Research Article https://doi.org/10.12973/eu-jer.13.2.747

European Journal of Educational Research
Volume 13, Issue 2, 747 - 762.
ISSN: 2165-8714
https://www.eu-jer.com/

Critical Thinking-Independent Learning: A Model of Learning to Improve Students' Critical Thinking Skills

Faiz Hasyim
Universitas Negeri Surabaya, INDONESIA

Tjipto Prastowo
Universitas Negeri Surabaya, INDONESIA

Budi Jatmiko*
Universitas Negeri Surabaya, INDONESIA

Received: August 6, 2023 ▪ Revised: October 14, 2023 ▪ Accepted: November 13, 2023

Abstract: This study aims to develop a Critical Thinking-Independent Learning (CTIL) model to improve students' critical thinking skills that meet valid, practical, and effective criteria. This type of research was Research and Development (R&D), with a one-group pre-test and post-test design without a controlled class. The CTIL model was validated by three experts in the field of science education and then tested on 156 students in three public junior high schools in Surabaya, Indonesia. The results revealed that: (a) The CTIL model was valid based on the model validity score with an average of 3.73 (very good category) and the validity score of the learning materials with an average of 3.72 (very good category); (b) The practical CTIL model based on the averaged observations of the implementation of learning using the CTIL model was 3.88 (very good category); and (c) The CTIL model was effective based on four determining factors, namely: an increase in critical thinking skills, which was statistically significant at alpha .05; the calculated mean *N*-gain of .72 (high category), similar mean values for all test classes in all schools participated in this study; the effect size of 3.07 (strong category); and the average student response of 84% (very good category).

Keywords: *Critical Thinking-Independent Learning, critical thinking skills, effectiveness, practicality, validity.*

To cite this article: Hasyim, F., Prastowo, T., & Jatmiko, B. (2024). Critical thinking-independent learning: A model of learning to improve students' critical thinking skills. *European Journal of Educational Research, 13*(2), 747-762. https://doi.org/10.12973/eu-jer.13.2.747

Introduction

By 2025, it is anticipated that the need for critical thinking skills (CTS) will have increased when efforts are made to improve the quality of human resources (World Economic Forum, 2020). To help students gain 21st-century learning abilities, including problem-solving, autonomous learning, critical and creative thinking, responsibility, and independent learning, it is necessary to have qualified human resources (Tucker, 2014). These abilities are especially crucial for overcoming obstacles in the volatility, uncertainty, complexity, and ambiguity (VUCA) age, increasingly characterized by uncertainty, the complexity of issues, and ambiguity or ambiguous circumstances (Latha & Christopher, 2020). In this day and age, one must make quick decisions based on empirical facts (Guo & Cheng, 2019).

A high level of CTS is essential for students because it can help students solve problems effectively (Chukwuyenum, 2013; Peter, 2012). Facione (2015) also argued that critical thinking involves logical and reflective thinking processes that enable one to observe, analyze, synthesize, and conclude based on facts. These views are in line with the statement of Ennis (2011), who considered CTS as an activity of logical and reflective thinking that focused on making decisions about what to do. This confirms that a high level of CTS is essential for students to be superior human resource candidates.

It is anticipated that students with a higher level of CTS would do better academically (Dwyer et al., 2014). In addition, students with good CTS will also be able to use their minds to think abstractly, have an open attitude, and build effective communication with others (Wahyuni et al., 2020). Therefore, using appropriate learning models, CTS must be implemented through science learning.

Facts in class prove that mastering CTS has not yet met expectations. This is in line with the results of the 2018 Program for International Student Assessment (PISA), particularly in terms of the scientific literacy abilities of 15-year-olds; Indonesia is ranked 71st out of 79 countries participating in the survey with a score of 396, below the average score of 489, obtained for Organization for Economic Co-operation and Development (OECD) countries. The scientific literacy ability of Indonesian children is still at Level One of six PISA levels. It means that Indonesian children can only provide

* **Corresponding author:**
Budi Jatmiko, Universitas Negeri Surabaya, Indonesia. ✉ budijatmiko@unesa.ac.id

routine scientific explanations. In comparison, critical thinking at Level Five corresponds to students who can think complexly, namely making explanations based on evidence and critical argumentation (OECD, 2019).

The PISA results are almost the same as the results of the Trends in International Mathematics and Science Study (TIMSS) in 2015, where Indonesia was in 44th position out of 47 participating countries. Based on TIMSS 2015 for science and cognitive aspects (understanding, application, and reasoning), Indonesian students scored 397, far below the international average score of 500. It means that Indonesian students' science knowledge is categorized as limited knowledge if given explicit and clear evidence. In addition, Indonesian students do not have scientific reasoning skills in compiling explanations based on argumentation and evidence using critical analysis (Mullis et al., 2015).

PISA and TIMMS questions are based on Higher Order Thinking Skills (HOTS), but Indonesian curriculum has not yet fully practiced HOTS. This is the reason why PISA and TIMMS scores are always low. The government's policy to increase PISA scores began with HOTS content in the National Examination questions, before being replaced with Minimum Competency Assessment and Character Survey. However, the learning model used in the learning process, as well as the curriculum, still does not train HOTS as a whole understanding. Therefore, an innovation is needed to change the curriculum with HOTS content implemented in learning and appropriate learning models.

The results of an initial study on CTS were carried out on 27 students (grade VIII) in Surabaya, Indonesia. The average score before learning (pre-test) for CTS produced is still low, namely 4.39 from a score range of 1-10. Overall, the average score of the interpretation indicator is 4.15 (low category), the analysis indicator is 4.40 (low category), the evaluation indicator is 4.00 (low category), the inference indicator is 4.42 (low category), the explanation indicator is 5.00 (medium category), and the self-regulation indicator is also in the low category, namely 4.38 (Hasyim et al., 2020). Purwanto et al. (2019) supported this information, which produced a low percentage of student's CTS with details of interpretation (33.84%), analysis (33.79%), evaluation (32.00%), inference (34.93%), explanation (36.59%), and self-regulation (34.29%). The results of the present study are parallel with those of Kamsinah et al. (2020), which showed that the percentage of students' CTS in the indicators of analysis, evaluation, explanation, and self-regulation is 25.00% (low category), only the interpretation is of 88.00% (in the high range).

Based on the data analysis and processes from previous studies, there is a gap between expectations and reality. Consequently, it is essential to have a good teaching method so that students' CTS increases. In general, the guided inquiry learning model is used to improve CTS, which has five phases (Joyce & Weil, 2015; Kuhlthau et al., 2007). The guided inquiry learning model can improve CTS, but it still has limitations in improving CTS based on Facione's CTS indicators (2015). Based on the previous results, guided inquiry still has limitations in improving CTS, especially in the indicators of analysis, evaluation, and self-regulation (Hasyim et al., 2020; Ratnasari et al., 2020). In addition, students also find it difficult to explain facts accompanied by correct arguments when using the guided inquiry learning model (Bunterm et al., 2014). The syntax of the guided inquiry learning model as a whole does not include indicator of self-regulation as described by Facione (2015).

One of the alternative learning models designed to fill this gap and to meet the needs of the 21st century is the CTIL model to improve CTS that meet valid, practical, and effective criteria. In this context, the CTIL model has a syntax with five phases, namely: 1) motivation, 2) identification and analysis of phenomena, 3) independent investigation, 4) data interpretation and communicating results, and 5) reflection. All phases of the CTIL model syntax are hypothesized to support improved students' CTS. As for CTS in this study, there are six indicators: interpretation, analysis, evaluation, inference, explanation, and self-regulation (Facione, 2015).

