JRERE 12 (1) 2023 1-11

Journal of Research and Educational Research Evaluation

http://journal.unnes.ac.id/sju/index.php/jere

Developing a Diagnostic Test of Critical Thinking Skills In Physics Lessons

Muh. Asriadi AM[1]✉, Samsul Hadi[2]

[1]Universitas Pendidikan Indonesia, Indonesia
[2]Universitas Negeri Yogyakarta, Indonesia

Article Info

History Articles
Received:
15 December 2022
Accepted:
10 January 2023
Published:
30 March 2023

Keywords:
Diagnostic Tests, Critical Thinking, Physics.

Abstract

A diagnostic test of critical thinking ability is a method to reveal the location of students' weaknesses and strengths in critical thinking skills, especially in physics subjects. This study aims to develop a diagnostic test of quality critical thinking skills in physics subjects regarding validity and reliability and the quality of items in diagnosing students' critical thinking skills. This research is development research whose development procedure uses a combination of the Ministry of National Education development model (2007) with the test development procedure from Mardapi (2017), which consists of 11 steps, namely: (1) identifying essential competencies that have not been completed, (2) determining possible root causes of problems, (3) developing test specifications, (4) writing tests, (5) assessing tests by experts, (6) conducting test trials, (7) analyzing the quality of test items, (8)) improving tests, (9) assembling test, (10) conducting the test, (11) interpreting the test result. The subjects of this study were MAN students in the city of Yogyakarta. Data were analyzed by CVR, EFA, and PCM tests. The results showed that the diagnostic test model for critical thinking skills was valid, reliable, and had good item quality in diagnosing students' critical thinking skills. The diagnostic test results show that MAN students' critical thinking ability in the city of Yogyakarta in physics subjects is more dominant in the low category with a percentage of 35.30%. As a result of knowing the description of students' critical thinking weaknesses, the findings of this study can be used as a reference, particularly for physics teachers, to improve the quality of physics learning.

✉Correspondence Address :
Department of Early Childhood Education, Cibiru Campus,
Universitas Pendidikan Indonesia, Indonesia 40625
E-mail : asriadi190197@gmail.com

p-ISSN 2252-6420
e-ISSN 2503-1732

INTRODUCTION

The 21st century is marked by the rapid development of information technology and science. They started with technology, science, society, economics, and education. Along with this, each country is expected to have human resources that can adapt to these developments (Asriadi & Istiyono, 2020). One of the critical skills needed in the world of work in the 21st century is the ability to think critically (Djatmiko et al., 2022). In PISA, this ability becomes an essential part of the assessment (AM & Hadi, 2021). Critical thinking can essentially be developed through education and training. The teacher, as a facilitator, designs and implements various learning strategies to train and stimulate students so that their critical thinking skills can be developed (Amin̈ dkk., 2020).

According to (Ennis, 2011), critical thinking is introspective thinking that focuses on deciding what to believe and do. Meanwhile, according to (Mason & Singh, 2016), the concept of critical thinking is perhaps one of the most important trends in education, and this trend is dynamically linked to how teachers teach and students learn. Critical thinking is a complex skill that allows a person to obtain information, collect data, and evaluate findings effectively (Fitriani et al., 2020). Critical thinking is a metacognitive process through intentional and self-regulating reflective evaluation (i.e., including sub-skills of analysis, evaluation, and inference) and disposition towards thinking (i.e., the degree to which a person is inclined or willing to perform specific thinking skills), increasing the likelihood of logical conclusions or solutions problem (Dwyer & Walsh, 2020). Fisher and Scriven define critical thinking as "the skillful and active interpretation and evaluation of observation, communication, information, and argumentation" (de Glopper, 2002). Critical thinking is the essence of innovative thinking that shows a person's level of thinking ability (Baker, 2020). Critical thinking includes scientific thinking processes that extend to identifying and defining problems, gathering information to study the problem in-depth, analyzing situations concerning the problem, generating potential solutions to the problems being addressed, and evaluating proposed solutions. (Warsah et al., 2021). Based on some of the points above, it can be concluded that critical thinking skills are a process that a person must do to obtain appropriate and reasonable results or decisions. These issues are usually made from facts, information, and knowledge. Training students to master critical thinking skills is intended so that attitudes and perceptions are obtained, actualized, integrated, and advanced knowledge gained (Himmatussolihah et al., 2020). To detect the development of students' critical thinking skills, they can be done by giving tests (Roohr & Burkander, 2020).

