



Kasuari: Physics Education Journal (KPEJ) Universitas Papua

Web: <http://jurnal.unipa.ac.id/index.php/kpej>



Development and Validation of Critical Thinking Skills Instruments on Mechanical Waves for Senior High School Student

Fithrotul Azizah^{1*}, Supriyono Koes Handayanto¹, Hari Wisodo¹, Trio Junira Fernando²,
Kuswanto³, Anisak Intan Eka Prani⁴

¹Departemen Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Malang

²SMAN 3 Malang

³SMAN 1 Turen

⁴MAN Kota Batu

*Corresponding author: fithrotul.azizah.2203218@students.um.ac.id

Abstract: The importance of developing Critical Thinking Skills (CTS) is claimed as a major education purpose. Recently, enhanced learning has been conducted to promote students' CTS through study on certain topics (specific domains). Based on the numerous studies on CTS, there is a need for an assessment instrument to test students' CTS. This research aims to develop an instrument of CTC on mechanical wave topics for physics class in senior high school students. The instrument consists of seven essay items and one forced-choice item, tested to 45 students from XI physics class of high school in Malang. The validity test obtained an excellent Cronbach alpha coefficient of $\alpha = 0.717$. There was also an inter-rater test for the scoring guideline reliability using the Intraclass Correlation Coefficient (ICC) with high reliability results ($\kappa = 0.909$). The results of this study show that the items in instrument is qualified to be used to measure students' CTS on mechanical wave topics.

Keywords: Assessment in physics, critical thinking skills, mechanical wave

Pengembangan dan Validasi Instrumen Soal Kemampuan Berpikir Kritis Materi Gelombang Mekanik untuk Siswa Sekolah Menengah Atas

Abstrak: Pengembangan kemampuan berpikir kritis diklaim sebagai salah satu tujuan utama pendidikan. Saat ini telah dilakukan penelitian terkait peningkatan pembelajaran dalam meningkatkan kemampuan berpikir kritis melalui pembelajaran pada materi tertentu atau dalam domain spesifik. Sejalan dengan banyaknya usaha untuk mengasah kemampuan berpikir kritis, maka dibutuhkan instrumen penilaian untuk menguji kemampuan berpikir kritis siswa. Penelitian ini memiliki tujuan mengembangkan instrumen soal kemampuan berpikir kritis materi gelombang mekanik untuk siswa sekolah menengah atas. Instrumen soal terdiri dari tujuh soal uraian dan satu soal forced-choice dimana soal tersebut diujikan kepada 45 siswa kelas XI Fisika SMA di Malang. Berdasarkan uji validitas yang telah dilakukan didapat koefisien cronbach alpha yang dapat diterima yaitu $\alpha = 0,717$. Dilakukan pula uji antar-rater untuk reliabilitas pedoman penilaian menggunakan *Intraclass Correlation Coefficient* (ICC) dengan hasil reliabilitas tinggi ($\kappa = 0,909$). Hasil penelitian ini menunjukkan bahwa instrumen soal yang dibuat layak digunakan dalam proses penilaian kemampuan berpikir kritis siswa pada materi gelombang mekanik.

Kata kunci: Asesmen fisika, gelombang mekanik, kemampuan berpikir kritis

INTRODUCTION

The development of Critical Thinking Skill (CTS) is claimed to be one of the main goals of science education (Putra et al., 2023; Sermeus et al., 2021; Viennot & Décamp, 2018). CTS involves students having a high level of curiosity about a problem so that they

try to find information to get the right understanding (Viennot & Décamp, 2018). CTS assists students in thinking logically, making valid conclusions and solving problems during the learning process (Sermeus et al., 2021; Tiruneh et al., 2017). This ability is also associated with one's success in learning and enhanced decision-making proficiency in everyday life problems (Butler et al., 2017). In the present research, CTS refer to the description described by Halpern (2014), i.e.: reasoning, argument analysis, hypothesis testing, probability and uncertainty analysis, and decision making and problem solving.

Several studies in Indonesia show that students' CTS are still in the low category (Benyamin et al., 2021; Fitriani et al., 2022; Priyadi et al., 2018; Saphira & Prahani, 2022; Susilawati et al., 2020). Low CTS are caused by several problems, namely: (1) teachers still use lecture method learning (Saphira & Prahani, 2022); (2) students had challenges to interpret research outcomes with the relevant theories (Fitriani et al., 2022); (3) students are unable to identify data from experiments or problems (Priyadi et al., 2018). In addition, Musyarrof et al., (2018) specifically analysed the weak CTS of high school students due to the low ability of students in the aspects of analysing problems, evaluating, and making decisions.