The aim of this study is to create a CTIL model that is valid, practical, and effective in improving students' CTS. The model's validity is determined by experts through validation processes; its practicality is determined by class observations using the CTIL model; and the average *N*-gain, paired t-test/Wilcoxon test, effect size, independent t-test/Mann-Whitney test, and percentage of student replies are used to assess its effectiveness.

Methodology

Research Design

Based on the Borg and Gall (2007) development model, this study is called research and development (R&D). The goal of this study was to create the CTIL model, which included valid, practical, and efficient criteria to enhance CTS (Plomp, 2013). The CTIL model has five phases: phase 1: motivation, phase 2: identification and analysis of phenomena, phase 3: independent investigation, phase 4: data interpretation and communicating results, and phase 5: reflection. The flowchart of model development is shown in Figure 1.

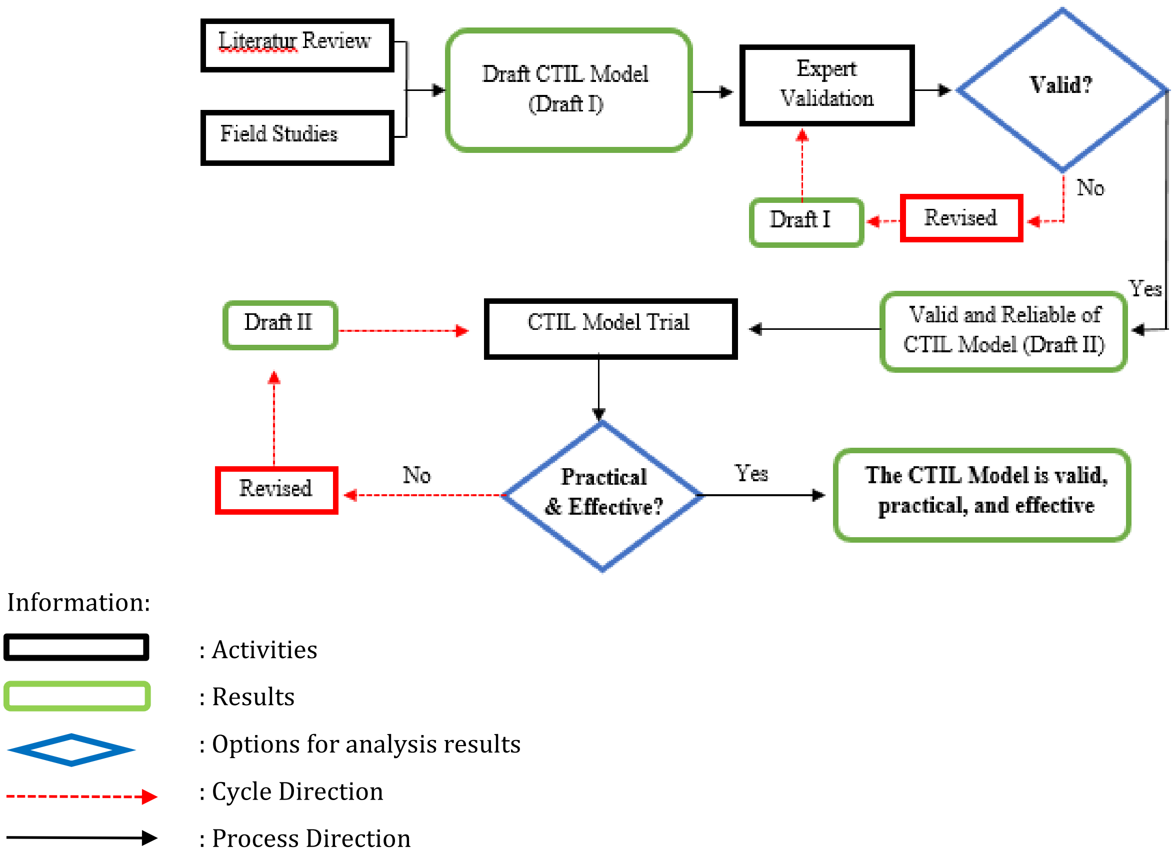

Figure 1. CTIL Model Development Design

The CTIL model's validity and practicality were assessed using expert validation results of learning materials supporting the CTIL model and observation of model implementation. Meanwhile, the pre-test and post-test results, together with the responses from the student response survey after employing the CTIL model to teach students, were used to evaluate the efficacy of the CTIL model.

Data Sample

This study used three public junior high schools (SMP Negeri) from the population of all SMP Negeri in Surabaya, Indonesia. The selection of three schools for data collection used a purposive sampling technique, namely based on the average score of the national exam in science subjects in 2019. The three schools are level A schools with high average national exam scores, namely SMP Negeri 16 Surabaya (56 students), level B schools with medium average national exam scores, namely SMP Negeri 21 Surabaya (53 students), and level C schools with low average national exam scores, namely SMP Negeri 36 Surabaya (47 students).

Instrument and Procedures

The CTIL model development process adapted the Borg and Gall (2007) development model with a flowchart (Figure 1). The preliminary stage of model development began with the needs analysis consisting of literature studies and field studies, which produce an initial draft of the CTIL model, completed with learning materials. Learning materials for the CTIL model comprised a learning syllabus, lesson plans, student teaching materials, the CTS test (CTST), and student feedback questionnaires. The second stage involved product validation by three science education experts to determine the validity and reliability of the CTIL model and its learning materials. Experts validated the draft CTIL model based on content and constructed validation instruments. The CTIL model and its learning materials' content validity, construct validity, and reliability were assessed after validation using the analytical findings. Input from experts was used to improve the draft CTIL model before it was tested in schools.

The third stage was a model test on 56 students from SMP Negeri 16 Surabaya, 53 from SMP Negeri 21 Surabaya, and 47 from SMP Negeri 36 Surabaya, Indonesia. The model testing phase's goal was to evaluate the viability and efficiency of the existing CTIL model. The CTST, the model implementation observation sheet, and student response questionnaires were the tools utilized in the model test to gauge the success of the CTIL model. The model test used a one-group pre-test and post-test design on topics: temperature and heat. Students were given a CTST pre-test before using the CTIL model, and a CTST post-test after the use of the CTIL model. The model test was carried out after students were given four learning meetings.

Data Analysis

The Validity of the CTIL Model

A validation sheet is used to test the CTIL model's content and construct validity. The validation sheet comprises an assessment sheet. The content validity of the model is seen from the ability of the statement items in the instrument to measure the achievement of content validity indicators in terms of the level of model development needs and the novelty of the model with respect to previous learning models. The ability of the statement items in the instrument to measure indicator achievements demonstrates the model's construct validity (van den Akker et al., 2013). The effectiveness of the CTIL model and related learning resources was evaluated using the evaluation findings of three scientific learning professionals.

Data from the validation process of the CTIL model were analyzed descriptively, qualitatively, and quantitatively. In this process, the validator gives a fair, objective assessment based on the instrument's scale. The results of assessment obtained from three validators were averaged and then compared with the validity criteria, as listed in Table 1.