Physics as a subject at the high school level has particular competencies. Most students consider physics one of the most challenging subjects at the high school level, especially those who take the science program (AM & Istiyono, 2022). One of the things that can overcome difficulties in learning physics is to improve the critical thinking skills of each student. So it is necessary to identify how many critical thinking skills in physics subjects are owned by students. The required tests are referred to as diagnostic tests.

Diagnostic tests can be used to accurately determine the strengths and weaknesses of students under certain conditions (Zaleha et al., 2017). The diagnostic test results can be used to analyze students' abilities individually or in groups to improve the learning process. Diagnostic tests can also be used to outline the form of assignments by students' character (Hadi et al., 2015). This test will help analyze students' learning difficulties and as an effort to improve learning activities (Triumiana & Sudarsono, 2017). Diagnostic tests are carried out after learning but before summative tests, which aim to provide treatment or improvement when problems are found (Rusilowati, 2015) so that the diagnostic test identifies the participant's weaknesses and then provides follow-up according to the problems experienced.

Diagnostic tests have various forms of multiple-choice testing, including single-layer, two-layer, three-layer, and four-layer multiple-choice diagnostic tests. However, in this study, a two-tier multiple choice is used or what is known as a two-tier multiple choice. The two-level test is a diagnostic instrument consisting of questions containing choice content and reasons for choosing the answer (Gurel, 2015). If the answer choices and reasons are given correctly, the answer to the question is correct. Consider the reasons or explanations behind the student's answer choices and how their choices relate to the misunderstanding of the target concept (Tam, 2016). Two-tier multiple-choice tests are relatively convenient for students and are more practical and valuable for teachers in reducing guesswork, allowing for easy large-scale management and evaluation, and providing insight into student reasoning. (Xiao et al., 2018). The advantages of this instrument are that it is easy to implement in large classes, has a high level of objectivity, has comprehensive material coverage, can be improved, and requires students to think about determining the right reasons. In addition, using reasoning when answering multiple-choice tests is a sensitive and effective way to add value to the completeness of their answer choice questions (A. Z. Abidin et al., 2019). A two-level multiple choice test can determine a student's ability to learn critical thinking skills in physics. It is known that the basic concepts of matter physics require critical reasoning from natural phenomena (Chang et al., 2007). Based on the previous description, this research can be used to diagnose the level of students' critical thinking skills. In addition, it can be used as a reference, especially for physics teachers, to improve the quality of physics learning by revealing the location of students' weaknesses and strengths in critical thinking skills in physics subjects. Therefore, the purpose of this study is to develop a construct of a diagnostic test instrument for critical thinking skills that are valid, reliable, and has good quality questions and diagnoses students' critical thinking skills in physics subjects.

METHODS

This type of research is development research. In this study, the development model combines the guidelines for developing diagnostic tests made by (Depdiknas RI, 2007) with test development procedures by Mardapi (2017). The combination referred to in this study is to combine the guidelines for developing diagnostic tests from the Ministry of National Education, which consists of 7 development steps, with the Mardapi test development model, which has nine development steps. Therefore, there are 11 steps of developing a diagnostic test model for critical thinking skills, namely 1) identifying incomplete essential competencies, 2) determining possible root causes of problems, 3) developing test specifications, 4) writing tests, 5) assessing tests by experts., 6) carry out test trials, 7) analyze the quality of test items, 8) improve tests, 9) assemble tests, 10) conduct tests, 11) interpret test results. The subjects in this study were students of class XI science program at MAN 1 Yogyakarta and MAN 2 Yogyakarta, each school being sampled by two classes. The instruments used are validation sheets, critical thinking ability diagnostic tests, interview guidelines, and test scoring rubrics. All instruments will be tested for validity and reliability using Content Validity Ratio (CVR), exploratory factor analysis, and Cronbach's Alpha formula. The data analysis technique used to test the quality of the items is the Partial Credit Model (PCM). This analytical technique will analyze the items using the polytomous type by expressing the level of interest in the questions and testing the quality of each item. In addition, the quantitative descriptive analysis was used to determine the results of diagnostic tests. This descriptive analysis will describe the results of diagnostic tests by classifying the position of students' strengths and weaknesses in quantitative critical thinking.

RESULTS AND DISCUSSION

Instrument Construct

The critical thinking ability diagnostic test instrument developed in this study modifies the Watson & Glaser Critical Thinking Skill Appraisal (WGCTA) model (Possin, 2014). by forming it into a two-tier multiple choice test. The final product of the critical thinking ability test model in physics subjects is a test designed to diagnose students' critical thinking skills, especially in physics subjects. The diagnostic test instrument for critical thinking skills is adjusted to the indicators on the physics material. The indicator of critical thinking ability uses critical thinking concepts used in the Watson-Glaser Critical Thinking Assessment (WGCTA) test, which consists of Conclusion (Inference), Recognition Assumption, Deduction (Deduction), Interpretation (Interpretation), Argument evaluation (Evaluation). of arguments).