There are several studies that have been conducted to improve students' CTS through learning models. Wartono et al., (2017) used an inquiry-discovery learning model to improve CTS in high school grade X students. Ferty et al., (2019) improved students' CTS through scaffolding-integrated PhET simulation technology. Koes-H et al., (2020) used model of flipped classroom class in inquiry learning to improve CTS in impulse and momentum material. Setiawan & Islami, (2020) applied problem-based learning for CTS of high school students. Putra et al., (2023) examined students' CTS through engineering design process in physics class.

Consistent with the numerous efforts to improve CTS, there is a need for develop instruments to test students' CTS (Lin, 2014; Negoro et al., 2020). There have been existing assessment instruments to measure CTS in the general domain such as the Cornell Critical Thinking Test (CCTT) (Ennis et al., 1989). and the Halpern Critical Thinking Assessment (HCTA) (Halpern, 2010). However, there is a need for critical thinking assessment instruments in specific domains. This is because there are indications that students' CTS depend on knowledge of specific material content, so a deep understanding of a specific domain is needed in order to be competent in completing thinking tasks (Davies, 2013). This is in line with Tiruneh research (Tiruneh et al., 2014), suggested that embedding critical thinking instructions in specific learning materials is considered the right way to help students become more capable of critical thinking. Therefore, an accurate and comprehensive critical thinking skill assessment instrument is needed to cover a certain material in a specific domain.

Furthermore, there are several researches that developed CTS instrument for specific domain of physics. The main research by Tiruneh et al., (2017) improve Critical Thinking in Electricity and Magnetism (CTEM) for undergraduate students. The CTEM instrument was then adapted by Perdana et al., (2019) on Critical Thinking on Kinetic Theory of Gases (CTKTG) material for high school students, Negoro et al (2020) on Momentum Critical Thinking (MCT) for pre-service teachers, Rusilowati et al., (2023) on Wave Critical Thinking (WCT) for high school students. Then research by Marisda et al., (2022) which developed an instrument for CTS in electricity and magnetism for pre-service teachers.

From literature studies and observational studies, it became clear that there is no question instrument that can specifically measure the CTS of high school students in mechanical wave material. Literacy studies show that students are expected to be capable of performing domain-specific CTS on physics content (Tiruneh, 2016). A specific test

instrument is needed as an assessment for CTS in mechanical wave topic. In this study, the development of critical thinking instruments through the 4D model (Define, Design, Develop and Disseminate) (Thiagarajan, 1974). The instrument developed is an essay question that is not only narrative but also applicable to daily life, and the researcher also developed a forced choice type question which includes several aspects of critical thinking at once. In this study, the validity test of the test instrument was also conducted, which consisted of the Cronbach alpha consistency test, the inter-rater reliability test using the Intraclass Correlation Coefficient (ICC) and the test of the difficulty level and the test of the difference of the questions. After conducting item design with the 4D model and testing the validity of the items, the further goal of the item development in this study is to be used as an instrument to test students' critical thinking skills in broader terms.

METHOD

The method of developing CTS on mechanical wave material apply the 4D model which contain four steps Define, Design, Develop and Disseminate (Thiagarajan, 1974). First define which defines the CTS that will be included in the questions. The definition stage or analysis of the needs of this question was carried out by studying literacy and previous research (Hariyanto et al., 2022). The second stage is design where the researcher dissects the criteria for critical thinking items in CTEM questions attached to Tiruneh (2016). In this stage also determined the type of questions to be developed, where researchers chose the type of questions with essay and forced-choice formats.

The third stage is developed, where the development of CTS in mechanical wave material is carried out according to the CTEM grid. Each item that has been made is then reviewed by expert lecturers in the field of physics. The last stage is disseminated, where the instruments of questions that have been reviewed and improved in the research stage are applied directly to the students (Thiagarajan, 1974). The question instrument was tested on class XI students at SMAN 3 Malang and SMAN 1 Turen who had received mechanical wave material. Finally, after collecting the empirical data, the quality of the questions was tested using quantitative descriptive methods related to validity, inter-rater, reliability, difficulty level and differentiation.