Table 1. Validity Criteria (Hariadi et al., 2021)

Score Intervals	Assessment Criteria	Information
3.30 ≤ V < 4.00	Very Valid	It may be used without modification
2.30 ≤ V < 3.30	Valid	Useful with minor changes
1.80 ≤ V < 2.30	Invalid	It may be used with major changes
1.00 ≤ V < 1.80	Totally Invalid	It cannot be utilized; more advice is needed

Calculation of the reliability of model validity instruments and learning materials is based on the Cronbach's alpha formula,

$$r = \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum {\sigma_t}^2}{{\sigma_t}^2}\right)$$

Information:

r = reliability

n = the number of question items tested

$\sum {\sigma_t}^2$ = total variance score of each item

${\sigma_t}^2$ = total variance (Yalçin & Ereş, 2018)

Using Cronbach's alpha analysis, the instrument dependability of the CTIL model validation sheet was determined (Fraenkel et al., 2012). Table 2 shows the Cronbach's alpha reliability ranges and requirements.

Table 2. Cronbach's Alpha Reliability Intervals (Hinton et al., 2014)

Interval Cronbach's Alpha (α)	Reliability Criteria
.90 ≤ α ≤ 1	Excellent reliability
.70 ≤ α < .90	High reliability
.50 ≤ α < .70	Moderate reliability
α < .50	Low reliability

The CTIL model is said to be valid if the validator's average score is ≥ 2.75 (adapted from Ratumanan & Laurens, 2011). If the validation results of the CTIL model are less than good category, then revisions are required to improve the model and learning materials. The proportion of agreement amongst observers is used to calculate the dependability of model instruments and learning materials. If the reliability of the model instrument and learning materials reaches .75 or higher, the instrument is said to be reliable and can then be used adequately (Lestari et al., 2021).

The Practicality of the CTIL Model

A tool was utilized as an observation sheet on the learning process in class using the CTIL model to evaluate its feasibility. The observation was carried out by two observers using observation sheets with a scale of 1-4. After class observation, the scores given by the observers were averaged and analyzed based on Table 3.

Table 3. Criteria for Evaluating the Critical Thinking Independent Learning Model's Viability (Hariadi et al., 2021)

Score Intervals	Assessment Criteria
3.30 ≤ V < 4.00	Very good
2.30 ≤ V < 3.30	Good
1.80 ≤ V < 2.30	Good Enough
1.00 ≤ V < 1.80	Not good

If the average score of learning implementation using the CTIL model is at least in the good category, the model is said to be practical.

The Effectiveness of the CTIL Model

To find out the increase in students' CTS, the CTST instrument was used. Based on *N*-gain averages, paired t-test, independent t-test, and student replies, the CTIL model's performance was evaluated. The pre-test and post-test scores were calculated using the normalized gain (*N*-gain) formula to get the average *N*-gain, as indicated in the following equation:

$$N - gain = \frac{post\text{-}test\ score - \ pre\text{-}test\ score}{\text{maximum score} - \ pre\text{-}test\ score}$$

The CTIL model is said to be effective for improving CTS if the average *N*-gain is at least in the moderate category (*N*-gain ≥ .30). The complete *N*-gain classification can be seen in Table 4.

Table 4. N-Gain Category (Hasyim et al., 2020)

***N*-gain Score**	***N*-gain Category**
.70 < *N*-gain	High
.30 ≤ *N*-gain ≤ .70	Moderate
N-gain < .30	Low

Before doing the paired t-test, the normality test is an assumption test that must be passed. The normality test establishes if the data sample was taken from a population with a normally distributed population. Because there were less than 100 data points, the Shapiro-Wilk test was used statistically to determine if the data were normal. The criterion for testing normality was that if $p > .05$, the test results were said to be normally distributed. Data analysis to determine whether or not there was an increase in CTS was used paired t-test with the formula $t = \frac{\bar{d}}{\frac{s}{\sqrt{n}}}$ where $\overline{d}$ indicates the mean deviation, s the standard deviation, and n the total number of samples; with the test criteria: If alpha < .05, ignore Ho, which asserts that there is no change between the pre-test and post-test means; otherwise, Ho is accepted. The Wilcoxon test is a non-parametric test that is used when the data are not regularly distributed (Hasyim et al., 2020).

After the paired t-test was carried out, the calculation of the effect size was carried out. Effect size is a method used to measure the effectiveness of a study (Lestari et al., 2021), equation of effect size = $\frac{mean\ of\ post\ test\ score - mean\ of\ pre\ test\ score}{standard\ deviation}$ is used to calculate the effect size and the categories are shown in Table 5.

Table 5. The Effect size Category

Interval	Category
> 1.00	Strong Effect
.51 – 1.00	Moderate Effect
.21 – .50	Modest Effect
0 – .20	Weak Effect

The next test is an independent t-test to assess if the average amount of improvement in the CTS in the two model test classes from three schools is consistent. If the data for the two test model classes are not normal and/or the variance is not the same or inhomogeneous, then a nonparametric test is performed, namely the Mann-Whitney test (Siswanto et al., 2018). The results from the paired t-test and the independent t-test were analyzed using the SPSS version 25. Student response surveys to learning activities were used to collect information regarding student perception. The descriptive quantitative and qualitative methodologies were used to analyze the student response data. Positive questions were used to analyze student responses on a Likert scale. Data from the list of student response questions were analyzed using the following equation.

$$P = \frac{\sum K}{\sum N} x100\%$$

Information:

P : percentage of student response scores

∑K : overall grade received by students

∑N : the highest combined grades a student can achieve

Student response data were used to determine student responses after learning activities. These data were then converted into percentages according to the criteria in Table 6.

Table 6. Criteria for Percentage of Student Responses (Wahyuni et al., 2020)

Percentage (%)	Assessment Criteria
81 – 100	Very good
61 – 80	Good
41 – 60	Good Enough
21- 40	Not good
0 – 20	Bad

If the CTIL model satisfies the three criteria given, it is considered effective: (a) At alpha .05, there is a statistically significant increase in CTS, (b) The average *N*-gain is at least substantial and does not differ across the two model test classes, (c) the minimum effect size is moderate category, and (d) the average student response is at least good.

Results

The CTIL model syntax consists of five phases, namely: a) motivation, b) identification and analysis of phenomena, c) independent investigation, d) data interpretation and communicating results, and e) reflection. Meanwhile, the CTS indicators take all indicators from Facione (2015), namely interpretation, analysis, evaluation, inference, explanation, and self-regulation. Table 7 shows the CTIL model's stages and how they relate to the trained CTS indicators (Facione, 2015).

Table 7. CTIL Model Phases and Trained CTS Indicators

No	Phases of the CTIL Model	Trained CTS Indicators
1	Motivation	Interpretation and self-regulation
2	Identification and analysis of phenomena	Analysis, inference, and self-regulation
3	Independent investigation	Interpretation and self-regulation
4	Data interpretation and communicating results	Interpretation, analysis, inference, explanation, and self-regulation
5	Reflection	Evaluation and self-regulation

Table 8 shows the CTIL model validity scores by three science education experts, which are then calculated on the average.

Table 8. CTIL Model Validation Results

Aspect	Score of the validator			Average	Criteria
	Expert 1	Expert 2	Expert 3		
Content	3.58	3.76	3.94	3.76	Very valid
Construct	3.68	3.79	3.68	3.72	Very valid
Average	3.63	3.77	3.81	3.73	Very valid
Cronbach's Alpha (α)				.97	Excellent reliability

Table 8 shows that the average score of content and construct validity is 3.73 (very valid) with a reliability of .97 (excellent reliability). On the other hand, Table 9 shows the complete validation results of the CTIL model learning materials by three science education experts.