This indicator is then adjusted to the physics material. The physics material chosen is mechanics, which consists of sub-materials of uniform, straight motion, uniformly changing straight motion, circular motion, parabolic motion, particle dynamics, work and energy, momentum & impulse, rotational dynamics, rigid body equilibrium, and elasticity. This material was chosen because it contains a series of basic sub-subjects that require reasoning and the use of mathematical formulae formulations. This material also has a difficulty level of easy, medium, and complex questions, so it is very suitable for diagnosing students' critical thinking skills at MAN in Yogyakarta. This test is in the form of two-level multiple choice consisting of 15 questions. The test is used to diagnose students' critical thinking skills in physics learning which consists of five aspects: inference, assumption recognition, deduction, interpretation, and argument evaluation.

Validation Analysis

Content Validity

The results of content validity given by experts (expert judgment) on the critical thinking ability diagnostic test instrument have met the valid criteria based on the CVR value of each item, which is 0.99. Meanwhile, the CVI value of the test instrument obtained a value of 0.99, which was in the excellent category. So it can be concluded that the critical thinking ability diagnostic test instrument has met the valid criteria and can then be used at the product trial stage.

Construct Validity

This study's analysis of construct validity (Construct Validity) used EFA (Exploratory Factor Analysis) analysis. Prerequisites for using factor analysis include the KMO & Bartlett test of Sphericity. KMO value is used to see whether the analyzed data is feasible. The criteria are that the KMO value must exceed 0.7 or at least 0.5, and the Bartlett significance value must be below 0.05. The KMO value of 0.762 is more significant than 0.7. The KMO value meets the requirements, and the significance value is 0.000 less than 0.05, so the cross-correlation matrix is not a unit matrix, and factor analysis can be performed. Based on the results of the EFA analysis, it is known that six items are not valid because they have a loading factor value below 0.5. These items are 5, 6, 15, 23, 24, and 30. Apart from these items, they all meet valid criteria and can be used. Invalid items will be excluded from the test instrument.

Reliability Analysis

Based on SPSS analysis, it is known that the reliability value of Cronbach's Alpha is 0.777. Based on the reliability criteria reference, if the value > 0.60-0.80 is included in the category of reliable/reliable. So based on these criteria, the diagnostic test instrument for critical thinking skills in physics subjects is said to be reliable.

Quality Analysis of Question Items

Based on the results of the developed test, it is known that the log-likelihood value is -4713.892, so it can be concluded that the data used can be suitable if using the PCM model. The critical thinking ability diagnostic test will provide a maximum of 38 information with a standard error of 0.7 if given to test takers with a moderate level of ability, which is 0.5. The lower limit of the interval is at theta -1.7, and the upper limit is at theta 2.3. These results indicate that a good test is given in the range of abilities from -1.7 to +2.3. In addition, ten items did not fit. From these results, it can be seen that in measuring the quality of the questions, the patterns of student responses will show the level of difficulty of the items and the student's ability. Items that do not fit are also determined based on the response patterns given by the respondents.

Test Diagnostic Results

The results of the diagnosis of students' critical thinking skills in physics subjects will be presented in the form grouped into 16 diagnostic categories. This category is based on the results of students' analysis of test questions. The test questions with the lowest scores on each indicator will describe the location of students' weaknesses in critical thinking. For the description of the categorization results, see table 1 below.

Table 1. Critical Thinking Ability Diagnostic Category

Category Diagnosing Weaknesses in Critical Thinking	Total Students	Percentage (%)
Conclusion	19	18.63
Conclusion, Interpretation & Evaluation of Arguments	3	2.94
Conclusion & Evaluation of Arguments	3	2.94
Conclusion & Interpretation	5	4.90
Conclusion & Introduction to Assumptions	4	3.92
Conclusion, Recognition of Assumptions, Interpretation, & Evaluation of Arguments	1	0.98
Conclusion, Deduction & Evaluation of Arguments	1	0.98
Assumption Introduction	14	13.73
Introduction to Assumptions & Evaluation of Arguments	2	1.96
Pengenalan Asumsi, Deduksi, & Interpretasi	1	0.98
Deduction	3	2.94
Deduction & Interpretation	1	0.98
Deduction & Evaluation of Argument	2	1.96
Interpretation	17	16.67
Interpretation & Evaluation of Argument	2	1.96
Evaluation of Argument	24	23.53

Based on the results listed in table 1, it is known that the weakness of students in critical thinking is the weakest in the argument evaluation indicator, with a percentage of 23.53%. Then students are also weak in the conclusion indicator with a percentage of 18.63%. Table 1 shows that the two indicators of critical thinking skills must be an essential concern by the teacher and the students if they want to improve their critical thinking skills, apart from grouping the diagnostic category of critical thinking skills. After that, a recapitulation of the score on each question is carried out, which can be seen in table 2.