RESULTS AND DISCUSSION

Result of this research is an instrument of CTS on mechanical wave topic. The instrument in this study purpose to define the CTS on domain specific of high school students, which enables the students to use it as a reference to assess the ability to analyse and conceptual understanding of the students.

Define

From the literature study and the results of field observations, the problem is that there is no question instrument that can specifically measure the CTS in mechanical wave material. The literacy study shows that students are proposed to capable of improving CTS related to specific domains in physics material (Tiruneh, 2016). It is needed to expand a standardised instrument items as a tool for evaluating students' CTS in a specific domain, specifically in mechanical wave topic.

Design

In designing a CTS grid based on Tiruneh (2016), questions were selected that were possible to be answered by high school students. Where each item has more than one general domain of CTS which is then developed in a specific domain. The question

instrument developed aims to discover the critical thinking of students and how they analyse phenomena related to mechanical waves and understand related mathematical equations. The indicators of critical thinking ability of mechanical wave material are explain in Table 1.

Table 1. Critical thinking indicator content of the mechanical wave

Critical Thinking Indicator	Specific Domains in Science	Specific Domain in Mechanical Wave
Reasoning	Evaluating the validity of data	Explain reasons based on related wave mechanic theory
	Interpreting experiment results	Explain the results of Melde's experiment
	Identify misuse and ambiguities in usage of definitions	Answer with indicating the errors in the data and the causes of the errors
Hypothesis testing	Interpretating the link between variables	Show the relation among variables in mechanical wave topic
Argument analysis	Identifying the main parts of an argument	Describe the identification of arguments that concern the application of waves to real-life cases
	Assessing the credibility of a information resource	Identify the key parts of a graph to draw conclusions from the relevant information
Likelihood and uncertainty analysis	Predicting the likelihood of certain conditions	Predict the expected probability of the resulting event when a variable value is included
	Utilizing consideration of possibilities to make decisions	Make decisions on scenarios when a variable is changed by considering other factors that will be affected
	Calculating the probability of various situations expected to occur with known likelihoods	Recognize the probability of an event occurring given the possibilities shown

Develop

In this phase, the development of CTS questions on mechanical wave material was conducted according to the CTEM content of Tiruneh (2016). The questions were adjusted to the ability of high school students, where researchers developed the ability to think about mechanical wave material consisting of seven essay questions and one forced choice question. Each item that has been made is then reviewed by expert lecturers in the physics department. The review of questions is according to the subsequent guidelines: (a) the feasibility of each item to measure CTS in a specific domain and the suitability of the items with the test participants (b) the accuracy of the information presented in each item (c) clarity of words, sentences, and diagrams on each item (Tiruneh et al., 2017). In this research, the items developed by researchers have received approval from expert lecturers, whereas during the process of developing the questions, expert lecturers have provided

feedback and guidance. First, the expert lecturer considered the CTEM question framework and adjusted it to examine the accuracy of the CT that appeared in the mechanical wave question. Then, expert lecturers review the content of physics, diagrams and sentences used in CT questions for mechanical waves.

Then the researcher developed an assessment guide which was also reviewed by expert lecturers. To create a scoring guide, the desired ideal answer was first reviewed, then the researcher created a series of possible student responses to determine the scoring scale. Where the scoring scale varies depending on the time it takes to answer an item. Figure 1 shows an example of a critical thinking skill question for mechanical waves of the essay type and Figure 2 for the forced choice type.

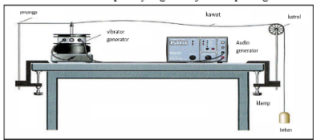
Disseminate

Following the revision of the questions and final validation by expert lecturers, an empirical test was conducted involving 11th grade high school students in Malang who had completed the mechanical wave material. Students who took this test did not receive special learning treatment for critical thinking. Before working on the questions, students were given verbal instructions regarding general directions for taking the test and an appeal to do the test well and serious. The test was given in a conducive classroom atmosphere and students were given 90 minutes to finish the test. About 70% of students were able to complete it after 70 minutes and the rest completed it after 90 minutes.

