Critical thinking skills assessment instrument in physics subjects: how to develop a four tier diagnostic test?

A A Waluyo*, Hartono, and Sulhadi

Physics Education, Faculty of Mathematics and Natural Science, Universitas Negeri Semarang, Indonesia

*Corresponding author: ardiwaluyo49@gmail.com

Abstract. Critical thinking is a thinking skills that involves a process of reasoning and reflective thinking that cannot be directly observed, so it requires a separate assessment instrument. This research is a development research that aims to produce an instrument for assessing critical thinking skills that can measure students’ critical thinking skills and illustrate the interconnectedness between aspects of critical thinking skills. The sample was selected using random sampling technique of 109 students in the trial stages and 136 students in the implementation stages. Data analysis was performed through several test that is content validity, construct validity, reliability, the level of students’ critical thinking skills, and analysis of variance. The validity test shows that the instrument contains 4 factors that match theorized and the content validity score shows that can be used. While the reliability test shows that the Cronbach Alpha coefficient of 0,820. The tier of critical thinking skills of high school students in Banyumas Regency is less. Analysis of the relationship aspects of critical thinking skills shows that each aspect of critical thinking skills are interconnected with the analysis aspect as the basic and the conclusion and evaluation aspects as the culmination of being mutually sustainable in producing solutions.

1. Introduction

Learning is a process that involves the thought process. This shows that the development of the efficacy of thinking is one of the goals of learning. The importance of developing thinking skills for students relates to future problems that will become more complex. One of the thinking skills needed to deal with the 21st century is critical thinking skills [1]. Critical thinking skills are a reasoning process and reflective thinking with the aim of determining what to believe or do [2]. There are 6 aspects of critical thinking skills. These aspects are interpretation, analysis, evaluation, conclusion, explanation, and self-regulation [3]. Critical thinking skills cannot be observed through an observation, so it takes an instrument that can describe critical thinking skills including showing a form of interconnectedness between each aspect. In order to show the interconnectedness between the aspects of critical thinking skills, the aspects of critical thinking skills must be measured simultaneously [4]. Problems that can be used to measure critical thinking skills must be open-ended [5]. In physics subjects, one of the materials that is open-ended is linear and parabolic motion [6]. The purpose of this research is to produce an assessment instrument that can measure the level of critical thinking skills in physics subjects and illustrate the form of interconnectedness between each aspect.
Assessment instruments developed only cover aspects of analysis, evaluation, conclusion, and explanation. Assessment instruments in form of a four-tier diagnostic test which is the development of a three-tier multiple choice diagnostic test. The adaptation is that each tier of diagnostic tests is used to measure every aspect of critical thinking skills. The first tier is used to uncover the concluding skills. The second tier is used to uncover skills in explaining. The third tier is used to uncover evaluating skills. The fourth tier is used to uncover the skills of analysis. The use of this four tier diagnostic test will ensure aspects of critical thinking skills are measured simultaneously [7].

2. Methods
This research is Research & Development (R & D) research, with steps seeing potentials and problems, collecting information and literature studies, designing the product, validating the design, revising the design, conducting product trial, conducting the revised product, implementing the revised product revision [8]. At the trial stage, the validity and reliability tests were carried out on the diagnostic tests developed so that they are feasible to use. The content validity was carried out using a validation questionnaire given to 4 validators. The construct validity tests were conducted with confirmation factor analysis on the data of assessment instrument test results [9]. The assessment instrument test was conducted by giving diagnostic test questions to 109 high school students in Banyumas Regency. The reliability tests carried out using the Alpha Cronbach formula in the assessment instrument test results data. At the implementation stage carried out the analysis of students’ critical thinking skill level and analysis of the related aspects of critical thinking skills by providing diagnostic test questions to 136 high school students in Banyumas Regency. Analysis of students’ critical thinking skill level carried out by confirming the scores obtained by students on each aspect with critical thinking skill level category [10]. The analysis of related aspects of critical thinking skills were conducted with variant analysis (ANOVA) on the cumulative score of each aspect of critical thinking skills.