Table 2. Recapitulation of Scores for Each Question

Item Number	Number of Respondents who scored			
	0	1	2	3
1	73	1	4	24
2	32	14	28	28
3	27	50	8	17
4	10	8	9	75
5	44	34	12	12
6	25	61	1	15
7	30	15	12	45
8	19	38	7	38
9	38	36	16	12
10	28	16	35	23
11	41	19	19	23
12	54	31	8	9
13	43	12	24	23
14	49	22	9	22
15	50	10	34	8

Based on the information from table 2, it is known that respondents who get a score of 0 are dominant on question number 1. Respondents who get a score of 1 are dominant on question number 6. Respondents who get a score of 2 are dominant on question number 10. Moreover, respondents who get a score of 3 are dominant on question number 4. In addition to question number 1, the respondent is more dominant in obtaining a score of 0. In question number 2, the respondent is more dominant, obtaining a score of 0. In question number 3, the respondent is more dominant in obtaining a score of 1. In question number 4, the respondent is more dominant in obtaining a score of 3. In question number 5, the respondent is more dominant, obtaining a score of 0. In question number 6, the respondent is more dominant, obtaining a score of 1. In question number 7, the respondent is more dominant, obtaining a score. 3. In question number 8, the respondent is more dominant, obtaining a score of 1 and 3. In question number 9, the respondent is more dominant. More dominantly get a score of 0. In question number 10, the respondent is more dominant in getting a score of 2. In question number 11, the respondent is more dominant in getting a score of 0. In question number 12, respondents are more dominant in obtaining a score of 0. In question number 13, respondents are more dominant in obtaining a score of 0. In question number 14, respondents are more dominant in obtaining a score of 0. In question number 15, respondents are more dominant in obtaining a score of 0. From these results, a diagnosis can be drawn that the critical thinking ability of MAN students is deficient.

DISCUSSION

Diagnostic test instruments can be used to conduct effective interventions on students individually or classically to evaluate the learning process (Duskri et al., 2014). Diagnostic tests provide information in the form of numbers as an indicator of student ability and describe student mastery in certain sub-ability (Hadi et al., 2015). This test instrument has gone through a series of development stages so that a set of test instruments that are valid, reliable, and have good quality items can be obtained. According to Hadi et al. (2022), the process of developing a test instrument that follows a systematic development procedure will produce a product that is feasible to use. It is also supported by Asriadi & Hadi (2021) that the development of a particular product must go through complex procedural steps. In developing a diagnostic test for critical thinking skills, more emphasis is placed on detecting students' weaknesses in critical thinking so that the test results can be used as material for improvement in learning. Because as stated by Arisoy & Aybek (2021), "good thinking" has become very important in today's world. A good thinker must have critical thinking skills and critical thinking in virtue. Therefore, it is essential to detect and improve students' critical thinking skills for the better.

In this study, the quality of the test instrument can be seen in four aspects, namely the level of content validity, construct validity, instrument reliability, and the quality of test items. Content validity is related to the rational analysis of the domain to be measured to determine the representation with the ability to be measured (M. Abidin & Retnawati, 2019). It has been theoretically completed by focusing on three aspects of the test: material, structural, and language/cultural (Retnawati, 2016). In this study, content validity follows the Content Validity Ratio (CVR) and Content Validity Index (CVI) formulas. According to Lawshe (1975), Validity verification only uses three rating scales, namely (1) essential, (2) helpful but not essential, and (3) not essential. The CVR value has a range from -1 to 1. If half of the SMEs (Subject Matter Experts) state it is essential, then the CVR value will be worth 0. CVR will be worth one if all SMEs (Subject Matter Experts) states that it is essential for an item (Syahfitri et al., 2019). Items that get a negative CVR value are invalid and must be eliminated. The overall test validity value can be determined using CVI (Content Validity Index) (Bashooir & Supahar, 2018). In this study, the diagnostic test instrument has an excellent level of content validity based on the CVR and CVI formulas.