Pada percobaan melde, Naila menggunakan seutas senar yang terikat di kedua ujungnya memiliki panjang 120 cm diukur dengan sebuah penggaris dengan ketelitian 0,5 mm. Kemudian senar tersebut diberi tegangan sebesar 96,0 N dan diberikan getaran hingga beresonansi. Sumber getaran (*audio generator*) menggunakan frekuensi sebesar 100 Hz. Ketika senar beresonansi, Naila mengukur jarak antara satu simpul kepada satu simpul terdekat bernilai 25 cm. Dapatkah Anda menyimpulkan bahwa saat senar beresonansi dengan keadaan seperti diatas, terukur nilai kecepatan gelombang pada senar tepat bernilai 50 m/s ?

Figure 1. Example of CTS Essay on Mechanical Wave (Item Number 5)

Rangkaian percobaan Melde terdiri dari tiga komponen utama yaitu audio generator, kawat tembaga dan beban bermassa tertentu seperti yang ditunjukkan pada gambar di bawah.



Pada percobaan pertama digunakan sumber frekuensi suara 50Hz yang berasal dari audio generator (dengan sumber listrik AC). Kemudian pada percobaan pertama digunakan beban dengan massa 50 gram dan kawat tembaga dengan diameter 0,4 mm. Dari set data pertama dihasilkan setengah gelombang pada kawat dengan panjang 35 cm.

Pilihlah dari pilihan a-f bagaimana setiap perubahan akan mempengaruhi apa yang terjadi pada hasil percobaan melde!

Setiap perubahan di bawah ini mengacu pada percobaan pertama yang dilakukan	Pengaruh yang dihasilkan pada percobaan selanjutnya
(1) Massa beban dilipatgandakan (massa beban yang baru adalah 100 gram)	(a) (b) (c) (d) (e) (f)
(2) Frekuensi sumber suara ditambahkan sehingga frekuensinya menjadi 50 Hz	(a) (b) (c) (d) (e) (f)
(3) Digunakan kawat tembaga dengan diameter lebih lebar, yaitu 0,6 mm	(a) (b) (c) (d) (e) (f)
(4) Digunakan sumber listrik DC (digunakan aki) pada audio generator	(a) (b) (c) (d) (e) (f)

Keterangan (dengan perubahan tersebut maka yang terjadi adalah) :

- Kecepatan gelombang pada tali bertambah
- Kecepatan gelombang pada tali berkurang
- Nilai panjang gelombang pada tali tetap
- Terbentuk lebih banyak perut dan simpul
- Tidak terjadi perubahan apapun (hasil percobaan sama dengan sebelumnya)
- Tidak muncul getaran pada sistem

Figure 2. Examples of CTS Forced Choice on Mechanical Wave (Item Number 3)

Instrument Validation and Reliability

Data was collected after students worked on critical thinking questions of mechanical wave material to conduct statistical analyses. The first statistical test carried out was to

measure the internal consistency of the questions using the Cronbach alpha test. 45 sets of scores of participants who took the test were taken to be processed and an acceptable Cronbach alpha coefficient was obtained, namely $\alpha = 0.717$. Where the Cronbach alpha value between 0.7 and 0.8 is considered acceptable, so the questions tested are valid and reliable (Cohen et al., 2017). If the items are confirmed valid then the items are qualified to be used to measure the intended skills and can be proceeded for further statistical tests.

Furthermore, the scoring guidelines for CTS questions used in this study were evaluated using inter-rater reliability. The researcher randomly selected 15 student test results to be assessed by three raters separately using the provided scoring guidelines. The statistical test used to measure inter-rater reliability is the ICC calculated with 95% confidence interval (IC). The ICC for this study was based on a two-way mixed-effect model because there were specifically selected raters (raters of interest). Then the type of ICC used is average measures because there are three raters to evaluate the reliability of the question (Koo & Li, 2016). Table 1 shows the range of values used for the interpretation of ICC test coefficients (Fleiss, 1999). In this study, the ICC test was analyzed using the SPSS software and then the coefficient of the ICC test results will be examined using the parameters presented in the Table 2.

Table 2. Interpretation of ICC Test Results (Fleiss, 1999)

ICC (κ)	Interpretation
0 – 0,39	Low reliability
0,4 – 0,74	Medium reliability
0,75 – 1	High reliability

of the ICC test on each item showed that the level of agreement between raters reached a coefficient range of 0.777 to 0.929. The details of the ICC test are summarised in table 3, where for question number 5 all answers assessed by the raters were given a value of '0' so that the ICC test could not be carried out. For the total score on the question, the inter-rater agreement rate was 0.909. The ICC Test results showed moderate to high inter-rater reliability for each item and for the total score. This proves the objectivity of the raters in the assessment process. The reliability test results also show that the answer scoring guidelines used by the raters can compare well the results of student performance on each item and for the whole question.

