



Critical Thinking Ability of High School Students in Daily Life Acid-Base Concept

Das Salirawati ^{a,1,*}, Erfan Priyambodo ^{a,2}, Anggiani R. E. Nugraheni ^{a,3}, Rahmat Basuki ^{b,c,4}

^a Department of Chemistry Education, Universitas Negeri Yogyakarta, Kampus Karangmalang, Yogyakarta 55281, Indonesia

^b Department of Chemistry, Indonesia Defense University, Kawasan IPSC Sentul, Bogor 16810, Indonesia

^c Bureau of School Curriculum Analysis, Design, and Interpretation, Chemistry Alunmy Forum (FoAM),

Universitas Gadjah Mada

¹ das_salirawati@uny.ac.id*; ² erfan@uny.ac.id; ³ anggianiratnaningtyas@uny.ac.id; ⁴ rhmtbsq@gmail.com

* corresponding author

| Article history | Abstract |
|------------------------------|---|
| Submission : 2021-01-14 | This study aimed to comprehend the quality of critical thinking questions and measure the critical thinking ability in the daily life acid-base topic of 2 nd -grade students from six public high schools, six private high schools, and combination of them in Yogyakarta at the 2018/2019 academic year based on the answer patterns of students to the developed critical thinking questions. The method was an ex-post facto research designed with a one-sample design. The level of critical thinking ability was measured by 15 critical thinking question with five option answers (1 answer key & 4 distractors) outlined from the seven critical thinking criteria referred to established commercial critical thinking test. The items were validated by three reviewers (content, evaluation, and chemistry education experts) with the contents validity index (CVI) 1.00 and excellent for use as a research instrument to measure critical thinking ability of the students. The reliability of the instrument was satisfactory with the Cronbach alpha 0.704 in the total participant and in each item. The level of critical thinking skills in daily life acid-base topic of 2 nd -grade Yogyakarta's senior high schools at the academic year 2018/2019 of six public high schools, six private high schools, and combination both of them was 40.2% (moderate), 34.8% (deficient), 37.5% (deficient), respectively. Item analysis showed that majority of the items (11 items, 73%) was of average (recommended) difficulty and 11 (73%) items had satisfactory and no revision needed Discrimination Index (DI) with Distractor Efficiency (DE) 67.5% of all distractors. |
| Revised : 2021-02-16 | |
| Accepted : 2021-03-03 | |
| Keyword | |
| Critical thinking ability | |
| Daily life acid-base concept | |
| High school student | |
| Yogyakarta | |



This work is licensed under a

[Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

©2021 Jurnal Pendidikan Sains Universitas Muhammadiyah Semarang

1. INTRODUCTION

In this disruptive era, teachers are urged to be able to arouse students' thinking abilities at the higher cognitive level so that the students are wanted to thinking critically, creatively, innovatively, and used to contend with challenges (Dilekli & Tezci, 2016; Hwang & Chen, 2017). However, in reality only a few teachers giving questions relate to the application of concept in chemistry for their students in the learning process (Danczak et al., 2020). The teacher should present more divergent questions so that students are accustomed to using logic and reasoning. The question "why, how, and explain" should be raised more than convergent questions (what, who,

when, and where) that only designed for short and simple answers and less involving mental processes (reasoning) (Eliasson et al., 2017; Kipper & Rüttemann, 2010).

In chemistry curriculum 2013, the teachers were mandated to be able to create challenging learning, to evoke a critical, logical, creative, and innovative attitude to students. Conversely, Indonesia's result in the OECD Program for International Student Assessment (PISA) 2015 shows that Indonesia only ranked 62th from 72 countries, even though, there was a slight improvement compared to 2013 (ranked 71st from 72 country) and worse than in 2009 (ranked 57th from 72 country) (PISA, 2016). However, the latest PISA result for Indonesian students in 2018 was quite low (ranked 74th from 79 country) in reading, mathematics, and science category. The poor results of the PISA 2018 report combine with Trends in International Mathematics and Science Study (TIMSS) 2015 report for Indonesia indicate that Indonesian students troubled dealing with high-level aspects problems that need critical thinking skills i.e. analysis, assessing arguments, deduction and induction, decision making and problem solving, even though they were quite excellent dealing with the theoretical and memorizing problems (Mullis et al., 2016). This evidence exhibit learning in Indonesia relatively has not able to invite students to raise their critical thinking skills to solve problems. Whereas, in chemistry classroom, critical thinking skills immensely required to advance a strong understanding of basic chemistry concepts, students should have ability to relate their understanding of the symbolic to the atomic and macroscopic levels in daily life phenomenon (Bain et al., 2014; Hernández et al., 2014).

