Analysis of critical thinking test instruments on the Light Waves concept (CT-LiWa): Rasch model

Raldy Aditya[1], Endi Suhendi[1*], Lina Aviyanti[1], Abu Nawas[2], Heni Rusnayati[1]

[1]Universitas Pendidikan Indonesia, Dr. Setiabudi St. No.229, Bandung, West Java, 40154, Indonesia
[2]The University of Adelaide, Adelaide, South Australia, 5005, Australia
*Corresponding author, email: endis@upi.edu

Article History
Received: 25 February 2025
Revised: 20 April 2025
Accepted: 10 June 2025

Keywords
Critical thinking
Light waves
Rasch model

Abstract
This research aims to develop an instrument that measures students' critical thinking skills in physics learning to test its validity and reliability using the Rasch Model. The instrument is called the Critical Thinking Test on The Light Waves Concept (CT-LiWa). The novelty of this research is to create a new instrument that is applied to the concept of light waves. The target of this instrument is high school students in one of the provinces in West Java. This research method uses the ADDIE model involving 72 high school students from West Java, Indonesia. The instrument used is a validation sheet and the CT-LiWa consists of 15 items in the form of descriptions. The trial results data were analyzed using the Rasch Model assisted by the Winstep program to obtain the quality of the instrument, namely reliability, validity, and level of difficulty. The results of the analysis show that the CT-LiWa test is not reliable but has a good validity of 27.6% in the fulfilled category with varying levels of difficulty. Therefore, the development of the CT-LiWa test can be implemented as an instruent to measure students' critical thinking skills on the concept of light waves.

Aditya, R., Suhendi, E., Aviyanti, L., Nawas, A., & Rusnayati, H. (2025). Analysis of critical thinking test instruments on the Light Waves concept (CT-LiWa): Rasch model. *Momentum: Physics Education Journal*, *9*(2), 337-349. https://doi.org/10.21067/mpej.v9i2.11662

1. Introduction

Critical thinking is one of the skills that everyone in the 21st century should master (Saputra et al., 2019; Teo, 2019; Aulia, 2022). Teaching critical thinking is an important goal of modern education, as it equips students with the competency necessary to reason about social affairs in a rapidly changing world (Ku, 2009). Critical thinking involves the ability to draw valid inferences, identify relationships, analyze probabilities, make predictions and logical decisions, and solve complex problems (Ennis, 1993; Halpern, 2014; Tiruneh et al., 2017). To develop these competencies, students must do more than just memorise information from textbooks but also need to learn to build skills in assessing information, developing alternative evidence, and arguing with strong reasons. These skills in critical thinking are not only important for students to excel in school but are also needed in social and interpersonal contexts where the right decisions must be made carefully and independently. (Ku, 2009). Instilling critical thinking skills in subject learning will facilitate the acquisition of critical thinking skills that can be applied to thinking tasks and transfer these skills to other problems in everyday life, especially in physics subjects (Lawson, 2004; Tiruneh et al., 2017).

Critical thinking skills can be improved by practicing critical thinking indicators in the learning process. Combining models, media, and well-planned, implemented, and evaluated learning approaches will produce good learning strategies. A good learning strategy will create active learning. If the active learning process is combined with training indicators of critical thinking skills, continuous learning will be able to improve students' critical thinking skills. The relationship between models, media, and interactive learning approaches is illustrated in Figure 1 (Aditya et al., 2023).

doi: 10.21067/mpej.v9i2.11662

ISSN: 2548-9135

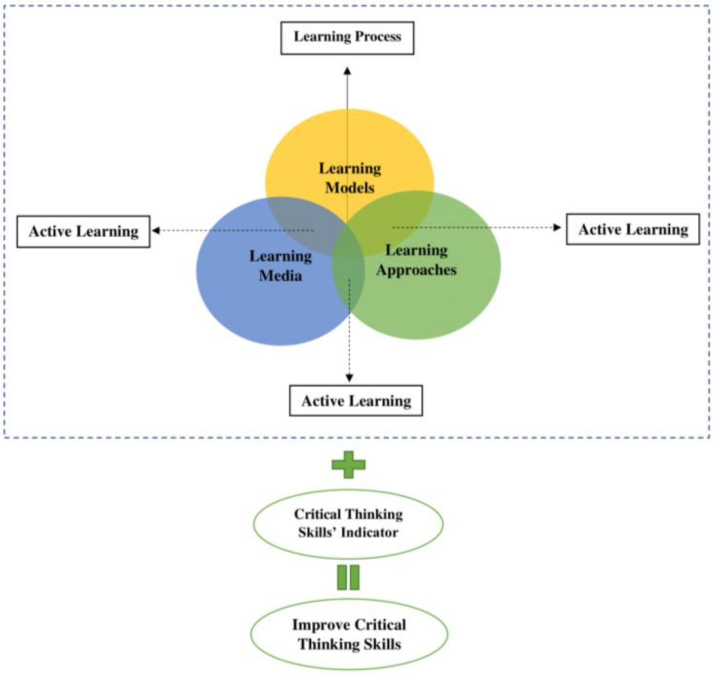


Figure 1. Framework of relationships between factors that improve critical thinking skills

Critical thinking skills must be continuously applied in the physics learning process, with the purpose of producing competent and skilled learners who can solve problems in everyday life (Kealey et al., 2005; Wartono et al., 2018). Based on research in one of the high schools in Sumatra province, shows that as many as 76.57% of students are in the low critical thinking category (Ariani, 2020). Students must have critical thinking skills in learning physics to understand complex and abstract physics concepts. With good critical thinking skills, students can understand and analyze complicated physics theories, thus developing a deeper understanding (Bhakti et al., 2023). Physics is ideally positioned to strengthen critical thinking skills, which can also be applied to other domains. Studying real-world physical phenomena through scientific practice trains disciplined thinking that relies on collecting empirical evidence, quantitative analysis, logical argumentation, and critical evaluation of claims (Recker, 2021; Jamil et al., 2024).