Table 9. Validation Results of CTIL Model Learning Devices

Tools of The CTIL Model	Score of the validator			Average	Criteria
	Expert 1	Expert 2	Expert 3		
Syllabus	3.71	3.65	3.81	3.72	Very valid
Lesson Plans	3.41	3.65	3.44	3.48	Very valid
Teaching Materials	3.41	3.65	3.44	3.48	Very valid
Student Worksheets	3.00	3.55	3.55	3.36	Very valid
CTST	3.50	3.80	3.70	3.66	Very valid
Response Questionnaire	3.40	3.66	3.58	3.54	Very valid
Average	3.71	3.65	3.81	3.72	Very valid
Cronbach's Alpha (α)				.94	Excellent reliability

According to Table 9, the CTIL model learning device's average validity score is 3.72 (very valid), and its reliability is .94 (excellent reliability). Therefore, the CTIL model and its learning materials are valid so that it can be continued at the model testing stage. Data regarding the practicality and effectiveness of the CTIL model are presented through model test results. The results of the practicality of the CTIL model are based on an assessment using an observation sheet by two observers. The observation sheet is used to evaluate how well the CTIL model of learning has been implemented. Table 10 shows the average score of two observers.

Table 10. Observation Results of Learning Using the CTIL Model in the Model Testing Stage

No	Phases of the CTIL Model	Average Score	Category
1	motivation	3.91	Very good
2	identification and analysis of phenomena	3.91	Very good
3	independent investigation	3.75	Very good
4	data interpretation and communicating results	3.90	Very good
5	reflection	3.91	Very good
	Average	3.88	**Very good**

The CTIL model application's observations have an average value of 3.88. As a result, the CTIL model is effectively applied. The effectiveness of the CTIL model is assessed in accordance with the findings of tests performed using the CTST instrument. The average level of growth (*N*-gain) and the pre- and post-test are then used to determine these findings. Tables 11, 12, and 13 provide data on the average pre-test, post-test, and *N*-gain scores from the three schools.

Table 11. Mean Scores from SMP Negeri 16 Surabaya for the Pre-Test, Post-Test, and N-Gain

Description	VIIC		VIIE	
	Pre-test	Post-test	Pre-test	Post-test
Lowest value	17	46	15	60
Highest score	50	100	58	96
Average value	29	81	41	83
Number of students	27	27	29	29
Average *N*-gain	**.73**		**.73**	
The average *N*-gain is 2 classes	**.73**			
Category	**High**			

Table 12. Mean Scores from SMP Negeri 21 Surabaya for the Pre-Test, Post-Test, and N-Gain

Description	VIIA		VIIE	
	Pre-test	Post-test	Pre-test	Post-test
Lowest value	13	46	8	54
Highest score	46	96	40	96
Average value	30	79	17	75
Number of students	24	24	29	29
Average *N*-gain	**.70**		**.70**	
The average *N*-gain is 2 classes	**.70**			
Category	**Medium**			

Table 13. Mean Scores from SMP Negeri 36 Surabaya for the Pre-Test, Post-Test, and N-Gain

Description	**VIIA**		**VIIB**	
	Pre-test	**Post-test**	**Pre-test**	**Post-test**
Lowest value	33	67	17	54
Highest score	67	100	75	100
Average value	52	86	48	87
Number of students	23	23	24	24
Average *N*-gain	**.71**		**.74**	
The average *N*-gain is 2 classes	**.73**			
Category	**High**			

Tables 11, 12, and 13 show that the average post-test score is higher than the pre-test score. The average *N*-gain of SMP Negeri 16 Surabaya is .73 in the high category, the average *N*-gain for SMP Negeri 21 Surabaya is .70 in the moderate category but is at the upper limit close to high, and the average *N*-gain for SMP Negeri 36 Surabaya is .73 is in the high category, so the average *N*-gain of all schools is .72 in the high category. First, a normality test is conducted using the Shapiro-Wilk test in two model test classes using the average *N*-gain data from three schools. This is done to determine whether there was statistically a statistically significant increase in the CTS score after learning with CTIL and whether the average level of improvement in the CTS score did not differ in the two model test classes in the three schools. The results of the normality test and the two-variance similarity test (test of homogeneity of variance) are shown in Tables 14 and 15.

Table 14. Normality Test Results

School	**Class**	**Shapiro-Wilk test**			**Conclusion**
		Data	**Statistic**	**Sig.**	
SMP Negeri 16 Surabaya	VIIC	*N*-gain	.914	.028	Non-normal
	VIIE	*N*-gain	.898	.009	Non-normal
SMP Negeri 21 Surabaya	VIIA	*N*-gain	.947	.237	Normal
	VIIE	*N*-gain	.960	.331	Normal
SMP Negeri 36 Surabaya	VIIA	*N*-gain	.925	.086	Normal
	VIIB	*N*-gain	.888	.012	Non-normal

Table 15. Homogeneity Test Results

School	**Class**	**Levene's test**		**Conclusion**
		Data	**Sig.**	
SMP Negeri 16 Surabaya	VIIC-VIIE	*N*-gain	.272	Homogen
SMP Negeri 21 Surabaya	VIIA-VIIE	*N*-gain	.930	Homogen
SMP Negeri 36 Surabaya	VIIA-VIIB	*N*-gain	.730	Homogen

Table 14 shows that the significance values for class VIIC and VIIE of SMP Negeri 16 Surabaya are .028 and .009. It may be deduced that the sample data for the two model test classes originate from aberrant populations since the two significant values are less than .05. SMP Negeri 16 Surabaya showed a significant value for class VIIA of .237 and class VIIE of .331, meaning that the sample data for the two model test classes came from the average population. SMP Negeri 36 Surabaya shows a significant value for class VIIA of .086. Because class VIIB is .012, the class test model data comes from an abnormal population, but the class sample data comes from a regularly distributed population. Homogeneity test results data based on Table 15 prove that the three schools have a significance value of more than .05, so all classes are homogeneous. Paired t-test followed schools with normally distributed samples. In contrast, the two test samples that were normally distributed and homogeneous were then continued with the independent t-test, namely SMP Negeri 21 Surabaya. At the same time, schools with samples that were not normally distributed and/or not homogeneous continued with the Wilcoxon test and the Mann-Whitney test, namely SMP Negeri 16 Surabaya and SMP Negeri 36 Surabaya.

To determine if there was an increase in CTS after learning using the CTIL model, paired t-test, Wilcoxon test and effect size were used. Table 16 displays the entire outcomes of the Wilcoxon test and the paired t-test using the SPSS version 25 program, while table 17 shows the standard deviation and effect size values.

Table 16. Paired t-Test and Wilcoxon Test Results

School	**Statistic test**	**Data**	**Sig.**	**Information**
SMP Negeri 16 Surabaya	Wilcoxon test	Pre-test-	< .001	There is a difference
SMP Negeri 21 Surabaya	Pared t-test	Post-test	< .001	There is a difference
SMP Negeri 36 Surabaya	Wilcoxon test		< .001	There is a difference

Table 17. Standard Deviation and Effect Size Values

Mean of Post-test Score	Mean of Pre-test Score	Standard Deviation	Effect Size
35.16	81.57	15.08	3.07

According to Table 16's paired t-test and Wilcoxon test findings, there is a significant difference between the pre-test and post-test scores of the two model test classes with a significance value of < .001, less than .05. It may be inferred from these findings and the positive *N*-gain values in Tables 11, 12, and 13 that the students' CTS scores increases as a consequence of utilizing the CTIL model to teach them. The result of the effect size test to know the effectiveness of the CTIL model is 3.07 with a strong effect category.