The need to advance the critical thinking skills of the high school (HS) students has driven to the construct and employment of a breadth of teaching innovations, and the development of a range of methods of assessing the impact of these interventions. Many of these assessment methods utilize validated, commercially available tests. Critical thinking skills tests that commonly commercially available were the California Critical Thinking Skills Test (CCTST) (Assessment Insight, 2013), the Watson-Glaser Critical Thinking Appraisal (WGCTA) (AssessmentDay, 2019), the Watson-Glaser Critical Thinking Appraisal Short Form (Pearson, 2018), the Cornell Critical Thinking Test Level Z (CCTT-Z) (The Critical Thinking Co, 2019), the Ennis-Weir Critical Thinking Essay Test (EWCTET) (Ennis, 1993), and the Halpern Critical Thinking Assessment (HCTA) (Halpern, 2010).

All of the tests above did not require any specific expertise and use a general knowledge background to make a reasonable attempt on the tests. Each test was guide by a manual containing specific instructions, norms, validity, reliability and item analysis. Several report of empirical research suggest that the WGCTA was the most outstanding test in use (Huber & Kuncel, 2016). However, recent trends reported the CCTST has achieved reputation amongst researchers since its inception. Whereas the commercial were popular, some report found these testing were inconsistent; some studies reported significant changes while others reported no significant changes in critical thinking (Behar-Horenstein & Niu, 2011). Previous publication of the CCTST or the WGCTA mentioned that some studies reporting increases, decrease, or no change in critical thinking skills over time (Carter et al., 2015). These reviews highlight the importance of experimental design when evaluating critical thinking. Review of the 27 studies found that only 7 of them exhibit significant changes in critical thinking (McMillan, 1987). McMillan concluded that the tests which were constructed by the researcher which address specific critical thinking outcomes were better compared with the critical thinking as a broad and generalized construct.

The evidence recommend that if these assessments would be implemented to the chemistry students, the context of the assessments should be in the field of chemistry so that the students may better engage with the assessment in a familiar context and reflect their actual critical thinking ability (Ennis, 1993; Halpern, 1998; McMillan, 1987). Some publication of critical thinking tests specified to chemistry students were found in the literature. Kogut (1996) developed exercises where students were required to note observations and underlying assumptions of chemical phenomena then develop hypotheses and experimental designs of particular topics such as trends in the periodic table or the ideal gas law. Similarly, the six questions consisting of a statement requiring an understanding of chemical knowledge was also developed by Jacob (2004). Students were expected to predict the conclusion was valid, possible or invalid and provide a short statement to explain their reasoning. Garratt et al. (2000) developed a book dedicated to developing chemistry critical thinking entitled 'A Question of Chemistry'.

In our best knowledge, there was very limited investigation performed to measure specific critical thinking ability in applying a chemistry concept in daily life of HS students. Thus, there was a requirement to develop a chemistry critical thinking instrument models which could be used to assist chemistry teacher, educators, and chemistry education researchers in evaluating the effectiveness of teaching interventions designed to develop the critical thinking skills of chemistry since in HS level. The developed instrument models should be in the specific topic so could accurately reflect the critical thinking skills of chemistry HS students. This study aimed to

measure the critical thinking ability of HS students on a specific topic which was the concept frequently applied in daily life: acid and base by developed critical thinking question.