Nevertheless, most efforts to address the development of critical thinking are still within the context of general critical thinking skills or taught separately from the subject matter (Ennis, 1989). So efforts are needed to develop critical thinking skills by embedding them in design or learning materials. As the emphasis on developing critical thinking skills applied to subject matter increases, the need for assessment becomes increasingly important (Facione, 1990; Pascarella & Terenzini, 2005; Halpern, 2010; Lin, 2014; Tiruneh et al., 2017). Therefore, developing critical thinking skills assessments is necessary to measure students' critical thinking skills, especially in physics lessons.

An instrument is needed to assess critical thinking skills accurately based on indicators (Halimatun et al., 2020; Sumarni et al., 2018). The physics critical thinking skills assessment instrument is a tool used to measure students' critical thinking skills applied to physics subjects. This assessment is designed to evaluate students' skills in analyzing, evaluating, and synthesizing physics information, as well as students' abilities in identifying assumptions, connecting concepts, making decisions based on evidence, and solving problems critically (Bhakti et al., 2023). Assessment of critical thinking skills in physics usually consists of a series of questions or tasks specifically designed to test students' critical thinking skills (Gunawan et al., 2019; Putranta & Wilujeng, 2019). The questions in this instrument can be contextual questions that ask students to identify relevant physics concepts, explain physical phenomena clearly and logically, or apply physics concepts in everyday life situations. In addition, this instrument can also include contextual questions that link physics concepts to real-life contexts. This contextual question aims to test students' ability to apply physics knowledge in concrete situations, such as physics problems in everyday life, experiments, or the application of technology (Szabo et al., 2020; Zakaria et al., 2019).

There are various forms of assessment that can be used as options to measure students' critical thinking skills, namely multiple-choice measures of critical thinking (The WGCTA and the CCTST) (Watson & Glaser, 1980; Ennis et al., 1985; Facione, 1990a), Open-ended measures of critical thinking (Norris & Ennis, 1989; Taube, 1997; Halpern, 2003; Tiruneh et al., 2017), and New trend: critical thinking measure of multi-response format (Halpern, 2007). This study addresses the gap in the existing literature by adapting light wave material into a critical thinking test instrument that previously used the topic of electricity and magnetism. This study offers a new perspective by using the Rasch model in analyzing the validity of the instrument. Thus, this study is expected to provide a more comprehensive contribution in understanding the level of students' critical thinking skills in light wave material.

Measuring critical thinking skills in physics learning must be done using appropriate measurement instruments, considering the importance of measuring critical thinking skills in physics. Therefore, research is needed to assess critical thinking skills in the physics domain. This research uses an open form of assessment that aims to obtain students' critical thinking skills in a structured and comprehensive manner. This form of test is very structured and is able to test students' ability to identify inherent reasoning weaknesses in an argumentative passage, as well as students' ability to defend their own arguments (Ennis, 2003). The aim of this research is to develop and validate a test that measures the acquisition of critical thinking skills in physics. However, physics is a broad domain that deals with various subdomains at advanced educational levels, such as electricity and magnetism, waves, etc. Therefore, it is impractical to target content from all subdomains (physics) to test critical thinking skills as this could result in many items, which is difficult to do in a reasonable amount of time. Thus, researchers only focus on the content of one subdomain in introductory physics, namely light waves. Researchers deliberately chose light waves because this is one of the basic physics concepts regarding waves and the phenomena are often found in everyday life.

In summary, the conceptual framework of this research is illustrated in Figure 2. Therefore, this research aims to develop an instrument to measure students' critical thinking skills on light wave subjects, CT-LiWa, and test its validity and reliability. From the research objectives, the research questions are: (1) how is the development of CT-LiWa?; (2) what is the validity and reliability of IMPAS?; and (4) what are the recommendations for further research?

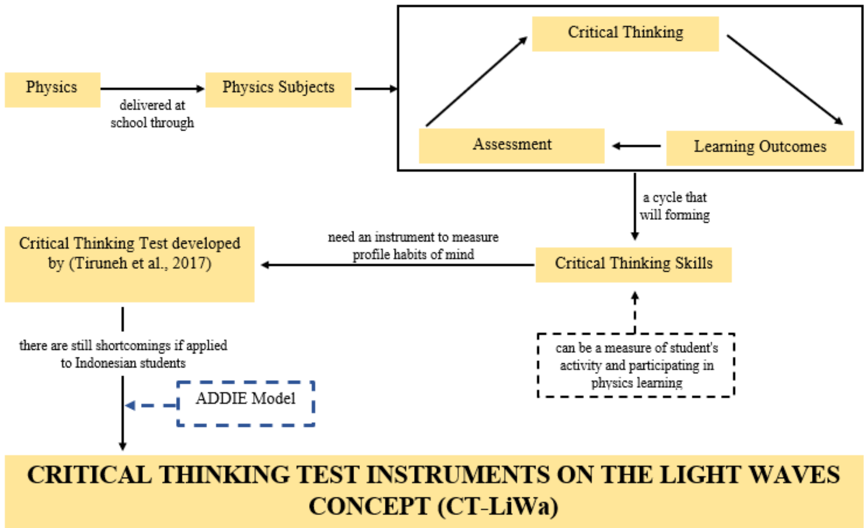


Figure 2. CT-LiWa Research Conceptual Framework

2. Method

2.1. Research Design

CT-LiWa is a developing assessment from the Critical Thinking (CT) Test to measure students' critical thinking skills in physics in Indonesia. CT-LiWa was developed using the ADDIE (Analyzing, Designing, Developing, Implementing, Evaluating) model. The applied research design is shown in Figure 3.