Furthermore, an independent t-test or Mann-Whitney test using the SPSS program version 25 is necessary to determine the consistency of the average degree of growth (*N*-gain average) of the CTS score across all courses. The complete test results are shown in Table 18.

Table 18. Independent t-Test and Mann-Whitney Test

School	Uji Statistic Test	Data	Sig.	Information
SMP Negeri 16 Surabaya	Mann-Whitney Test		.640	There is no difference
SMP Negeri 21 Surabaya	Independent t-Test	*N*-gain	.992	There is no difference
SMP Negeri 36 Surabaya	Mann-Whitney Test		.655	There is no difference

Table 18 informs the significance values of the results of the independent t-test and the Mann-Whitney test from three schools of .640, .992, and .655; these results are all greater than .05, meaning that there is no significant difference between the average *N*-gain in the two model test classes from each school. The two model test classes from each school experienced an increase in the CTS score at the same level of improvement. According to the test findings, the average rate of CTS score improvement across the two model test classes of each institution is comparable (consistent).

We need information on what students think about the benefits and drawbacks of the CTIL model. The information was collected through the responses written in the questionnaire provided for students at the end of their learning time (Figure 2).

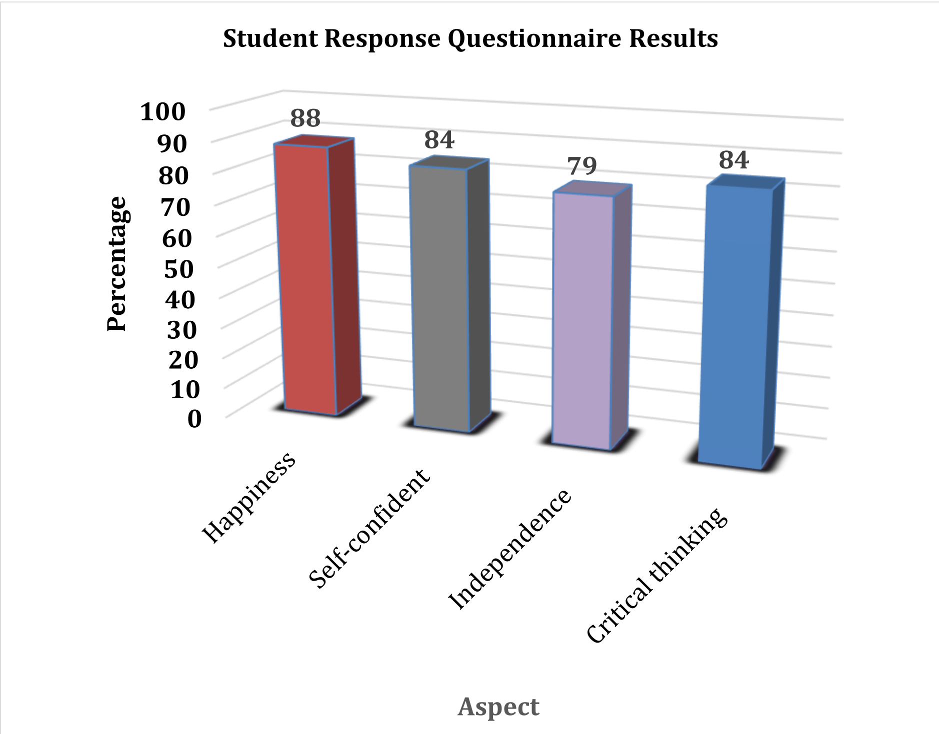

Figure 2. Student Response Questionnaire Results

Figure 2 displays the findings from students' replies after instruction utilizing the CTIL model. Student responses include 4 indicators, namely: happiness, self-confidence, independence, and critical thinking. The average answer of 84% across all metrics places the findings in the very excellent category. These outcomes demonstrate that students reacted well to the CTIL model employed in instruction. Students feel happy, confident, independent, and critical in learning.

Discussion

The CTIL model developed must meet valid, practical, and effective criteria (Jatmiko et al., 2018; Mashluhah et al., 2019; Pratiwi et al., 2019; Prayogi et al., 2018; Setiani et al., 2019; Siswanto et al., 2018; Wahyuni et al., 2020). This research begins with developing the CTIL model design and its learning materials. The first requirement is that the developed CTIL model must meet valid criteria. Therefore, three science learning experts validated the CTIL model and its learning materials. The validity of the CTIL model is divided into two aspects, namely content validity and construct validity, as presented in Table 8. The average content validity score for the CTIL model is 3.76, which is in the very excellent range. The CTIL model's mean construct validity score is 3.72, which is very good, while the reliability percentage is .97 categorized as excellent reliability. Model content validity includes model development needs, model updating, theoretical support, model planning and implementation, and model learning environment. The construct validity of the model includes the rationality of the model, the theoretical and empirical support of the model, the model syntax, the social system, the reaction principle, the support system, the instructional and accompaniment impacts, and the evaluation of the CTIL model. The criteria for content and construct validity are according to van den Akker's statement that content validity is viewed from the need for conformity between the model and design based on relevant, up-to-date knowledge, while construct validity is viewed from the consistency and logic of all components in the development of learning models (2013). The mean score of the overall model validity is 3.73, which is in the very good category, meaning that the CTIL model can be used without revision.

The validity of the CTIL model learning tools is described in full in Table 9. The learning materials include a learning syllabus, lesson plans, student teaching materials, student activity sheets, CTST, and student response questionnaires. The mean validity score for the syllabus was 3.72 in the very good category, the learning implementation plan was 3.48 in the very good category, the student teaching materials were 3.48 in the very good category, the student activity sheets were 3.36 in the very good category, the CTS test was 3.66 in the very good category, and the student response questionnaire of 3.54 is in the very good category. The mean score of the validity of the CTIL model learning tool is 3.72, with a very good category, meaning that the CTIL model learning tool can be used without revision. Well-designed learning materials are essential in setting up effective learning. Learning materials must be well designed because they function as lesson plans for teachers, learning resources, and evaluation after learning (Fink, 2012; Prayogi et al., 2018).

The second requirement is the practicality of the CTIL model. The practicality of the CTIL model and its devices is based on the results of observations by two observers on the implementation of learning using the CTIL model (Table 10). The aspects observed by the observer are by the CTIL model syntax criteria: motivation, identification and analysis of phenomena, independent performance, interpretation of data and communicating results, evaluation and reflection. The mean score of the CTIL model based on observer ratings for the motivation phase was 3.91, which was in the very good category. The identification and analysis phase of phenomena was 3.91, which was very good. The independent performance phase was 3.75, which was in the very good category. The data interpretation phase and communicating results were 3.90, which was in the very good category, and the evaluation phase and reflection of 3.91 are in the very good category. Overall, the average score of the two observers for all aspects is 3.88, which is in the very good category, meaning that learning in each phase of the CTIL model syntax is going relatively well. The teacher carries out each learning phase according to the plan that has been written in the lesson plan.