Construct validity refers to the extent to which test items can measure what they want based on a particular concept or defined concept definition (Sustekova et al., 2019). the theoretical process must be carried out starting from the concept of the variable to be measured, formulating the structure, determining dimensions and indicators, to determining and writing the instrument. (Burkholder et al., 2021). The formulation of construction must be based on the synthesis of theories about the concept of the variable to be measured through a process of logic and careful analysis and comparison (Siwi et al., 2020). In this study, construct validation used exploratory factor analysis or EFA (Exploratory Factor Analysis). Exploratory factor analysis detects and evaluates potential sources of variation or covariance in measurements (Faizah et al., 2019). Exploratory factor analysis explores empirical data to find and detect features and relationships between variables without specifying a data model (Retnawati, 2016). In this analysis, the researcher does not have a priori theory to formulate hypotheses but only tests the test items that are feasible and valid in measuring the critical thinking ability variable based on the material construction that has been determined. From the results of the exploratory factor analysis, 13 test items could not be used because they had an MSA (Measure of Sampling Adequacy) value and a loading factor value of less than 0.05. The test items that did not pass the analysis were due to errors in the process of writing test items. That is because the test items used are inappropriate for measuring critical thinking skills. This finding follows the findings of Fajrianthi dkk. (2016) that the choice of the wrong topic in the writing of the items will not be able to measure the critical thinking ability variable. It is also supported by Lia dkk. (2020) that value of the loading factor that does not match the criteria is caused by the incompatibility of the item with the variable being measured.

Instrument reliability testing is related to the instrument's consistency in repeatedly measuring a variable on the same subject and conditions (Retnawati, 2016). In the results of this study, the diagnostic test instrument has met the criteria for reliability. A reliable instrument will be consistent in measuring a variable even though it is tested repeatedly at different times (Kriswantoro et al., 2021). The item quality test was tested using PCM (Partial Credit Model) analysis. This model focuses on contiguous categories when estimating the threshold (i.e., difficulty) between ordered response categories.

Given that an item has K-ordered response categories, PCM estimates a threshold of K 1 for the item (Desjardins & Bulut, 2018). PCM can provide information about students' difficulty levels, which can be obtained through their errors at each difficulty level (Shantika & Istiyono, 2019). The use of the PCM model in measuring the quality of test items is based on research conducted by (Wasis, 2013) stated that the use of PCM in measuring ability in physics subjects can produce an estimate of ability, which is more accurate than other types of assessment, based on the complexity of each option.

Diagnostic testing allows service providers, in this case, teachers or schools to diagnose, monitor, and treat conditions or anticipate changes in student behavior and abilities during the learning process (CLPNA Self-Study Course, 2017). According to research results by Wijaya & Hp (2013), diagnostic tests based on indicators and learning difficulties diagnostic maps are used to diagnose students' learning difficulties. In practice, diagnostic tests in the classroom have two objectives, namely: (1) to identify learning targets that students have not mastered; and (2) to find the causes or reasons that make students unable to master the learning targets (Prihatni et al., 2016). The results of the diagnostic test show that students' critical thinking skills are in the deficient category. When described more specifically based on the constituent indicators, it shows that of the five indicators of critical thinking skills, namely Conclusion, Introduction to Assumptions, Deduction, Interpretation, and Argument Evaluation, all of them are in the shallow categories. That is because so far, students' individual abilities have not been paid attention to, so students with weaknesses will continue to be left behind in learning. Implementing learning in the classroom is generally carried out by looking at students as individuals with the same average ability and habits. The learning process between students who have high abilities is undoubtedly not the same as students who have moderate abilities, even with students whose abilities are less (Türkoguz, 2020). The difficulties and weaknesses of students in learning physics are related to the mastery of a collection of knowledge in the form of facts, concepts, or principles based on the discovery process (di Uccio et al., 2019). The teacher's ability to detect student difficulties and weaknesses is a challenge in physics learning activities to lead students to achieve success in learning. Each student in the class has different characteristics. Individual differences among students can be caused by their way of thinking. Therefore individual differences need to be considered by teachers in learning activities (Ketabi et al., 2021). Teachers must pay attention to individual students; students' weaknesses and difficulties must be considered. Physics learning is currently more emphasis on the sequence of subject matter, not on the thinking process and cognitive psychology of students, so in learning, many students have difficulties and misconceptions (Zaleha et al., 2017). Because teachers play an essential role in overcoming student difficulties and improving the learning process in the classroom, the results of diagnostic tests will be essential input material in improving and improving the quality of physics learners.