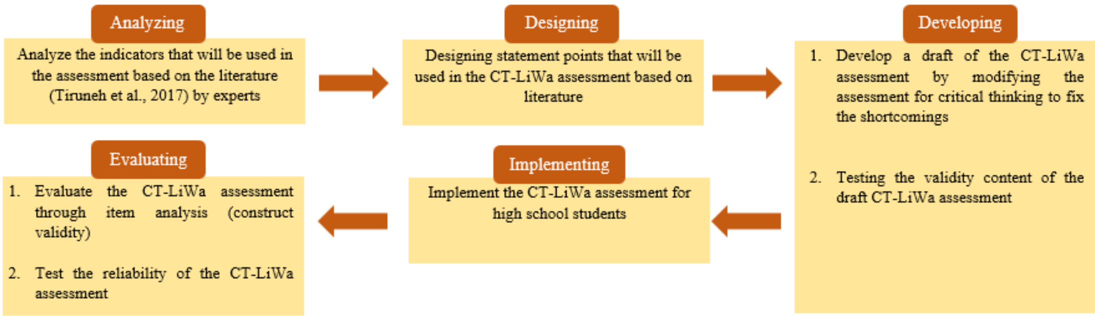


Figure 3. The CT-LiWa assessment Research and Development Design

2.2. Research Participant

This study involved three physics lecturers and two physics teachers to validate the instruments to be developed. This instrument was tested on 72 students aged 15-18 in one of the public high schools in Malang City using a simple random sampling, it is purposive sampling method. This amount is sufficient to achieve the research objectives. This is because to embed psychometrics in question construction, a minimum of 50 participants are involved (Sapnas & Zeller, 2002). In addition, with Rasch Model analysis, at a 95% confidence level, around 64-144 participants are needed (Linacre, 1994).

2.3. Research Instrument

The instruments used are validation sheets and CT-LiWa in essay form which have been designed and developed previously, as in Table 1. The validation sheet instruments are used at the development stage to determine the feasibility of the CT-LiWa instrument that has been designed. The validation instrument consists of 12 items for assessment aspects regarding content, construction and language validation. Based on the validation sheet, the designed CT-LiWa has been validated and declared suitable for testing.

2.4. Research Data Collecting and Analysis

The CT-LiWa validation questionnaire was addressed to three physics lecturers and two physics teachers as experts. After the assessment was developed, the validity of the CT-LiWa design was tested by experts to determine the feasibility of each statement item. The experts' answer data was converted into quantitative data based on Lawshe (1975). Validity is determined by calculating the Content Validity Ratios (CVR) values from expert assessments. The CVR value is calculated based on the Lawshe (1975) equation as in Equation 1.

$$CVR = \frac{n_e - N/2}{N/2} \quad (1)$$

Equation description:

N = The number of experts judging

n_e = The number of items that esential based on expert judgment

The trial results data were analyzed using the Rasch Model assisted by the Winstep program. From this analysis, the quality of the instrument is obtained, namely reliability, validity and level of difficulty. The reliability of the instrument with the Rasch Model can be shown by the summary statistics output. The results of the analysis show three types of reliability values, namely person reliability, item reliability and Cronbach's Alpha. To find out the interaction between person and item as a whole, it can be seen from Cronbach's Alpha (Adams et al., 2018; Fisher, 2007; Samsudin et al., 2021). The validity of the instrument can be shown by the output items, namely fit order and unidimensional. For output items, the values seen are the logit of the MnSq outfit, ZStd output, and

PT Measure Correlation (Arsad et al., 2013; Park & Liu, 2021; Samsudin et al., 2020). For unidimensional, the validity value is shown by raw variance explained by measures and unexplained variance. The level of difficulty of the instrument with the Rasch Model is shown by the item measure output. With these measure items, instrument items can be grouped based on level of difficulty. Interpretation of the results is presented (Sumintono & Widhiarso, 2014): instrument validity (Table 1), reliable person and item (Table 2), logit person, and logit item (Table 3).

Table 1. Interpretation of Instrument Validity

Interpretation	Raw variance explained by measures
Fulfilled	>20%
In accordance	>40%
Special	>60%

Table 2. Interpretation for Item Reliability and Person Reliability

Interpretation	Raw variance explained by measures
value ≥ 0.95	Excellent
0.95 > value ≥ 0.91	Very Good
0.91 > value ≥ 0.81	Good
0.81 > value ≥ 0.68	Moderate
0.68 > value	Weak

Table 3. Interpretation Logit Value

Item Logit Value (A)	Criteria	Person Logit Value (B)	Criteria
A ≥ +1.37	Very difficult	B ≥ 1.80	Height
0.0 ≥ A ≤ +1.37	Difficult	B ≤ 1.80	Currently
0.0 ≤ A ≥ -1.37	Currently	B ≤ -1.29	Low
A ≤ -1.37	Easy		

3. Results and Discussion

3.1. Analyzing

At this stage the researcher obtained a number of references related to the CT-Test, how to measure it, the CT-Test in physics education, and the Rasch Model as a theory for analyzing instruments. This reference is used as a basis for carrying out the next stages. From this stage it can be concluded that critical thinking is an important skill that individuals must have and can be trained in learning, including by using special assessments to measure critical thinking skills but is rarely done in schools due to the lack of special instruments. This stage is carried out by analyzing indicators related to students' critical thinking skills in physics based on experts. Based on research and expert statements, there are five indicators of critical thinking skills that are measured. Each indicator is taken from three specific domains and included in the CT-LiWa assessment developed by Dufresne. The CT-LiWa assessment has a structure as shown in Table 4.

Table 4. Assessment CT-LiWa Structure in measuring Critical Thinking Skills

Critical Thinking Skills Indicator	Specific Domains of Critical Thinking	Sub-Topic Light Wave	Score
Reasoning	detect ambiguity and misuse of definitions	double slit interference	2
	recognize errors of measurement	single slit diffraction	2
	interpret the results of an experiment	polarization of light	2
Hypothesis testing	draw valid inferences from a given tabular or graphical information	polarization of light	2
	interpret a relationship between variables	double slit interference	3
	recognize the need for more information in drawing conclusions	single slit diffraction	3
Argument analysis	judge the credibility of an information source	double slit interference	5
	identify relevant information missing in an argument	polarization of light	6
	identify key parts of an argument	single slit diffraction	3
Likelihood and uncertainty analysis	compute expected values in situations with known probabilities	polarization of light	5
	use probability judgments to make decisions	double slit interference	2
	predict the probability of event	single slit diffraction	3