The third requirement in the development of the CTIL model is effectiveness. The effectiveness of the CTIL model was evaluated through several aspects, including mean n-gain (Tables 11, 12, and 13), paired t-test and Wilcoxon test (Table 17), effect size, independent t-test and Mann-Whitney test (Table 18), as well as student responses to learning with the CTIL model (Figure 2). From Table 11, Table 12, and Table 13, it was found that the post-test scores were higher than the pre-test scores, and the average n-gain for all classes was at least in the moderate category. Besides that, Table 17 shows that there is an increase in the CTS score which is statistically significant at alpha 5% after learning using the CTIL model. Meanwhile, Table 18 shows no difference in the level of improvement in the CTS scores of the two model test classes in 3 (three) schools. This means that the level of progress in the CTS for VIIC and VIIE students of SMP Negeri 16 Surabaya is not different; class VIIA and VIIE of SMP Negeri 21 Surabaya are no different; and class VIIA and VII B of SMP Negeri 36 Surabaya are no different. So, the level of improvement in the two model test classes for each school is consistent. These results prove that the students' CTS increased statistically significantly at 5% alpha after learning with the CTIL model with the average n-gain in the moderate minimal category and did not differ in the two model test classes in three schools. In other words, learning with CTIL can effectively improve students' CTS (Jatmiko et al., 2018). This means that the CTIL model can be used to train CTS indicators. This statement is in line with the idea of Živkovic (2016), which says that students must be prepared to think critically with the learning model the teacher uses so that students' CTS increases.

The increase in CTS for all model test classes is inseparable from the role of each phase of the CTIL model syntax. The first phase is motivation, supported by ARCS theory (Attention, Relevance, Confidence, and Satisfaction). So that curiosity and interest in learning arise, students must pay attention (Arend, 2012). The first syntax is also based on meaningful learning theory, namely processing new information related to previously learned knowledge (Moreno, 2009). Teachers who can encourage students to be involved in developing problems in learning can increase internal motivation and student persistence so that they have the high initiative in learning. In addition, student independence in observing and identifying student problems can increase motivation because when students are aware of what they are doing, they can change their behavior towards a better one (Çetin, 2015).

The second phase, namely the identification and analysis of phenomena, is supported by meaningful learning theory, which states that the processing of new information is related to previously learned knowledge (Moreno, 2009). The second syntax is also based on top-down process theory; with the help of the teacher, students analyze complex problems to solve (Slavin, 2018). The following learning theory is the dual code theory; namely, the information presented in verbal and visual form will be remembered more than information presented in one way (Slavin, 2018). The last theory that underlies the second syntax is the zone of proximal development (ZPD) which explains that student learning concepts are best when they are still in the student's closest development zone (Slavin, 2018). The results of Çetin's research (2015) state that student independence in observing and identifying student problems can increase motivation because when students know what they are doing, they can change their behavior towards a better one. In addition, identifying and solving problems can improve students' higher-order thinking skills (Temel, 2015).

The third phase, independent performance, is supported by constructivist learning theory, which states that the learning process occurs through adapting new experiences to previous experiences. This theory encourages students to think and discover things through exploratory activities to achieve meaningful new understandings (Arend, 2012; Slavin, 2018). The following learning theory is information processing theory. This theory explains how to receive environmental stimuli, organize data, solve problems, find concepts, and use verbal and visual symbols (Moreno, 2009; Slavin, 2018). Scaffolding theory also explains that students should be given complex, challenging, and realistic tasks and sufficient assistance to solve these tasks (Slavin, 2018). The next theory is self-regulated learning which provides the most comprehensive opportunity for students to manage their learning effectively in various ways to achieve optimal learning outcomes. Self-regulated learning involves metacognition, motivation, and behavior in the learning process (Schunk & Zimmerman, 2013). Rutten et al. (2015) confirmed that scientific investigations positively affect students to control their learning process through experimentation.

The fourth phase, namely data interpretation and communicating the results, is supported by dual code theory as in the second phase. Social interaction theory asserts that students' opportunities to share alternative views or ideas can help students see ideas in a different way (Eggen & Kauchak, 2012). Modeling theory states that students can learn through explanations and observations from others (Moreno, 2009). The last theory that underlies the fourth syntax is the zone of proximal development (ZPD) which explains that student learning concepts are best when they are still in the student's closest development zone (Slavin, 2018). Based on the research results of Hesse et al. (2015) stated that CTS with discussion can be trained on students, where groups of students build knowledge by working together in a social environment to learn and solve a problem or produce a product. Discussions from arguments show that students' CTS is better than CTS students who do not use arguments (Laal, 2013). Cronje et al. (2013) added the importance of argumentation skills in learning because it supports the improvement of CTS and encourages students' reasoning abilities.

The fifth phase is evaluation and reflection supported by feedback theory. This theory explains that the most effective way to teach CTS is to give students lots of practice covering various forms of problems, providing feedback on the correct answers and how students can answer these problems (Moreno, 2009). The next theory is self-regulated learning which provides the most comprehensive opportunity for students to manage their learning effectively in various ways to achieve optimal learning outcomes (Schunk & Zimmerman, 2013). Based on the results of Woolfolk's research (2016), critical thinking is evaluating conclusions by examining the problem, the evidence, and the solution logically and systematically. Evaluation of ideas makes students produce more original ideas and improves problem solving abilities (Gregory et al., 2013). Students who have self-reflection are in a good category, and the problem-solving skills they acquire are also in the good category (Jäkel & Schreiber, 2013). The existence of feedback given by the teacher to students will result in the knowledge gained being better (Arends, 2012).

After learning using the CTIL model, students fill out student response questionnaires. The purpose of the questionnaire is to identify the benefits and drawbacks of the CTIL model. According to the findings of the student response survey shown in Figure 1, each of the four indicators (happiness, self-confidence, independence, and critical thinking) is at least in the excellent range for student replies. This supports the effectiveness test results that the CTIL model can facilitate the improvement of students' CTS.

Based on the discussion above, the CTIL model is proven to be valid, practical, and effective for improving the CTS of junior high school students. The results of this study prove that theoretically and empirically, the CTIL model can be used as a learning model to improve the CTS of junior high school students.

Conclusions

This study aims to develop a valid, practical, and effective CTIL model to improve the junior high school students' CTS. The results show that the CTIL model meets the validity requirements, as indicated by the mean CTIL model validity score of 3.73, which is in the very good category, and the validity of the learning device is of 3.72, which is in the very good category. In addition, the CTIL model proves to be practical, as seen from the model's mean score of 3.88, which is categorized in the very good category. The model also proves to be effective, as indicated by the average *N*-gain value of all schools of .72, which is in the high category, the results of the paired t-test, and the Wilcoxon test with a significance value of < .001, indicating an increase in the CTS score. The effect size of 3.07, classified into strong effect category. In addition, the results of the independent t-test and the Mann-Whitney test show significance values of .640, .992, and .655, meaning that the level of increase in the CTS score in the two model test classes for each school is similar to one another or the level of increase is consistent. Student responses also give positive responses, with an average response percentage of 84%, which is very good. Therefore, the CTIL model is proven theoretically and empirically to increase the CTS of the three public junior high school students involved in the present work. The current results imply that the CTIL learning model can be used as an alternative to improve the CTS of public junior high school students.

Recommendations

To strengthen the present results, it is necessary to conduct a further study with a larger sample size and more schools involved. Such a study may provide consistent results and/or more valid, practical, and effective results. The CTIL model is recommended to be used thoroughly in schools to facilitate the improvement of students' CTS. In class, teachers should prepare learning tools and investigations independently based on the CTIL model.

Limitations

The CTIL model was fully implemented in good learning environment for public junior high school students with a total of 156 students of Grade VII being involved. Learning materials of only temperature and heat are included in the model.

