuban. *Jurnal Riset Pendidikan Fisika*, 5(1), 24–29. <http://journal2.um.ac.id/index.php/jrpf/article/view/15904>
- Asmawati, E. Y. S., Rosidin, U., & Abdurrahman, A. (2018). Efektivitas instrumen asesmen model creative problem solving pada pembelajaran Fisika terhadap kemampuan berpikir kritis siswa. *JPF (Jurnal Pendidikan Fisika) FKIP UM Metro*, 6(2), 128-143. <https://ojs.fkip.ummetro.ac.id/index.php/fisika/article/view/1318>
- Butler, P. J., Dong, C., Snyder, A. J., Jones, A. D., & Sheets, E. D. (2008). Bioengineering and Bioinformatics Summer Institutes: Meeting modern challenges in undergraduate summer research. *CBE Life Sci Educ.*, 7(1), 45–53. <https://doi.org/10.1187/cbe.07-08-0064>
- Disman, D., Kurniawati, S., & Sukamto, R. A. (2020). Peningkatan berpikir kritis melalui metode pembelajaran berbasis masalah dengan animasi analisis grafis. *Jurnal Pendidikan Ekonomi Indonesia*, 2(1), 59-66. <https://ejournal.upi.edu/index.php/JPEI/article/view/28345>
- Frijters, S., ten Dam, G., & Rijlaarsdam, G. (2006). Effects of dialogic learning on value loaded critical thinking. *Learning and Instruction*, 18(1), 66–82. <https://doi.org/10.1016/j.learninstruc.2006.11.001>
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Kluwer Inc.
- Hendryadi, H. (2014). Content validity (Validitas isi). *Teorionline Personal Paper*, 01, 1–5. <https://teorionline.files.wordpress.com/2014/07/010614-content-validity.pdf>
- Hikmat, H., & Efendi, R. (2011). Representasi momentum dan impuls melalui diagram. In *Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA*, pp. 207-211. Retrieved on June 13, 2022, from

http://file.upi.edu/Direktori/FPMIPA/JUR._PEND._FISIKA/AHMAD_SAMSUDIN/Publikasi/26PFis_Hikmat.pdf

- Luthvitasari, N., Made D. P., N., & Linuwih, S. (2012). Implementasi pembelajaran Fisika berbasis proyek terhadap keterampilan berpikir kritis, berpikir kreatif, dan kemahiran generik sains. *Journal of Innovative Science Education*, 1(2), 92-97. <https://journal.unnes.ac.id/sju/index.php/jise/article/view/630>
- Makhrus, M., Ayub, S., Wahyudi, W., Verawati, N. N. S. P., & Busyairi, A. (2020). Kemampuan berpikir kritis mahasiswa pada materi Optik saat proses pembelajaran dengan CCM-CCA berbantuan eksperimen virtual. *Kappa Journal Program Studi Pendidikan Fisika*, 4(2), 143-148. <https://doi.org/10.29408/kpj.v4i2.2632>
- Mardapi, D. (2012). *Pengukuran penilaian & evaluasi pendidikan*. Nuha Medika.
- Mukti, T. S., & Istiyono, E. (2018). Instrumen penilaian kemampuan berpikir kritis peserta didik SMA negeri mata pelajaran Biologi kelas X. *BIOEDUKASI: Jurnal Pendidikan Biologi*, 11(2), 105–110. <https://jurnal.uns.ac.id/bioedukasi/article/view/21624>
- Murniati, N., Febyasha, R., & Irwandi, I. (2020). Kemampuan berpikir kritis dan hasil belajar kognitif siswa menggunakan model Pembelajaran Kooperatif tipe Think Pair Share (TPS) di MAN 2 Kota Bengkulu. *Diklabio: Jurnal Pendidikan dan Pembelajaran Biologi*, 4(1), 94-105. <https://doi.org/10.33369/diklabio.4.1.94-105>
- Nur'asiah, R. R. F., Siahaan, P., & Samsudin, A. (2015). Deskripsi instrumen tes keterampilan berpikir kritis materi Alat Optik. In *Symposium Nasional dan Pembelajaran Sains 2015 (SNIPS 2015)*, pp. 497–500. https://ifory.id/proceedings/2015/z4pZjcjkq/snips_2015_rd._risma_farissa_nurasiah_63ad41e19ce5254d4ee1531011f52977.pdf
- Pradana, S. D. S., Parno, P., & Handayanto, S. K. (2017). Pengembangan tes kemampuan berpikir kritis pada materi Optik Geometri untuk mahasiswa Fisika. *Jurnal Penelitian dan Evaluasi Pendidikan*, 21(1), 51-64. <https://doi.org/10.21831/pep.v21i1.13139>
- Regulation of the Minister of Education and Culture No. 23 of 2016 concerning the Standard for Educational Assessment*. (2016).
- Retnawati, H. (2016). *Analisis kuantitatif instrumen penelitian*. Parama Publishing.
- Rosengrant, D., Etkina, E., & Van Heuvelen, A. (2007). An overview of recent research on multiple representations. *AIP Conference Proceedings*, 883(1), 149–152. <https://doi.org/10.1063/1.2508714>
- Sankey, M. (2005). Multimodal design and the neomillennial learner. In *Proceedings of OLT 2005 Conference*, Brisbane, Australia, 27 September 2005, pp. 251-259. <https://research.usq.edu.au/item/9x99z/multimodal-design-and-the-neomillennial-learner>
- Sari, D. K. (2020). Analisis instrumen penilaian kemampuan pemodelan matematis pada kelas Fisika menggunakan Rasch model. *MEGA: Jurnal Pendidikan Matematika*, 1(1), 46–52. <https://doi.org/10.59098/mega.v1i1.182>
- Subali, B., & Suyata, P. (2012). *Pengembangan item tes konvergen dan divergen dan penyelidikan validitasnya secara empiris*. Diandra Pustaka Indonesia.