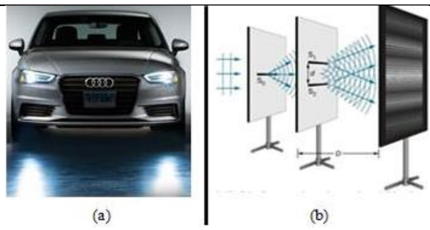
Critical Thinking Skills Indicator	Specific Domains of Critical Thinking	Sub-Topic Light Wave	Score
Problem-solving and decision-making	recognize the features of a problem and adjust solution plan accordingly	single slit diffraction	5
	examine the relevance of procedures in solving scientific problems	double slit interference	5
	evaluate solutions to a problem & make sound decisions on the basis of evidence	polarization of light	7

3.2. Disigning

The CT-LiWa assessment was designed by modifying the CTEM created by Tiruneh, (2017). The language transfer from English to Indonesian is carried out by referring to the EYD in a standard and effective manner to make it easier for students to understand the questions on the instrument. Critical thinking skills assessment is designed based on information from the previous stage. This instrument is known as a critical thinking test instrument on the light waves concept (CT-LiWa) with an open-ended form. The designed instrument consists of 5 indicators, each of which consists of 3 specific domains of critical thinking skills and is tested on 3 sub-concepts of light waves, namely double-slit interference, single-slit diffraction, and light polarization. The following is an example of an item from CT-LiWa that was developed, Table 5.

Table 5. An example of a question

No	Critical thinking skills indicators	Specific domains of critical thinking	Question Indicator	Items
1	Reasoning	detect ambiguity and misuse of definitions	Given a picture of a double-slit interference experiment accompanied by a hypothesis/suspected experimental result. Students can provide opinions about the causes of ambiguity or errors in student thinking in the question.	A student conducted Young's double-slit interference experiment using monochromatic light. The student learned that interference occurs when two or more coherent waves meet while traveling through the same medium or path. The student obtained experimental results that were beyond his expectations.  Students predict that light will interfere to produce a dark and light pattern like the picture (b). In fact, the results of the experiment show a dark and light pattern like the picture (c). Based on this, answer the following questions: Can you find any ambiguity in the student's thinking? What causes ambiguity or error in the student's thinking? Answer: The student probably still assumes that light is a particle, so that light seems to act through the two slits without interfering and forming a pattern like image b. In fact, this experiment shows that light is a wave. Then because it passes through the slit, diffraction occurs so that the waves that leave the two small slits spread out and the combination of waves (interference) creates a dark and light pattern like image c
2	Hypothesis testing	Recognize the need for more information in drawing conclusion	Given a comparison of images of car headlights with a light interference scheme. Students can provide a hypothesis about why two car headlights do not produce an interference pattern.	Students gain knowledge that waves can experience interference. If the waves come from two slits and travel the same distance so that they are in phase, then the waves experience constructive interference. But if the waves travel different distances so that they are not in phase, then the waves experience destructive interference. In relation to this definition, students pay attention to the headlights of a car.

No	Critical thinking skills indicators	Specific domains of critical thinking	Question Indicator	Items
				 The student wondered, why does the light from two car headlights (image a) which are far apart not produce an interference pattern like in image b? Answer: The reason why there is no interference pattern from two car headlights that are far apart is because: 1. The two car headlights do not come from a coherent light source (the same wavelength and frequency, and have the same phase relationship with each other all the time.) 2. The light source must be monochromatic (single wavelength)

3.3. Developing

The content validity test was carried out on the CT-LiWa assessment draft produced in the previous stage. The content validity test was carried out by experts by answering questions on the CT-LiWa validation questionnaire. Then the answers from the experts are processed by calculating the CVR. The CVR calculation results show that all CT-LiWa question items are suitable for use. This is because the CVR results of these questions are within the accepted range with a significance level of 0.05 (Lawshe, 1975). However, the number of question items used was 15 items and underwent several revisions based on expert advice. The results of the revised instrument are arranged in a format as in Figure 4.

3.4. Implementing

After being developed, the CT-LiWa assessment design was tested on high school students in Bekasi Regency, West Java, Indonesia. The data on filling out the CT-LiWa instrument for each student is then assessed and processed at the Evaluation stage.

3.5. Evaluating

The evaluation stage is carried out by analyzing the items in the CT-LiWa assessment questions and the assessments carried out. After validating the construct, the analysis results are as follows:

3.5.1. Instrument Validity with Unidimensionality

One assessment instrument that can be used to analyze the level of validity of the question instruments created is unidimensionality (Clark & Watson, 2019; Raof et al., 2021). This analysis states whether the instrument the researcher developed can measure what it should measure, in this case students' critical thinking skills. Unidimensionality results are presented in Figure 4.

```
     Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units
                                              Eigenvalue   Observed   Expected
Total raw variance in observations     =        20.7102 100.0%          100.0%
  Raw variance explained by measures   =         5.7102  27.6%           29.9%
    Raw variance explained by persons  =         2.5121  12.1%           13.2%
    Raw Variance explained by items    =         3.1981  15.4%           16.7%
  Raw unexplained variance (total)     =        15.0000  72.4% 100.0%    70.1%
    Unexplned variance in 1st contrast =         2.5274  12.2%  16.8%
```

Figure 4. Validity instrumen results

Based on Figure 4, it can be seen that the validity measurement results refer to the raw variance explained by the measured value of 27.6% with the criterion fulfilled because it is greater than 20% (see Table 1). This means that the instrument can be said to be suitable for use in further research if viewed from the unidimensionality aspect. According to Hagell, (2014), Rasch Model analysis uses Principal Component Analysis of residual values which can measure the diversity of instruments developed to measure something that must be measured. Apart from that, there is an unexplained variance of 12.2%. According to Sumintono, (2018), the instrument developed and tested should ideally have unexplained variance not exceeding 15%.

3.5.2. Instrument reliability item and person

In this section, item and person analysis can provide information on the reliability of the question items, person reliability, and the relationship between item-person reliability on the instrument being developed. The following results are presented in Figure 5.