Table 3. Summary of Statistical Test Results for CTS Questions on Mechanical Waves

Items	ICC (κ)	Level of Difficulty	Level of Discrimination
No 1	0,817	0,11	0,25
No 2	0,929	0,43	0,97
No 3	*	0,52	0,96
No 4	0,849	0,58	0,75
No 5	-	0,12	0,25
No 6	0,808	0,36	0,75
No 7	0,915	0,35	0,79
No 8	0,777	0,38	0,44

* question number 3 is a question with a forced-choice format

- the raters gave a score of '0' for all answers to question 5

Item 3 is a forced-choice question (see Figure 2) therefore the scoring is based on binomial assessment with correct score 1 and incorrect score 0. Furthermore, forced choice used in item number 3 was used in this study to reduce bias in the answers given by students (Shavelson et al., 2019). In the case of item 5 (see Figure 1), this is a narrative question wherein all raters gave a score of zero because students answered with straightforward mathematical calculations without considering the presence of more or less probabilities that must occur in the context of a real experiment. In this case for question number 5, students are expected to calculate various probabilities that are expected to occur in situations with known probabilities (namely in the form of an error value on the tools used in the experiment).

The test of the level of difficulty and level of discrimination of each item was also carried out as an additional test to assess the characteristics of CTS questions on mechanical waves. Level of difficulty is the percentage of correct answers from students that indicate a certain index of difficulty. According to (Bai et al., 2017) the level of difficulty is the ratio between the number of students who answered correctly and incorrectly. However, presenting the level of difficulty is not enough, it must be continued with the level of discrimination (Quaigrain & Arhin, 2017). Level of discrimination provides information about the differences in each student's CTS ability based on the test.

Level of difficulty test in this study is carried out to obtain the proportion of correct answer scores on a question that has been done by test takers. For example, question number 1 has a range of 0 to 4, for number 2 the score is from 0 to 6 and for number 7 the score is from 0 to 2. Because the questions used in this test are included in the type of open-ended questions, the researchers used the formula from the Evaluation and Examination Service of the University of Iowa to test the level of difficulty and level of discrimination of the items. The equation for calculating level of difficulty (P):

$$P = \frac{\sim fX - nX_{min}}{n(X_{max} - X_{min})}$$

where $\sim fX$ is the total points of all student scores on a number, n is the number of students, X_{min} is the smallest score on the question, and X_{max} is the largest score on the question. In the level of difficulty test using 45 student test results, where a range level of difficulty was obtained from 0.11 to 0.58. In detail, there are no items with low level of difficulty, there are six questions with medium level of difficulty and two questions with high level of difficulty.

Furthermore, for the level of discrimination test stage of the items, first grouping test participants with high (upper) and low (lower) scores is carried out. The grouping is done to get an idea of how well a question can distinguish individual abilities at various levels. In general, the grouping of high and low scores is to take 27% of the highest and lowest scores of all participants' scores, but in this study the number of test participants' scores processed was limited ($N = 45$), so a presentation of 22% of the highest and lowest scores was used to test the level of discrimination. Therefore, researcher took ten student scores for the upper category and ten student scores for the lower category then analysed the results to determine the level of discrimination (D). The equation for calculating level of discrimination (D):

$$\text{Level of discrimination (D)} = P_U - P_L$$

where P_U is the difficulty level for the upper group and P_L is the difficulty level for the lower group. As shown in Table 2, level of discrimination result in the range of 0.25 to 0.97. In detail, there are two questions with the interpretation of sufficient differential power, one question with the interpretation of good differential power and five questions with the interpretation of excellent differential power. To clarify the role of sub-scores in

an item to show the differentiating power, here is one illustration of the distribution of the results of number six (score range 0 to 4) for the high score group (upper) and the low score group (lower).