Ethics Statements

SMP Negeri 16 Surabaya, SMP Negeri 21 Surabaya, and SMP Negeri 36 Surabaya examined and authorized the human subject's research. The participants provided their written informed consent to participate in this study.

Authorship Contribution Statement

Hasyim: Making concept and research design, performing data collection and acquisition, performing data analysis, drafting and editing the manuscript. Prastowo: Editing and reviewing, providing technical and material supports, supervising. Jatmiko: Editing and reviewing, suggesting critical revision, providing technical and material supports, supervising, final approval.

References

Arends, R. I. (2012). *Learning to teach* (9th ed.). Mc-Graw Hill.

Borg, W. R., & Gall, M. D. (2007). *Educational research: An introduction* (8th ed.). Pearson Cloth.

Bunterm, T., Lee, K., Ng Lan Kong, J., Srikoon, S., Vangpoomyai, P., Rattanavongsa, J., & Rachahoon, G. (2014). Do different levels of inquiry lead to different learning outcomes? A comparison between guided and structured inquiry. *International Journal of Science Education*, *36*(12), 1937-1959. https://doi.org/10.1080/09500693.2014.886347

Çetin, B. (2015). Academic motivation and self-regulated learning in predicting academic achievement in college. *Journal of International Education Research*, *11*(2), 95-106. https://doi.org/10.19030/jier.v11i2.9190

Chukwuyenum, A. N. (2013). Impact of critical thinking on performance in mathematics among senior secondary school students in Lagos State. *IOSR Journal of Research and Method in Education*, *3*(5), 18–25. https://doi.org/10.9790/7388-0351825

Cronje, R., Murray, K., Rohlinger, S., & Wellnitz, T. (2013). Using the science writing heuristic to improve undergraduate writing in biology. *International Journal of Science Education*, *35*(16), 2718-2731. https://doi.org/10.1080/09500693.2011.628344

Dwyer, C. P., Hogan, M. J., & Stewart, S. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity*, *12*, 43-52. https://doi.org/10.1016/j.tsc.2013.12.004

Eggen, P., & Kauchak, D. (2012). *Educational psychology: Windows on classrooms* (9th ed.). Pearson.

Ennis, R. H. (2011). *The nature of critical thinking: An outline of critical thinking dispositions and abilities.* University of Illinois. https://bit.ly/TheNatureofCriticalThinking

Facione, P. A. (2015). *Critical thinking: What it is and why it counts*. California Academic Press. https://bit.ly/CriticalThinkingFacione

Fink, S. B. (2012). The many purposes of course syllabi: Which are essential and useful? *Syllabus*, *1*(1), 1-12.

Fraenkel, J. C., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.

Gregory, E., Hardiman, M., Yarmolinskaya, J., Rinne, L., & Limb, C. (2013). Building creative thinking in the classroom: From research to practice. *International Journal of Educational Research*, *62*, 43–50. https://doi.org/10.1016/j.ijer.2013.06.003

Guo, X., & Cheng, L. (2019). Challenges, core competence development and future prospects of appraisers in the VUCA era. In W. Striełkowski (Eds.), *4th International Conference on Modern Management, Education Technology and Social Science (MMETSS 2019) Challenges*, (pp. 555–561.). Atlantis Press. https://doi.org/10.2991/mmetss-19.2019.112

Hariadi, B., Sunarto, M. J. D., Amelia, T., Lemantara, J., Prahani, B. K., & Jatmiko, B. (2021). Higher order thinking skills for improved learning outcomes among indonesian students: A blended web mobile learning (BWML) model. *International Journal of Interactive Mobile Technologies*, *15*(07), 4-16. https://doi.org/10.3991/ijim.v15i07.17909

Hasyim, F., Prastowo, T., & Jatmiko, B. (2020). The use of android-based phet simulation as an effort to improve students' critical thinking skills during the Covid-19 pandemic period. *International Journal of Interactive Mobile Technologies*, *14*(19), 31-41. https://doi.org/10.3991/ijim.v14i19.15701

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp 37–56). Springer. https://doi.org/10.1007/978-94-017-9395-7_2

Hinton, P. R., McMurray, I., & Brownlow, C. (2014). *SPSS explained*. Routledge. https://doi.org/10.4324/9781315797298

Jäkel, F., & Schreiber, C. (2013). Introspection in problem solving. *Journal of Problem Solving*, *6*(1), Article 4. https://doi.org/10.7771/1932-6246.1131

Jatmiko, B., Prahani, B. K., Munasir, Supardi, Z. A. I., Wicaksono, I., Erlina, N., Pandiangan, P., Althaf, R., & Zainuddin. (2018). The comparison of OR-IPA teaching model and problem based learning model effectiveness to improve critical thinking skills of pre-service physics teachers. *Journal of Baltic Science Education*, *17*(2), 300-319. https://doi.org/10.33225/jbse/18.17.300

Joyce, B., & Weil, M. (2015). *Models of teaching* (8th ed.). Pearson Education Inc.

Kamsinah, D. L., Abdullah, & Suryajaya. (2020). Analysis of critical thinking skills in junior high school students. *Journal of Advances in Education and Philosophy*, *4*(6), 234-237. https://doi.org/10.36348/jaep.2020.v04i06.002

Kuhlthau, C. C., Maniotes, L. K., & Caspari, A. K. (2007). *Guided inquiry: Learning in the 21st century*. Libraries Unlimited.

Laal, M. (2013). Positive interdependence in collaborative learning. *Procedia Social and Behavioral Science*, 93(21), 1433-1437. https://doi.org/10.1016/j.sbspro.2013.10.058

Latha, S., & Christopher, P. (2020). Vuca in engineering education: Enhancement of faculty competency for capacity building. *Procedia Computer Science*, *172*, 741-747. https://doi.org/10.1016/j.procs.2020.05.106

Lestari, T., Supardi, Z. A. I., & Jatmiko, B. (2021). Virtual classroom critical thinking as an alternative teaching model to improve students' critical thinking skills in pandemic coronavirus disease era. *European Journal of Educational Research*, *10*(4), 2003-2015. https://doi.org/10.12973/eu-jer.10.4.2003

Mashluhah, M. I., Prahani, B. K., Suryanti, S., & Jatmiko, B. (2019). The effectiveness of OrDeP2E learning model to train the natural science problem-solving skills of primary school students. *Journal of Physics: Conference Series*, *1157*(2), Article 022011. https://doi.org/10.1088/1742-6596/1157/2/022011

Moreno, R. (2009). *Educational psychology*. John Wiley & Sons, Inc.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2015). *TIMSS 2015 international results in Mathematics*. The TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. https://bit.ly/TIMSSresult2015

Organisation for Economic Co-operation and Development. (2019). *PISA 2018 assessment and analytical framework.* OECD Publishing. https://bit.ly/Pisaresult2018

Peter, E. E. (2012). Critical thinking: Essence for teaching mathematics and mathematics problem solving skills. *African Journal of Mathematics and Computer Science Research*, *5*(3), 39–43. https://bit.ly/3S3tMnC

Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *Educational design research – Part A: An introduction* (pp. 10-51). Netherlands Institute for Curriculum Development (SLO).