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	2.3	15.0	-2.42	.94	.98	.13	.82	.02
SEM	.3	.0	.16	.03	.08	.12	.13	.12
P.SD	1.6	.0	.99	.21	.47	.74	.80	.72
S.SD	1.6	.0	1.01	.21	.48	.75	.81	.73
MAX.	8.0	15.0	.26	1.15	1.92	1.50	3.34	1.81
MIN.	1.0	15.0	-3.33	.57	.43	-.63	.11	-.75

REAL RMSE 1.05 TRUE SD .00 SEPARATION .00 Person RELIABILITY .00
MODEL RMSE .96 TRUE SD .25 SEPARATION .26 Person RELIABILITY .06
S.E. OF Person MEAN = .16

MINIMUM EXTREME SCORE: 32 Person 44.4%
Person RAW SCORE-TO-MEASURE CORRELATION = .96
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .66 SEM = .96
STANDARDIZED (50 ITEM) RELIABILITY = .00

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	6.0	72.0	.00	.59	.99	.21	.82	-.03
SEM	1.6	.0	.32	.05	.04	.14	.10	.17
P.SD	6.1	.0	1.19	.17	.16	.53	.38	.64
S.SD	6.4	.0	1.23	.18	.17	.55	.39	.66
MAX.	26.0	72.0	1.77	1.03	1.21	1.18	1.56	.86
MIN.	1.0	72.0	-3.20	.36	.67	-.65	.25	-1.12

REAL RMSE .63 TRUE SD 1.01 SEPARATION 1.59 Item RELIABILITY .72
MODEL RMSE .62 TRUE SD 1.02 SEPARATION 1.65 Item RELIABILITY .73
S.E. OF Item MEAN = .32

Item RAW SCORE-TO-MEASURE CORRELATION = -.96
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000

Figure 5. Reliability instrument item-person

Figure 5 presents reliable information about person-items and the relationship between person-items through Cronbach's alpha value. Person reliability gets a score of 0.00 with unreliable criteria, while item reliability gets a score of 0.72 with moderate criteria. This states that the consistency of students' answers is inconsistent but the quality of the item instruments is in the moderate category. The results presented state that the Cronbach Alpha value of 0.66 is in the sufficient category. In this way, the instrument developed is included in the sufficient category for measuring students' critical thinking skills.

3.5.3. Logit Item and Person Measure

This analysis presents the level of students' critical thinking skills and the level of difficulty of each item. This can also be the basis for the validity and reliability results discussed previously. Therefore, further analysis of item and person logit values is necessary to clarify previous results. The following logit value results are presented in Figure 6.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
5	1	72	1.77	1.03	1.03	.34	.52	.05	.19	.19	97.5	97.5	s5
8	2	72	1.00	.75	1.07	.32	.92	.26	.23	.26	95.0	95.0	s8
15	2	72	1.00	.75	.84	-.08	.47	-.30	.33	.26	95.0	95.0	s15
1	3	72	.53	.63	1.03	.21	.61	-.30	.33	.31	92.5	92.4	s1
3	3	72	.53	.63	1.15	.46	1.56	.86	.23	.31	92.5	92.4	s3
7	3	72	.53	.63	.85	-.19	.64	-.26	.38	.31	92.5	92.4	s7
10	3	72	.53	.63	.67	-.65	.25	-1.12	.47	.31	92.5	92.4	s10
13	3	72	.53	.63	.77	-.38	.33	-.88	.43	.31	92.5	92.4	s13
14	3	72	.53	.63	.86	-.16	.48	-.55	.39	.31	92.5	92.4	s14
2	4	72	.18	.56	1.08	.32	1.25	.57	.32	.36	87.5	90.0	s2
6	7	72	-.58	.45	1.11	.49	1.14	.45	.39	.45	82.5	83.6	s6
4	9	72	-.95	.42	1.15	.72	1.20	.66	.43	.49	77.5	79.8	s4
9	10	72	-1.12	.40	.87	-.57	.69	-1.02	.59	.51	80.0	78.1	s9
12	11	72	-1.28	.39	1.21	1.06	1.20	.76	.45	.53	72.5	76.6	s12
11	26	72	-3.20	.36	1.15	1.18	1.07	.35	.66	.71	55.0	67.9	s11
MEAN	6.0	72.0	.00	.59	.99	.21	.82	-.03			86.5	87.9	
P.SD	6.1	.0	1.19	.17	.16	.53	.38	.64			11.0	8.3	

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
27	8	15	.26	.57	1.30	1.42	1.29	.76	.20	.42	46.7	67.7	27
28	6	15	-.42	.60	1.03	.20	.92	-.08	.46	.46	66.7	73.1	28
44	5	15	-.79	.63	1.14	.55	1.13	.43	.38	.48	66.7	75.4	44
65	5	15	-.79	.63	.92	-.15	.85	-.22	.54	.48	80.0	75.4	65
5	4	15	-1.21	.67	1.03	.19	.94	.06	.48	.49	86.7	78.6	05
10	4	15	-1.21	.67	.78	-.54	.70	-.45	.64	.49	86.7	78.6	10
11	4	15	-1.21	.67	.78	-.54	.70	-.45	.64	.49	86.7	78.6	11
54	4	15	-1.21	.67	1.24	.72	2.04	1.69	.24	.49	73.3	78.6	54
4	3	15	-1.70	.74	.71	-.56	.46	-.75	.71	.48	86.7	84.2	04
21	3	15	-1.70	.74	.95	.05	1.29	.61	.46	.48	86.7	84.2	21
31	3	15	-1.70	.74	1.71	1.44	1.65	1.01	.03	.48	73.3	84.2	31
53	3	15	-1.70	.74	1.75	1.50	1.84	1.19	-.01	.48	73.3	84.2	53
2	2	15	-2.35	.87	1.53	.96	.89	.22	.25	.46	80.0	89.3	02
3	2	15	-2.35	.87	1.49	.91	.85	.18	.28	.46	80.0	89.3	03
6	2	15	-2.35	.87	.59	-.60	.30	-.64	.74	.46	93.3	89.3	06
7	2	15	-2.35	.87	1.32	.68	.64	-.08	.40	.46	80.0	89.3	07
13	2	15	-2.35	.87	1.92	1.42	3.34	1.81	-.25	.46	80.0	89.3	13
22	2	15	-2.35	.87	.64	-.50	.34	-.57	.71	.46	93.3	89.3	22