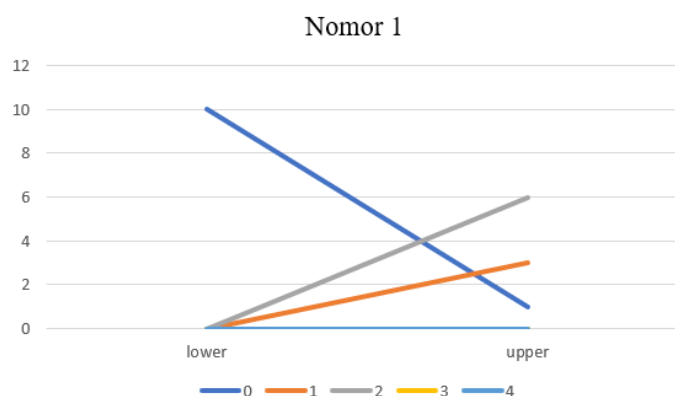


Figure 3. Distribution of Scores of Upper and Lower Score Groups for Item 1

In order to explain the low level of discrimination, a graph is made on question item number 1 as illustrated in the figure 3. In that graph the higher sub-scores (scores 1 and 2) increased from the lower to the upper group, the lower sub-scores (score 0) decreased from the lower to the upper group. Similar results were observed for the other seven items, even for those with low power level of discrimination (Tiruneh et al., 2017). The analysis of the upper and lower score groups supports the level of discrimination of the items. Furthermore, in research Schmidt & Embretson (2003), stated that the low differential power is due to the item difficulty index, where items that are either very easy or very difficult will lack differential power. In this case, questions item number 1 and number 5 are items that have a high level of difficulty, causing a low differentiating power.

Procedure in the development and validation of CTS question items for mechanical waves in this study is basically in line with the guidelines used in previous research on the CTS question development (Negoro et al., 2020; Perdana et al., 2019; Rusilowati et al., 2023). Even though the procedure of item development and validation has existed in previous studies, this study has been able to present an assessment framework that focuses on mechanical wave material that can measure CTS in four indicators, including reasoning, hypothesis testing, argument analysis and likelihood and uncertainty analysis (Halpern, 2010). Despite the fact that the questions made for this mechanical wave material are relatively difficult, the question instruments made have been able to accommodate the CTS indicators properly for the specific domain.

CONCLUSIONS

The instrument of critical thinking skill questions and assessment guidelines for mechanical wave material is declared feasible based on the Cronbach alpha coefficient reliability test $\alpha = 0.717$ so that it is included in the reliable category. The instrument of guideline for scoring was also declared quite reliable based on ICC analysis with an inter-rater agreement rate of 0.909. Based on these quantitative data, it is evident that the critical thinking test instrument for mechanical wave material produces good inter-rater agreement and a reliability coefficient with a significant value. This study has produced an assessment framework in the form of eight items of CTS that can be used for mechanical wave material at the high school level. Suggestions for further research are the need for continued research for the application of critical thinking skills questions on mechanical wave material that have been developed in this study. The further research might be the learning

outcomes linked to students' critical thinking skills in a specific domain of mechanical wave topics.