3. Results and Discussion
3.1. Four-tier diagnostic test
The four-tier diagnostic test developed consists of 40 questions that aim to measure 4 critical aspects of thinking in the form of explanation, analysis, conclusion, and evaluation. The first tier is a multiple choice question with 2 answer options. The second tier is a multiple choice question with 3 closed answer options and 1 open answer option. The third tier is a multiple choice question with 4 answer options. While the fourth tier is the description of how to get an answer at the first tier to third tier. The format of the developed four-tier diagnostic test can be seen in Figure 1.
3.2. Validity and reliability of four-tier diagnostic tests

The validity test of critical thinking skill assessment instruments consists of a content validity test and a construct validity test. The validity of the contents is obtained through instrument validation questionnaires given to 4 people, 2 lecturers and 2 high school physics teachers. The grade given to critical thinking performance assessment instruments by all validators is in the range of \(2.33 < N \leq 4.00\). In general the grades in this range indicates that the instrument can be used [10].

The construct validity of the assessment instrument is obtained by analyzing the test results of the assessment instrument using the analysis of confirmatory factors. Confirmatory factor analysis was conducted with One-Sample Kolmogrov-Smirnov Test, MSA KMO assumption test, Anti-Image Correlation (AIC), Communality, and Total Variance Explained.

In the Kolmogrov-Smirnov One-Sample Test, all aspects of critical thinking skills have a value of significance greater than 0.050. This means that the data is distributed normally so that further testing can be performed [9]. In the KMO MSA, the MSA in the Anti-Image Correlation table, and the extraction score on the communal table have a minimum limit score of 0.50 [9]. If there is a score that is below the minimum score, it must be issued and re-analyzed. In the instrument developed there are some questions that do not qualify up to 40 questions remaining 35 questions which is then done factor analysis. Confirmatory factor analysis aims to reduce items that do not correspond to previously...
determined theoretical factors [10]. The results of the factor analysis of the assessment instrument are shown by Figure 2.

![Scree Plot](image)

**Figure 2.** Diagram scree plot total variance explained:
(a) 35 item question and (b) 12 item question.

Figure 2a. is a Total Variance Explained diagram of 35 question points that shows there are 9 points that have an eigen value greater than 1.00, so it can be said that a factor of 9 factors is formed. The factor formed exceeds the number of factors used as the basis for the theory on the preparation of instruments, so the number of items must be reduced. While Figure 2b. shows a total variance explained diagram of 12 question items after reduction. Based on the analysis of confirmatory factors the assessment instruments obtained 12 valid questions constructed.

In reliability testing the size of the alpha coefficient was 0.820. This suggests that the instrument scoring critical thinking skills is reliable [12]. In other words critical thinking skill assessment instruments have good internal consistency.

### 3.3. Student's critical thinking skills level

Critical thinking skills are measured by implementing 12 items of assessment instruments that have been tested for validity and reliability. The result shows all aspects of critical thinking skills are in the less category. The uniformity of the category tier in these aspects of thinking skills is due to the failure of students to analyze basic information that is the background of the problem. Basic information is an important foundation in one's success in answering questions [13]. The assessment instrument developed, directly combines the assessment of aspects of critical thinking skills with the background of the same problem. That is, against the background of this same problem will ensure students use every aspect of their critical thinking on the same basic information. So, when students fail to analyze basic information it will result in other aspects. Combining multiple choice questions and description questions on assessment instruments will ensure students use their thought process to answer questions [14]. So when a student answers correctly on every four-tier diagnostic question due to the processing
of basic information he has with a thought process. Based on the above description and the results obtained, the instruments developed can be used to measure students' critical thinking skills.