Figure 6. Item and personlogit values

Through the logit values obtained according to Figure 6, students' critical thinking skills and the level of difficulty of the questions can be analyzed in depth (see Table 3). In the logit items, the question with the highest level of difficulty is S5 with a logit value of 1.77 (very difficult) while the easiest question is S11 with a logit value of -3.20 (easy). Then for students who have very high critical thinking skills, namely 27 with a digit value of 0.26 (currently) respectively, while the student's critical thinking skills very weak, namely 72 with a logit value of -4.78 (low). This can clarify that the reliability of the questions created also has an impact on students' critical thinking skills. The mapping of students' literacy abilities with questions is presented in Figure 7.

```
MEASURE     Person - MAP - Item                                GUTTMAN SCALOGRAM OF RESPONS
                 <more>|<rare>                                 Person |Item
   2                   +                                              |11       111 1
                       |                                              |129462137034855
                       |  s5                                          |---------------
                       |                                           27 +101010001111010  27
                       |                                           28 +101101100100000  28
                       |                                           44 +100110001000100  44
                       |                                           65 +111000000110000  65
                       |                                            5 +110001010000000  05
                       |S                                          10 +111000000001000  10
                       |                                           11 +111000000010000  11
   1                   +  s15     s8                               54 +100100010000001  54
                       |                                            4 +110010000000000  04
                       |                                           21 +101000000000010  21
                       |                                           31 +000101100000000  31
                       |                                           53 +000100100001000  53
                       |  s1      s10     s13     s14     s3     s7   2 +000110000000000  02
                       |                                            3 +001010000000000  03
                   .   |                                            6 +110000000000000  06
                       |  s2                                        7 +011000000000000  07
```

Figure 7. Mapping student profiles and question items

Based on Figure 7, it can be seen that the level of critical thinking skills of students and the level of difficulty of the questions they can answer. If we look at students, no one students have very high abilities and can answer all the questions. In the question items section, S5 is the most difficult question and is followed by questions S15 and S8. Students have diverse critical thinking skills so they have different answer patterns. From this pattern, it is estimated that there are no students cheating.

3.5.4. Item-Person Fit Order

This analysis has a role in analyzing the suitability of items and people known as outliers. In this analysis three main components can be used as a basis for deciding whether to accept or not, namely, (1) MNSQ is accepted if, 0.5 < MNSQ < 1.5, (2) ZSTD is accepted if, -2.0 < ZSTD < +2.0, and (3) Pt mean Corr is accepted if, 0.4 < Pt Mean Corr < 0.85 (Sumintono, 2018). The results are presented in Figure 8.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
5	1	72	1.77	1.03	1.03	.34	.52	.05	.19	.19	97.5	97.5	s5
8	2	72	1.00	.75	1.07	.32	.92	.26	.23	.26	95.0	95.0	s8
15	2	72	1.00	.75	.84	-.08	.47	-.30	.33	.26	95.0	95.0	s15
1	3	72	.53	.63	1.03	.21	.61	-.30	.33	.31	92.5	92.4	s1
3	3	72	.53	.63	1.15	.46	1.56	.86	.23	.31	92.5	92.4	s3
7	3	72	.53	.63	.85	-.19	.64	-.26	.38	.31	92.5	92.4	s7
10	3	72	.53	.63	.67	-.65	.25	-1.12	.47	.31	92.5	92.4	s10
13	3	72	.53	.63	.77	-.38	.33	-.88	.43	.31	92.5	92.4	s13
14	3	72	.53	.63	.86	-.16	.48	-.55	.39	.31	92.5	92.4	s14
2	4	72	.18	.56	1.08	.32	1.25	.57	.32	.36	87.5	90.0	s2
6	7	72	-.58	.45	1.11	.49	1.14	.45	.39	.45	82.5	83.6	s6
4	9	72	-.95	.42	1.15	.72	1.20	.66	.43	.49	77.5	79.8	s4
9	10	72	-1.12	.40	.87	-.57	.69	-1.02	.59	.51	80.0	78.1	s9
12	11	72	-1.28	.39	1.21	1.06	1.20	.76	.45	.53	72.5	76.6	s12
11	26	72	-3.20	.36	1.15	1.18	1.07	.35	.66	.71	55.0	67.9	s11
MEAN	6.0	72.0	.00	.59	.99	.21	.82	-.03			86.5	87.9	
P.SD	6.1	.0	1.19	.17	.16	.53	.38	.64			11.0	8.3	

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
27	8	15	.26	.57	1.30	1.42	1.29	.76	.20	.42	46.7	67.7	27
28	6	15	-.42	.60	1.03	.20	.92	-.08	.46	.46	66.7	73.1	28
44	5	15	-.79	.63	1.14	.55	1.13	.43	.38	.48	66.7	75.4	44
65	5	15	-.79	.63	.92	-.15	.85	-.22	.54	.48	80.0	75.4	65
5	4	15	-1.21	.67	1.03	.19	.94	.06	.48	.49	86.7	78.6	05
10	4	15	-1.21	.67	.78	-.54	.70	-.45	.64	.49	86.7	78.6	10
11	4	15	-1.21	.67	.78	-.54	.70	-.45	.64	.49	86.7	78.6	11
54	4	15	-1.21	.67	1.24	.72	2.04	1.69	.24	.49	73.3	78.6	54
4	3	15	-1.70	.74	.71	-.56	.46	-.75	.71	.48	86.7	84.2	04
21	3	15	-1.70	.74	.95	.05	1.29	.61	.46	.48	86.7	84.2	21
31	3	15	-1.70	.74	1.71	1.44	1.65	1.01	.03	.48	73.3	84.2	31
53	3	15	-1.70	.74	1.75	1.50	1.84	1.19	-.01	.48	73.3	84.2	53
2	2	15	-2.35	.87	1.53	.96	.89	.22	.25	.46	80.0	89.3	02
3	2	15	-2.35	.87	1.49	.91	.85	.18	.28	.46	80.0	89.3	03
6	2	15	-2.35	.87	.59	-.60	.30	-.64	.74	.46	93.3	89.3	06
7	2	15	-2.35	.87	1.32	.68	.64	-.08	.40	.46	80.0	89.3	07
13	2	15	-2.35	.87	1.92	1.42	3.34	1.81	-.25	.46	80.0	89.3	13
22	2	15	-2.35	.87	.64	-.50	.34	-.57	.71	.46	93.3	89.3	22
MEAN	1.2	15.0	-3.47	1.37	.98	.13	.82	.02			86.5	87.9	
P.SD	1.6	.0	1.39	.51	.47	.74	.80	.72			10.3	6.8	