REFERENCES

- Bai, Xue, & Ola, A. (2017). A Tool for Performing Item Analysis to Enhance Teaching and Learning Experiences. *Issues in Information Systems*, 18(1), 128–136.
- Benyamin, B., Qohar, Abd., & Sulandra, I. M. (2021). Analisis Kemampuan Berpikir Kritis Siswa SMA Kelas X dalam Memecahkan Masalah SPLTV. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, 5(2), 909–922. <https://doi.org/10.31004/cendekia.v5i2.574>.
- Butler, H. A., Pentoney, C., & Bong, M. P. (2017). Predicting Real-World Outcomes: Critical Thinking Ability is a Better Predictor of Life Decisions than Intelligence. *Thinking Skills and Creativity*, 25(1), 38–46. <https://doi.org/10.1016/j.tsc.2017.06.005>.
- Cohen, L., Manion, L., & Morrison, K. (2017). *Research Methods in Education*. Routledge. <https://doi.org/10.4324/9781315456539>.
- Davies, M. (2013). Critical Thinking and the Disciplines Reconsidered. *Higher Education Research & Development*, 32(4), 529–544. <https://doi.org/10.1080/07294360.2012.697878>.
- Ennis, R. H., Millman, J., & Tomko, T. N. (1989). *Cornell Critical Thinking Tests Level X & Level Z* (3rd ed.). Midwest Publications.
- Ferty, Z. N., Wilujeng, I., Jumadi, & Kuswanto, H. (2019). Enhancing Students' Critical Thinking Skills through Physics Education Technology Simulation Assisted of Scaffolding Approach. *International Seminar on Science Education (ISSE)*, Yogyakarta: 13 October 2018.
- Fitriani, A., Zubaidah, S., & Hidayati, N. (2022). The Quality of Student Critical Thinking: A Survey of High Schools in Bengkulu, Indonesia. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 8(2), 142–149. <https://doi.org/10.22219/jpbi.v8i2.18129>.
- Fleiss, J. L. (1999). *The Design and Analysis of Clinical Experiments*. New Jersey, U.S: Wiley. <https://doi.org/10.1002/9781118032923>
- Halpern, D. F. (2010). *The Halpern Critical Thinking Assessment: Manual*. Modling, Austria: Schuhfried GmbH.
- Halpern, D. F. (2014). *Thought and Knowledge: An Introduction to Critical Thinking* (5th ed.). New York, NY: Psychology Press.
- Hariyanto, B., MZ, I., SU, W., & Rindawati. (2022). 4D Model Learning Device Development Method of the Physical Geography Field Work Guidance Book. *Matec Web of Conferences*, 372, 05008. <https://doi.org/10.1051/mateconf/202237205008>
- Marisda, H D., Handayani, Y., Dhiqfaini Sultan, A., Hasin, A., & Nur Samsi, A. (2022). Development Assessment of Thinking Skills Interpretation, Analysis, and Inference Prospective Physics Teacher. *Physics Education Journal*, 5(1), 33–41. <http://jurnal.unipa.ac.id/index.php/kpej>.
- Koes-H., S., Putri, F. S., Purwaningsih, E., & Salim, A. Y. (2020). The Influence of Flipped Classroom in Inquiry Learning to Student's Critical Thinking Skills in Impulse and Momentum. *The 3rd International Conference on Mathematics and Science Education (ICoMSE)*, Malang: 26–28 August 2019.
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>.

- Lin, S.-S. (2014). Science and Non-Science Undergraduate Students' Critical Thinking and Argumentation Performance in Reading a Science News Report. *International Journal of Science and Mathematics Education*, 12(5), 1023–1046. <https://doi.org/10.1007/s10763-013-9451-7>.
- Musyarrof, A. F., Nugroho, S. E., & Masturi, M. (2018). The Analysis of Students' Critical Thinking Weakness in Senior High School on Physics Learning. *Proceedings of the International Conference on Science and Education and Technology 2018 (ISET 2018)*. <https://doi.org/10.2991/iset-18.2018.8>.
- Negoro, R. A., Rusilowati, A., Aji, M. P., & Jaafar, R. (2020). Critical Thinking in Physics: Momentum Critical Thinking Test for Preservice Teacher. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, 9(1), 73–86. <https://doi.org/10.24042/jipfalbiruni.v9i1.4834>.
- Perdana, R., Riwayani, R., Jumadi, J., Rosana, D., & Soeharto, S. (2019). Specific Open-Ended Assessment: Assessing Students' Critical Thinking Skill on Kinetic Theory of Gases. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, 8(2), 127–140. <https://doi.org/10.24042/jipfalbiruni.v0i0.3952>.
- Priyadi, R., Mustajab, A., Tatsar, M. Z., & Kusairi, S. (2018). Analisis Kemampuan Berpikir Kritis Siswa SMA Kelas X MIPA dalam Pembelajaran Fisika. *JPFT (Jurnal Pendidikan Fisika Tadulako Online)*, 6(1), 53–55. <https://doi.org/10.22487/j25805924.2018.v6.i1.10020>.
- Putra, P. D. A., Sulaeman, N. F., Supeno, & Wahyuni, S. (2023). Exploring Students' Critical Thinking Skills Using the Engineering Design Process in a Physics Classroom. *The Asia-Pacific Education Researcher*, 32(1), 141–149. <https://doi.org/10.1007/s40299-021-00640-3>.
- Quaigrain, K., & Arhin, A. K. (2017). Using Reliability and Item Analysis to Evaluate a Teacher-Developed Test in Educational Measurement and Evaluation. *Cogent Education*. 4(1), 1–11. <https://doi.org/10.1080/2331186X.2017.1301013>.
- Rusilowati, A., Negoro, R. A., Aji, M. P., & Subali, B. (2023). Development of Waves Critical Thinking Test: Physics Essay Test for High School Student. *European Journal of Educational Research*, 12(4), 1781–1794. <https://doi.org/10.12973/eu-jer.12.4.1781>.
- Saphira, H. V., & Prahani, B. K. (2022). Profile of Senior High School Students' Critical Thinking Skills and the Need of Implementation PBL Model Assisted by Augmented Reality Book. *Jurnal Pendidikan Sains Indonesia*, 10(3), 579–591.
- Schmidt, K. M. & Embretson, S. E. (2003). Item Response Theory and Measuring Abilities. In J. A. Schinka & I. B. Weiner (Eds.), *Handbook of Psychology: Research Methods in Psychology* (Vol. 2, pp. 429–444). Hoboken, NJ: Wiley.
- Sermeus, J., De Cock, M., & Elen, J. (2021). Critical Thinking in Electricity and Magnetism: Assessing and Stimulating Secondary School Students. *International Journal of Science Education*, 43(16), 2597–2617. <https://doi.org/10.1080/09500693.2021.1979682>.
- Setiawan, H. J., & Islami, N. (2020). Improving Critical Thinking Skills of Senior High School Students Using the Problem Based Learning Model. *Universitas Riau International Conference on Science and Environment (URICSE)*, Pekanbaru: 11-13 September 2020.
- Shavelson, R. J., Zlatkin-Troitschanskaia, O., Beck, K., Schmidt, S., & Marino, J. P. (2019). Assessment of University Students' Critical Thinking: Next Generation Performance Assessment. *International Journal of Testing*, 19(4), 337–362. <https://doi.org/10.1080/15305058.2018.1543309>.