An important finding and a novelty in this research lie in the product produced, namely a good and reliable critical thinking ability diagnostic test instrument that can be used to diagnose students' abilities in physics subjects. The second finding that needs attention for educational institutions is that critical thinking skills are still meager. Students' difficulties and weak in learning physics related to mastering a collection of knowledge in the form of facts, concepts, or principles based on the discovery process. In addition, students whose critical thinking skills are deficient have the most dominant weakness in the aspect of argument evaluation or internal ability or the ability to assess the strength and relevance of arguments related to specific problems. All of these findings conclude that the critical thinking ability diagnostic test instrument effectively diagnoses students' critical thinking weaknesses in physics subjects.

CONCLUSION

The diagnostic test instrument for thinking skills in physics subjects was developed based on the adjustment of critical thinking ability indicators with indicators on physics material. The indicator of critical thinking ability uses critical thinking concepts used in the Watson-Glaser Critical Thinking Assessment (WGCTA) test, which consists of Conclusion (Inference), Recognition Assumption, Deduction (Deduction), Interpretation (Interpretation), Argument evaluation (Evaluation). of arguments). This indicator is then adjusted to the physics material. The diagnostic test instrument for critical thinking skills in physics subjects has met the criteria of being valid, reliable, and has good quality items in diagnosing critical thinking skills of MAN students in Yogyakarta.

The test results of students' critical thinking skills in physics subjects are in a low category at 35.30%. This test instrument helps know and analyze the weaknesses of students' critical thinking skills in learning physics. The results of this test can be used as a reference to determine the ability of each student, especially the ability to think critically. Detection of critical thinking skills is essential for the sustainability of a better learning process because it is a competency that must be possessed in the 21st-century learning era.

This research also does not still have limitations, including because this test was carried out in a COVID-19 pandemic. Of course, the results would be different if this test was carried out outside a pandemic. The focus of the test is only on diagnosing critical thinking skills so that it has not revealed the misconceptions that students often experience in physics subjects.

Suggestions from research for further research, namely the results of this test, can be used as a reference to determine the ability of each student, especially the ability to think critically. In addition, it will guide teachers and students in improving and adapting a better way of learning, especially in physics subjects requiring critical thinking skills. For schools, the findings of this study can be used as a reference for formulating policies to form a sound learning system. For further research development, the researcher suggests that this diagnostic test be used as a reference in making similar tests to diagnose other abilities such as creative thinking, communication skills, and collaboration.

REFERENCES

Abidin, A. Z., Istiyono, E., Fadilah, N., & Dwandaru, W. S. B. (2019). A computerized adaptive test for measuring the physics critical thinking skills in high school students. *International Journal of Evaluation and Research in Education (IJERE)*, *8*(3), 376.

Abidin, M., & Retnawati, H. (2019). A Diagnosis of Difficulties Answering Questions of Circle Material on Junior High School Students. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *23*(2), 144–155.

AM, M. A., & Hadi, S. (2021). Implementation of Item Response Theory at Final Exam Test in Physics Learning: Rasch Model Study: *Proceedings of the 6th International Seminar on Science Education (ISSE 2020)*, *541*, 336–342.

AM, M. A., & Istiyono, E. (2022). Multiple Representation Ability of High School Students in Physics: A Study of Modern Response Theory. *Thabiea : Journal of Natural Science Teaching*, *5*(1), 85–97.

Amin̈, S., Utaya, S., Bachri̇, S., & Sumarmi̇, S. (2020). Effect of Problem Based Learning on Critical Thinking Skill and Environmental Attitude. *Journal for the Education of Gifted Young Scientists*, *8*(2), 743–755.

Arisoy, B., & Aybek, B. (2021). The Effects of Subject-Based Critical Thinking Education in Mathematics on Students' Critical Thinking Skills and Virtues. *Eurasian Journal of Educational Research*, *92*, 99–120.

Asriadi, M., & Hadi, S. (2021). Analysis of the Quality of the Formative Test Items for Physics Learning Using the Rasch Model in the 21st Century Learning. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, *6*(2), 158.

Asriadi, M., & Istiyono, E. (2020). Exploration of Creative Thinking Skills of Students in Physics Learning. *Journal of Educational Science and Technology (EST)*, *6*(2), 151–158.

Baker, G. S. (2020). *CRITICAL THINKING: A profound handbook to improve your memory and logical skills, level up your decision making and problem solving*. Independently Publisher.

Bashooir, K., & Supahar, S. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM. *Jurnal Penelitian dan Evaluasi Pendidikan*, *22*(2), 219–230.

Burkholder, E., Wang, K., & Wieman, C. (2021). Validated diagnostic test for introductory physics course placement. *Physical Review Physics Education Research*, *17*(010127), 1–11.