Figure 8. Item-person fit order

Figure 8 presents the results of misfit mapping between person and item to measure students' literacy abilities. In the item section, the number of item logits from the mean and standard deviation is obtained: 0.99 + 0.16 = 1.15. In this section, the researcher uses the ZSTD criteria so that it can be obtained that all fit order items are accepted because they are between the values ZSTD. This indicates that all question items are acceptable. Then look at students' abilities referring to grades ZSTD is also in the accepted category, namely between +2.0 and -2.0. That way, between items and people there is no misfit when viewed from the value ZSTD. A deeper review was carried out using the MNSQ value, where no one item have outlier value, so all items were in the accepted category. At the same time, for one person, there were some students with scores above 1.5, namely 54(2.04), 53 (1.84), and 13(3.34) which means the three students have different abilities from other students.

The results of this study provide recommendations for educators in using the CT-LiWa test as a test instrument in measuring students' critical thinking skills, especially in the topic of light waves. Although this study provides valuable insights, there are several limitations such as not all critical thinking indicators formulated by Tiruneh can be adapted to light wave material. Adjustments and considerations are needed in choosing critical thinking indicators to be measured according to needs. For further research, it is recommended to be applied in learning and test the effectiveness of CT-LiWa in different educational environments, such as at various levels of education or in a more diverse curriculum, in order to gain a broader understanding of its application and reliability.

4. Conclusion

The quality of the CT-LiWa assessment which was developed and analyzed using the Rasch Winstep Model assistance program based on trials on 72 high school students in West Java, Indonesia, is not reliable but has good validity. This is because students have very diverse critical thinking skills. Of the 15 items tested, all items were able to differentiate individual abilities because they were in areas that were not outliers or in the -2SD to +2SD range. Therefore, the CT-LiWa assessment can be used as a recommended instrument for measuring individual critical thinking skills.

Author Contributions

All authors have equal contributions to the paper. All the authors have read and approved the final manuscript.

Funding

No funding support was received.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Adams, D., Sumintono, B., Mohamed, A., & Noor, N. S. M. (2018). E-learning readiness among students of diverse backgrounds in a leading Malaysian higher education institution. *Malaysian Journal of Learning and Instruction, 15*(2), 227–256. https://doi.org/10.32890/mjli2018.15.2.9

Aditya, R., Suhendi, E., & Samsudin, A. (2023). Exploring high school students' critical thinking skills using active learning: A systematic literature review. *International Journal of Engineering Research & Technology (IJERT), 12*(6), 1–8. https://www.ijert.org/exploring-high-school-students-critical-thinking-skills-using-active-learning-a-systematic-literature-review IJERT

Ariani, T. (2020). Analysis of students' critical thinking skills in physics problems. *Kasuari: Physics Education Journal (KPEJ), 3*(1), 1–17. https://doi.org/10.37891/kpej.v3i1.119 ResearchGate+1

Arsad, N., Kamal, N., Ayob, A., Sarbani, N., Tsuey, C. S., Misran, N., & Husain, H. (2013). Rasch model analysis on the effectiveness of early evaluation questions as a benchmark for new students' ability. *International Education Studies, 6*(6), 185–190. https://doi.org/10.5539/ies.v6n6p185

Aulia, E. (2022). Effects of 21st century learning on the development of critical thinking, creativity, communication, and collaboration skills. *Journal of Nonformal Education, 8*(1), 46–53. https://doi.org/10.15294/jne.v8i1.33334

Bhakti, Y. B., Arthur, R., & Supriyati, Y. (2023). Development of an assessment instrument for critical thinking skills in physics: A systematic literature review. *Journal of Physics: Conference Series, 2596*(1), Article 012067. https://doi.org/10.1088/1742-6596/2596/1/012067

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment, 31*(12), 1412–1427. https://doi.org/10.1037/pas0000626

Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher, 18*(3), 4–10. https://doi.org/10.3102/0013189X018003004

Ennis, R. H. (1993). Critical thinking assessment. *Theory Into Practice, 32*(3), 179–186. https://doi.org/10.1080/00405849309543594

Ennis, R. H. (2003). *Critical thinking assessment.* In D. Fasko (Ed.), *Critical thinking and reasoning: Current research, theory and practice* (pp. —-). Hampton Press.

Facione, P. A. (1990a). The California Critical Thinking Skills Test: College level. California Academic Press.

Facione, P. A. (1990b). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report). American Philosophical Association. https://www.insightassessment.com

Fisher, W. P. (2007). Rating scale instrument quality criteria. Rasch Measurement Transactions, 21(1), 1095.

Gunawan, G., Harjono, A., Herayanti, L., & Husein, S. (2019). Problem-based learning approach with supported interactive multimedia in physics course: Its effects on critical thinking disposition. Journal for the Education of Gifted Young Scientists, 7(4), 1075–1089. https://doi.org/10.17478/jegys.627162

Hagell, P. (2014). Testing rating scale unidimensionality using the principal component analysis (PCA)/t-test protocol with the Rasch model: The primacy of theory over statistics. Open Journal of Statistics, 4(6), 456–465. https://doi.org/10.4236/ojs.2014.46044

Halimatun Sa, L., Siahaan, P., Suhendi, E., Samsudin, A., Aminudin, A. H., Rais, A., Sari, I., & Rachmadtullah, R. (2020). Critical Thinking Instrument Test (CTIT): Developing and analyzing Sundanese students' critical thinking skills on physics concepts using Rasch analysis. International Journal of Psychosocial Rehabilitation, 24(8), 3276–3290. https://doi.org/10.37200/IJPR/V24I8/PR281423

Halpern, D. F. (2003). The "how" and "why" of critical thinking assessment. In D. Fasko (Ed.), Critical thinking and reasoning: Current research, theory, and practice (pp. 1–23). Hampton Press.