- Susilawati, E., Agustinasari, A., Samsudin, A., & Siahaan, P. (2020). Analisis Tingkat Keterampilan Berpikir Kritis Siswa SMA. *Jurnal Pendidikan Fisika Dan Teknologi*, 6(1), 11–16. <https://doi.org/10.29303/jpft.v6i1.1453>.
- Thiagarajan, S. (1974). *Instructional Development for Training Teachers of Exceptional Children: A Sourcebook*. University of Minnesota: the Center for Innovation in Teaching the Handycapped. <https://api.semanticscholar.org/CorpusID:148925881>.
- Tiruneh, D. T. (2016). Learning Environments for Critical Thinking: The Added Value of Instructional Design Research (Doctoral dissertation). Retrieved from Katholieke Universiteit Leuven Bibliography Database. (Accession No. LIRIAS1769627)
- Tiruneh, D. T., De Cock, M., Weldelessie, A. G., Elen, J., & Janssen, R. (2017). Measuring Critical Thinking in Physics: Development and Validation of a Critical Thinking Test in Electricity and Magnetism. *International Journal of Science and Mathematics Education*, 15(4), 663–682. <https://doi.org/10.1007/s10763-016-9723-0>.
- Tiruneh, D. T., Verburgh, A., & Elen, J. (2014). Effectiveness of Critical Thinking Instruction in Higher Education: A Systematic Review of Intervention Studies. *Higher Education Studies*, 4(1), 1–17. <https://doi.org/10.5539/hes.v4n1p1>.
- Tiruneh, D. T., Weldelessie, A. G., Kassa, A., Tefera, Z., De Cock, M., & Elen, J. (2016). Systematic Design of a Learning Environment for Domain-Specific and Domain-General Critical Thinking Skills. *Educational Technology Research and Development*, 64(3), 481–505. <https://doi.org/10.1007/s11423-015-9417-2>.
- Viennot, L., & Décamp, N. (2018). Activation of a Critical Attitude in Prospective Teachers: From Research Investigations to Guidelines for Teacher Education. *Physical Review Physics Education Research*, 14(1), 1–19. <https://doi.org/10.1103/PhysRevPhysEducRes.14.010133>.
- Wartono, W., Hudha, M. N., & Batlolona, J. R. (2017). How are the Physics Critical Thinking Skills of The Students Taught by Using Inquiry-Discovery through Empirical and Theoretical Overview. *Eurasia Journal of Mathematics, Science and Technology Education*, 14(2), 691–697. <https://doi.org/10.12973/ejmste/80632>.