Chang, H., Chen, J., Guo, C., Chen, C., Chang, C., Lin, S., Su, W., Lain, K., Hsu, S., Lin, J., Chen, C., Cheng, Y., Wang, L., & Tseng, Y. (2007). Investigating Primary and Secondary Students' Learning of Physics Concepts in Taiwan. *International Journal of Science Education*, *29*(4), 465–482.

CLPNA Self-Study Course. (2017). *Diagnostic Tests and Laboratory Values*. College of Licensed Practical Nurse of Alberta.

de Glopper, K. (2002). Fisher, Alec and Scriven, Michael (1997). Critical Thinking. Its Definition and Assessment. *Argumentation*, *16*(2), 247–251.

Depdiknas RI. (2007). Tes Diagnostik. *Direktorat Jenderal Manajemen Pendidikan Dasar dan Menengah, Direktorat Pembinaan Sekolah Menengah Pertama*, 24.

Desjardins, C. D., & Bulut, O. (2018). *Handbook of Educational Measurement and Psychometrics Using R*. CRC Press Taylor & Francis Group.

di Uccio, U. S., Colantonio, A., Galano, S., Marzoli, I., Trani, F., & Testa, I. (2019). Design and validation of a two-tier questionnaire on basic aspects of quantum mechanics. *Physical Review Physics Education Research*, *15*(010137), 1–25.

Djatmiko, I. W., Hadi, S., & Am, Muh. A. (2022). *The Impact of the Enrollment Zoning System on the Educational Quality of Vocational Secondary Schools:*

Duskri, M., Kumaidi, K., & Suryanto, S. (2014). Pengembangan Tes Diagnostik Kesulitan Belajar Matematika di SD. *Jurnal Penelitian dan Evaluasi Pendidikan*, *18*(1), 44–56.

Dwyer, C. P., & Walsh, A. (2020). An exploratory quantitative case study of critical thinking development through adult distance learning. *Educational Technology Research and Development*, *68*(1), 17–35.

Ennis, R. H. (2011). The Nature of Critical Thinking: An Outline of Critical Thinking Dispositions and Abilities. *Inquiry: Critical Thinking Across the Disciplines*, *26*(2), 4–4.

Faizah, U., Zuchdi, D., & Alsamiri, Y. (2019). An authentic assessment model to assess kindergarten students. *Reid (Research and Evaluation in Education)*, *5*(2), 103–119. Fajrianthi, F., Hendriani, W., & Septarini, B. G. (2016). Pengembangan Tes Berpikir Kritis Dengan Pendekatan Item Response Theory. *Jurnal Penelitian dan Evaluasi Pendidikan*, *20*(1), 45–55.

Fitriani, A., Zubaidah, S., Susilo, H., & Al Muhdhar, M. H. I. (2020). PBLPOE: A Learning Model to Enhance Students' Critical Thinking Skills and Scientific Attitudes. *International Journal of Instruction*, *13*(2), 89–106.

Gurel, D. K. (2015). A Review and Comparison of Diagnostic Instruments to Identify Students' Misconceptions in Science. *Eurasia Journal of Mathematics, Science & Technology Education, 11*(5), 989–1008.

Hadi, S., Haryanto, H., Am, Muh. A., Marlina, M., & Rahim, A. (2022). Developing Classroom Assessment Tool using Learning Management System-based Computerized Adaptive Test in Vocational High Schools. *Journal of Education Research and Evaluation*, *6*(1), 143–155.

Hadi, S., Ismara, K. I., & Tanumihardja, E. (2015). Pengembangan Sistem Tes Diagnostik Kesulitan Belajar Kompetensi Dasar Kejuruan Siswa SMK. *Jurnal Penelitian dan Evaluasi Pendidikan*, *19*(2), 168–175.

Himmatussolihah, Ashadi, & Elfi Susanti, V. H. (2020). Critical Thinking Skills of 10th Grade Students and the Effect on Learning Achievement. *Proceedings of the 3rd International Conference on Learning Innovation and Quality Education (ICLIQE 2019)*, *397*, 953–956.

Ketabi, S., Alavi, D. S. M., & Ravand, D. H. (2021). Diagnostic Test Construction: Insights from Cognitive Diagnostic Modeling. *International Journal of Language Testing*, *11*(1), 22–35.

Kriswantoro, K., Kartowagiran, B., & Rohaeti, E. (2021). A Critical Thinking Assessment Model Integrated with Science Process Skills on Chemistry for Senior High School. *European Journal of Educational Research*, *10*(1), 285–298.

Lawshe, C. H. (1975). A Quantitative Approach To Content Validity. *Personnel Psychology*, *28*(4), 563–575.