3.4. Analysis of interconnected forms aspects of critical thinking skills
The relationship of each aspect of critical thinking skills is demonstrated by the significance of the average difference of each aspect using variant analysis (ANOVA). Variant analysis shows a calculated F value of 7.565. That is, the results obtained in one aspect will affect the other aspect. The interconnectedness of aspects of critical thinking skills is demonstrated by the analysis of homogenous subsets.

| No | Critical Thinking Aspects | Subset for alpha = 0.05 |
|----|---------------------------|------------------------|
|    |                           | 1     | 2     | 3     |
| 1  | Evaluation                | 41,60 |
| 2  | Analysis                  | 42,96 | 42,96 |
| 3  | Explanation               | 49,72 | 49,72 |
| 4  | Conclusion                | 54,65 |

Table 1 shows aspects with insignificant average differences. The result shows that the evaluation aspect and the analysis aspect have insignificant differences in average values. The analysis aspect and the explanation aspect has an insignificant difference in average value. The explanation aspect and the concluding aspect has an insignificant difference in average value. Insignificant differences in average scores indicate that there is connected when students take the test. Simply put, the above results show that aspect of analysis is the basic of critical thinking skills [15]. While the concluding aspect and evaluating aspect is a culmination of critical thinking.

4. Conclusion
The research that has been conducted resulted in a set of critical thinking skills assessment instruments using problems in the physics subjects in the form of 12 item questions of four-tier diagnostic tests. Based on validity and reliability tests, critical thinking skill assessment instruments are worth using with the content validity score being in the usable category and the construct validity test shows there are 4 factors formed on the instrument according to the theoretical framework used. While the reliability test shows that the Cronbach Alpha coefficient of 0.820. The critical thinking skills of 136 high school students in Banyumas Regency on the aspects of explanation, analysis, conclusion, and evaluation are all in the category of less and those aspects are interconnected with the analysis aspect is basic and the aspect of conclusion and evaluation is the culmination of critical thinking and being mutually sustainable in producing solutions.

References
[1] The Partnership for 21st Century Skills 2015 P21 Framework Definition United State of America: U. S. Department of Education
[2] Ennis R H 2011 The Nature of Critical Thinking: An Outline of Critical Thinking Dispositions and Abilities Artikel Sixth International Conference on Thinking (Illinois: University of Illinois)
[3] Facione P A 2011 Critical Thinking: What It Is and Why It Counts California: Measured Reasons and The California Academic Press
[4] Tiruneh D T, De Cock M, Weldeslassie A G, Elen J, and Janssen R 2017 Int. J. Sci. Math Educ. 15 663
[5] Harjo B, Kartowagiran B, and Mahmudi A 2019 Int. J. Instr. 12 149
[6] Festiyed, Djamas D, and Ramli R 2019 J. Phys.: Conf. Ser. 1185
[7] Kaniawati I, Fratiwi N J, Danawan A, Suyana I, Samsudin A, and Suhendi E 2019 *J. Turk. Sci. Educ.* **16** 110
[8] Sugiyono 2007 *Statistika Untuk Penelitian* (Bandung: Alfabet)
[9] Hartanto 2016 *J. Ilm. Couns.* **6**
[10] Suharsimi A 2013 *Dasar-dasar Evaluasi Pendidikan* (Jakarta: Bumi Aksara)
[11] Arifin W N and Yusoff M S B 2016 Confirmatory Factor Analysis of the Universiti Sains Malaysia Emotional Quotient Inventory Among Medical Students in Malaysia *SAGE Open* 1-9
[12] McNeish D 2018 *Psychol. Methods* **23** 412
[13] Kong S C 2014 *Sci. Direct : Comput. Educ.* **78** 160
[14] Crossley S, Allen L K, Snow E L, McNamara D S 2015 *Proceeding Int. Conf. Learn. Anal. Knowl.* 203
[15] Ken C, Sukkamart A, and Sisan B 2018 *J. Int. Stud.* **11** 37