Halpern, D. F. (2007). Halpern Critical Thinking Assessment using everyday situations: Background and scoring standards. Claremont McKenna College.

Halpern, D. F. (2010). Halpern Critical Thinking Assessment: Manual. Schuhfried GmbH.

Halpern, D. F. (2014). Thought and knowledge: An introduction to critical thinking (5th ed.). Psychology Press.

Jamil, M., Abdul Hafeez, F., & Muhammad, N. (2024). Critical thinking development for 21st century: Analysis of physics curriculum. *Journal of Social & Organizational Matters, 2*(3), 1–10. https://doi.org/10.56976/jsom.v3

Kealey, B. T., Holland, J., & Watson, M. (2005). Preliminary evidence on the association between critical thinking and performance in principles of accounting. *Issues in Accounting Education, 20*(1), 33–49. https://doi.org/10.2308/iace.2005.20.1.33

Ku, K. Y. L. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity, 4*(1), 70–76. https://doi.org/10.1016/j.tsc.2009.02.001

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Lawson, A. E. (2004). The nature and development of scientific reasoning: A synthetic view. *International Journal of Science and Mathematics Education, 2*(3), 307–338. https://doi.org/10.1007/s10763-004-3224-2

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions, 7*(4), 328. https://www.researchgate.net/publication/235361463

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/s10763-013-9442-5

Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Midwest Publications.

Park, M., & Liu, X. (2021). An investigation of item difficulties in energy aspects across biology, chemistry, environmental science, and physics. *Research in Science Education, 51*(1), 43–60. https://doi.org/10.1007/s11165-019-9819-y

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research* (Vol. 2). Jossey-Bass.

Putranta, H., & Wilujeng, I. (2019). Physics learning by PhET simulation-assisted using problem-based learning (PBL) model to improve students' critical thinking skills in work and energy chapters in MAN 3 Sleman. *Asia-Pacific Forum on Science Learning and Teaching, 20*(1), Article 7. https://www.eduhk.hk/apfslt/download/v20_issue1_files/putranta.pdf

Raof, S. A., Musta'amal, A. H., Zamzuri, F. K., & Salleh, M. H. (2021). Validity and reliability of students' perceptions on OBE approach in Malaysian VC using Rasch model. *Journal of Innovation in Educational and Cultural Research, 2*(2), 44–50. https://doi.org/10.46843/jiecr.v2i2.30

Recker, J. (2021). *Scientific research in information systems: A beginner's guide* (2nd ed.). Springer. https://doi.org/10.1007/978-3-030-85432-4

Samsudin, A., Cahyani, P. B., Purwanto, Rusdiana, D., Efendi, R., Aminudin, A. H., & Coştu, B. (2021). Development of a multitier open-ended work and energy instrument (MOWEI) using Rasch analysis to identify students' misconceptions. *Cypriot Journal of Educational Sciences, 16*(1), 16–31. https://doi.org/10.18844/cjes.v16i1.5504

Samsudin, A., Fratiwi, N. J., Ramalis, T. R., Aminudin, A. H., Coştu, B., & Nurtanto, M. (2020). Using Rasch analysis to develop multi-representation tier instrument on Newton's law (MOTION). *International Journal of Psychosocial Rehabilitation, 24*(6), 8542–8556. https://doi.org/10.37200/IJPR/V24I6/PR260875

Sapnas, K. G., & Zeller, R. A. (2002). Minimizing sample size when using exploratory factor analysis for measurement. *Journal of Nursing Measurement, 10*(2), 135–154. https://doi.org/10.1891/jnum.10.2.135.52552

Saputra, M. D., Joyoatmojo, S., Wardani, D. K., & Sangka, K. B. (2019). Developing critical-thinking skills through the collaboration of jigsaw model with problem-based learning model. *International Journal of Instruction, 12*(1), 1077–1094. https://doi.org/10.29333/iji.2019.12169a

Sumarni, W., Supardi, K. I., & Widiarti, N. (2018). Development of assessment instruments to measure critical thinking skills. *IOP Conference Series: Materials Science and Engineering, 349*(1), 012066. https://doi.org/10.1088/1757-899X/349/1/012066

Sumintono, B. (2018). Rasch model measurements as tools in assessment for learning. In *Proceedings of the 1st International Conference on Education Innovation (ICEI 2017)* (pp. 11–15). Atlantis Press. https://doi.org/10.2991/icei-17.2018.11

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

Szabo, Z. K., Körtesi, P., Guncaga, J., Szabo, D., & Neag, R. (2020). Examples of problem-solving strategies in mathematics education supporting the sustainability of 21st-century skills. *Sustainability, 12*(23), 10113. https://doi.org/10.3390/su122310113

Taube, K. T. (1997). Critical thinking ability and disposition as factors of performance on a written critical thinking test. *The Journal of General Education, 46*(2), 129–164. https://doi.org/10.1353/jge.2007.0021

Teo, P. (2019). Teaching for the 21st century: A case for dialogic pedagogy. *Learning, Culture and Social Interaction, 21*, 170–178. https://doi.org/10.1016/j.lcsi.2019.03.009

Tiruneh, D. T., De Cock, M., Weldeslassie, 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

Wartono, W., Hudha, M. N., & Batlolona, J. R. (2018). 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

Watson, G., & Glaser, E. M. (1980). *Watson–Glaser critical thinking appraisal*. Psychological Corporation.

Zakaria, N. H., Phang, F. A., & Pusppanathan, J. (2019). Physics on the go: A mobile computer-based physics laboratory for learning forces and motion. *International Journal of Emerging Technologies in Learning (iJET), 14*(24), 167–183. https://doi.org/10.3991/ijet.v14i24.12063