Lia, R. M., Rusilowati, A., & Isnaeni, W. (2020). NGSS-oriented chemistry test instruments: Validity and reliability analysis with the Rasch model. *Reid (Research and Evaluation in Education)*, *6*(1), 41–50.

Mardapi, D. (2017). *Penelitian, Pengukuran dan Evaluasi Pendidikan* (2nd ed.). Parama Publishing.

Mason, A. J., & Singh, C. (2016). Surveying college introductory physics students' attitudes and approaches to problem solving. *European Journal of Physics*, *37*(5), 055704.

Possin, K. (2014). Critique of the Watson-Glaser Critical Thinking Appraisal Test: The More You Know, the Lower Your Score. *Informal Logic*, *34*(4), 393.

Prihatni, Y., Kumaidi, K., & Mundilarto, M. (2016). Pengembangan Instrumen Diagnostik Kognitif Pada Mata Pelajaran IPA di SMP. *Jurnal Penelitian dan Evaluasi Pendidikan*, *20*(1), 111–125.

Retnawati, H. (2016). *ANALISIS KUANTITATIF INSTRUMEN PENELITIAN (Panduan Peneliti, Mahasiswa, dan Psikometrian)*. Parama Publishing.

Roohr, K. C., & Burkander, K. (2020). Exploring Critical Thinking as an Outcome for Students Enrolled in Community Colleges. *Community College Review*, *48*(3), 330–351.

Rusilowati, A. (2015). Pengembangan Tes Diagnostik Sebagai Alat Evaluasi Kesulitan Belajar Fisika. *Prosiding Seminar Nasional Fisika Dan Pendidikan Fisika (SNFPF) Ke-6 2015*, *6*(1), 1–10.

Shantika, E. G., & Istiyono, E. (2019). A Diagnosis of Students' Errors In Answering The Mathematics Test In Senior High School. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *23*(2), 129–143.

Siwi, E. J., Anindyarini, R., & Nahar, S. (2020). Item parameters of Yureka Education Center (YEC) English Proficiency Online Test (EPOT) instrument. *Research and Evaluation in Education*, *6*(1), 51–65.

Sustekova, E., Kubiatko, M., & Usak, M. (2019). Validation of Critical Thinking Test on Slovak Conditions. *EURASIA Journal of Mathematics, Science and Technology Education*, *15*(12), 1–10.

Syahfitri, J., Firman, H., Redjeki, S., & Sriyati, S. (2019). Development and Validation of Critical Thinking Disposition Test in Biology. *International Journal of Instruction*, *12*(4), 381–392.

Tam, H. P. (2016). On the Identification of Students' Misconceptions in a Two-Tier Item. In M.-H. Chiu (Ed.), *Science Education Research and Practices in Taiwan* (pp. 147–162). Springer Singapore.

Triumiana, D. A., & Sudarsono, F. (2017). Pengembangan Instrumen Tes Diagnostik Mata Pelajaran Fisika Kelas XI SMA Negeri 2 Bantul Semester Genap 2017/2018. *WIDYA DHARMA Jurnal Penelitian dan Evaluasi Pendidikan*, *5*(2), 10.

Türkoguz, S. (2020). Investigation of Three-Tier Diagnostic and Multiple Choice Tests on Chemistry Concepts with Response Change Behaviour. *International Education Studies*, *13*(9), 10–22.

Warsah, I., Morganna, R., Uyun, M., Hamengkubuwono, H., & Afandi, M. (2021). The Impact of Collaborative Learning on Learners' Critical Thinking Skills. *International Journal of Instruction*, *14*(2), 443–460.

Wasis, W. (2013). Model Penskoran Partial Credit Pada Butir Multiple True-False Bidang Fisika. *Jurnal Penelitian dan Evaluasi Pendidikan*, *15*(1), 1–21.

Wijaya, M. H., & Hp, A. (2013). Pengembangan Tes Diagnostik Mata Pelajaran IPA SMP. *Jurnal Penelitian dan Evaluasi Pendidikan*, *17*(1), 19–36.

Xiao, Y., Han, J., Koenig, K., Xiong, J., & Bao, L. (2018). Multilevel Rasch modeling of two-tier multiple choice test: A case study using Lawson's classroom test of scientific reasoning. *Physical Review Physics Education Research*, *14*(2), 020104.

Zaleha, Z., Samsudin, A., & Nugraha, M. G. (2017). Pengembangan Instrumen Tes Diagnostik VCCI Bentuk Four-Tier Test pada Konsep Getaran. *Jurnal Pendidikan Fisika dan Keilmuan (JPFK)*, *3*(1), 36.