Pratiwi, S., Prahani, B. K., Suryanti, S., & Jatmiko, B. (2019). The effectiveness of PO2E2W learning model on natural science learning to improve problem solving skills of primary school students. *Journal of Physics: Conference Series*, *1157*(3), Article 032017. https://doi.org/10.1088/1742-6596/1157/3/032017

Prayogi, S., Yuanita, L., & Wasis. (2018). Critical inquiry based learning: A model of learning to promote critical thinking among prospective teachers of physic. *Journal of Turkish Science Education*, *15*(1), 43–56. https://bit.ly/3SeWfXs

Purwanto, Y. P. B., Masykuri, M., Suparmi, & Elisanti, E. (2019). Analysis of science students critical thinking skill in junior high school. *Journal of Physics: Conference Series*, *1233*, Article 012086. https://doi.org/10.1088/1742-6596/1233/1/012086

Ratnasari, N., Sarwanto, & Prayitno, B. A. (2020). The role of students' critical thinking skills in junior high schools on chapter organism and its environment. *Journal of Physics: Conference Series*, *1511*, Article 012064. https://doi.org/10.1088/1742-6596/1511/1/012064

Ratumanan, G. T., & Laurens, T. (2011). *Evaluasi hasil belajar pada tingkat satuan pendidikan* [Evaluation of learning outcomes at the educational unit level]. Unesa University Press.

Rutten, N., van der Veen, J. T., & van Joolingen, W. R. (2015). Inquiry-based whole-class teaching with computer simulations in physics. *International Journal of Science Education*, *37*(8), 1225-1245. https://doi.org/10.1080/09500693.2015.1029033

Schunk, D. H., & Zimmerman, B. J. (2013). Self-regulation and learning. In W. M. Reynolds, I. B. Weiner, & I. B. Weiner (Eds.). *Handbook of psychology: Educational psychology* (pp. 45–68). John Wiley & Sons, Inc.

Setiani, R., Sanjaya, I. G. M., & Jatmiko, B. (2019). ARICESA as an alternative learning model to improve learning motivation and understanding of student concepts. *International Journal of Instruction, 12*(2), 383-398. https://doi.org/10.29333/iji.2019.12225a

Siswanto, J., Susantini, E., & Jatmiko, B. (2018). Practicality and effectiveness of the IBMR teaching model to improve physics problem solving skills. *Journal of Baltic Science Education, 17*(3), 381-394. https://doi.org/10.33225/jbse/18.17.381

Slavin, R. E. (2018). *Educational psychology: Theory and practice* (12th ed.). Pearson.

Temel, V. (2015). The problem solving skill of the teachers in various branches. *Educational Research and Reviews, 10*(5), 641-647. https://doi.org/10.5897/ERR2014.2059

Tucker, S. Y. (2014). Transforming Pedagogies: Integrating 21st century skills and web 2.0 technology. *Turkish Online Journal of Distance Education*, *15*(1), 166-173. https://bit.ly/3HufQOH

van den Akker, J., Bannan, B., Kelly, A. E., Nieveen, N., & Plomp, T. (2013). Formative evaluation in educational design research. In T. Plomp & N. Nieveen (Eds.), *Educational design research – Part A: An introduction* (pp. 152-159). Netherlands Institute for Curriculum Development (SLO).

Wahyuni, S., Erman, Sudikan, S. Y., & Jatmiko, B. (2020). Edmodo-based interactive teaching materials as an alternative media for science learning to improve critical thinking skills of junior high school students. *International Journal of Interactive Mobile Technologies*, *14*(09), 166-181. https://doi.org/10.3991/ijim.v14i09.13041

Woolfolk, A. (2016). *Educational psychology* (13th ed.). Pearson.

World Economic Forum. (2020). *The future of jobs report 2020*. https://bit.ly/worldeconomicforum1

Yalçin, M. T., & Ereş, F. (2018). A study of validity and reliability on the Instructional Capacity Scale. *Universal Journal of Educational Research, 6*(1), 57-67. https://doi.org/10.13189/ujer.2018.060105

Živkovic, S. (2016). A Model of critical thinking as an important attribute for success in the 21st century. *Procedia - Social and Behavioral Sciences*, *232*, 102-108. https://doi.org/10.1016/j.sbspro.2016.10.034

Appendix

OBSERVATION SHEET OF IMPLEMENTATION OF CTIL MODEL

Instruction:

1. The following lists the aspects of activity during classroom learning activities using the CTIL model.
2. Put a tick (√) in the column that you think is appropriate. Please read each statement carefully. Check the score if your opinion,

   score 1: not good; score 3: good; and

   score 2: pretty good; score 4: very good.

Observed aspect	Done		Score			
	Yes	No	4	3	2	1
A. INTRODUCTION						
Phase 1. motivation						
1. The teacher motivates to attract students' attention using examples of events related to the material, instilling the importance of being independent in learning and building students' self-confidence to learn (self-beliefs).						
2. The teacher does an apperception followed by discussion by asking some critical questions to students.						
3. The teacher informs the learning objectives and emphasizes the importance of studying the meeting material.						
B. CORE ACTIVITIES						
Phase 2. identification and analysis of phenomena						
1. The teacher asks each student to identify and analyze the phenomena contained in the available student worksheets.						
2. The teacher stimulates students to mention the possibilities that occur with the phenomenon (arranging hypotheses) and writing them in the available worksheets.						
3. The teacher convinces students that they are able to prove their hypothesis (self-efficacy).						
Phase 3. independent investigation						
1. The teacher asks each student to independently carry out practical activities by the activities in the student worksheets, then writes the practicum result data on the available student worksheets.						
2. The teacher reminds students to control their own practicum activities (self-instruction).						
Phase 4. interpretation of data and communicating results						
1. The teacher asks students to analyze the results of the practicum data.						
2. The teacher stimulates students to argue to strengthen the results of data analysis and make conclusions.						
3. The teacher asks students to present the practicum results and responds to student presentations to develop follow-up questions from student findings.						
4. The teacher confirms students' understanding during the discussion session and explains the material.						
5. The teacher asks to check the understanding of each student's concept by writing down what has been learned on the available student worksheets (self-recording).						
C. CLOSING						
Phase 5. reflection						
1. The teacher gives students the opportunity to work on the CTST independently.						
2. The teacher confirms student answers and reviews student understanding.						
3. The teacher asks students to write a reflective self-evaluation based on their mastery of temperature and heat (self-reflection).						
Notes:						

Surabaya, ... 2022

Observer

(...)

STUDENT RESPONSE QUESTIONNAIRE SHEET TO THE CTIL MODEL

Instructions

1. In the science lesson that just passed, you have participated in learning activities with a different model.
2. Give your response to the learning activity, by ticking (√) in the column containing:

   STS : strongly disagree; S : agree;

   TS : disagree; SS : totally agree.

No.	Statement	Score			
		STS	TS	S	SS
1.	The new science learning is fun for me.				
2.	The way the teacher teaches, the contents of the books/teaching materials and worksheets are new to me.				
3.	The science lesson I just followed provoked me to ask further questions.				
4.	The new science lesson taught me to practice and study independently.				
5.	The science lesson I just took made me able to present my ideas in front of the class.				
6.	When I learned the science that I just followed, I was able to refute other people's opinions, which I considered to be incorrect.				
7.	The science lesson I just followed makes it easier for me to interpret phenomena about science material.				
8.	Science learning that I just followed makes it easier for me to conclude phenomena about natural science material.				
9.	The science lesson I just followed makes it easier for me to analyze phenomena about science material.				
10.	The science lesson I just followed makes it easier for me to evaluate (assess) phenomena about science material.				
11.	The science lesson I just attended has improved my ability to argue against the results of my own practicum and in class discussions.				
12.	Science learning that I just followed makes it easier for me to organize myself, in understanding science material.				

Surabaya, ... 2022

Respondent

(..)