2. METHOD

2.1. Research Instrument

The research instrument was 15 multiple choice problems with five option answer represent 7 criteria of critical thinking developed based on critical thinking skills assessed by commercially critical thinking test (Assessment Insight, 2013; AssessmentDay, 2019; Ennis, 1993; Halpern, 2010; Pearson, 2018; The Critical Thinking Co, 2019) and syllabus of 2013 chemistry curriculum (Kemendikbud, 2017). The multiple choice questions with five option answer was specialized by literacy concerning chemical phenomena in daily life related to the acid and base concept.

The items developed prioritize the representation of critical thinking skills, aside the distribution of cognitive skill. The content refers from various references and packaged into a concise and contextual question so that the students were required to thoroughly criticize to answer it. The 7 criteria and the indicators of each question item were presented in Table 1.

Table 1. Critical thinking criteria and indicators of each question

| No | Critical Thinking Criteria | Number of Question | Question Indicators |
|----|---|--------------------|--|
| 1 | distinguish fact, non-facts, and opinion | 1 | Determine facts that relate to some chemicals and food ingredients with acid or base properties and pH well. |
| | | 2 | Explain scientifically the reason for dealing with bee stings. |
| | | 3 | Predict the possible occurrence of an action that violates the prohibition on the use of chemicals based on the information provided. |
| 2 | distinguish definitive and temporary conclusion | 4 | Conclude correctly the relationship between the acidic properties of oranges and their function in maintaining healthy teeth. |
| | | 5 | Conclude precisely the relationship between the type of acid solution, its concentration, and pH value. |
| 3 | to test the level of confidence | 6 | Shows evidence in the form of events that support the truth of the role of soap in overcoming bee stings. |
| | | 7 | Proves that one of the solutions contained in the glass cleaner solution has base properties by calculating its pH value accurately. |
| 4 | distinguish relevant and irrelevant information | 8 | Determine precisely relevant information about the base properties of soap and other substances that are around us. |
| | | 9 | Determine relevant information relating to how to lower the pH of drinking water that is too base. |
| 5 | make a decisions | 10 | Determine the exact cause of the failure of the strawberry harvest based on the actions that have been taken. |
| | | 11 | Determine the choice of drug based on scientific reasons related to the function of the compound in the drug. |
| 6 | identify cause and effects | 12 | Find the exact cause of the type of citrus that does not show acidic properties, but neutral based on the results of the experiments conducted. |
| | | 13 | Determine the right choice of pH meter according to the object to be measured the pH. |
| 7 | consider other insights | 14 | Analyze the causes of a natural indicator cannot be used as a determinant of the end-point of titration based on consideration of various concepts held. |
| | | 15 | Raises the idea of overcoming the problem through the application of a neutralization reaction by properly utilizing the chemicals around them. |

Validation of the instrument was piloted based on Yusoff (2019) by calculating Content Validity Index (CVI) following six step of content validation: (a) preparing content validation form, (b) selecting panel of experts, <https://jurnal.unimus.ac.id/index.php/JPKIMIA/index>

(c) early instrument review, (d) revision, (e) confirmation from experts, (f) providing score each item, and (g) calculating CVI. The selected panel of experts were content expertise (Dr. Isana Supiah YL, Department of Chemistry Education, Universitas Negeri Yogyakarta), assessment expertise (Dr. Sri Yamtinah, Department of Chemistry Education, Universitas Sebelas Maret), and chemistry education expertise (Dr. Maria Paristiwati, Department of Chemistry Education, Universitas Negeri Jakarta). After calculating CVI, the valid instrument was then administered to a group of students. Reliability testing is performed to see if the indicator used is reliable or not by referring to the coefficient of Cronbach's Alpha; with the following basic decision making (Taber, 2018): (2) if Cronbach's Alpha ≥ 0.6 then the construct used reliably; (b) if Cronbach's Alpha < 0.6 then the constructs used are not reliable.

2.2. Sample Size Selection

The population of this research was the natural science specialization of 2nd-grade HS students in Yogyakarta (consist of six public HS and six private HS) in the academic year 2018/2019. The six public HSs were SMA N 3, SMA N 4, SMA N 6, SMA N 8, SMA N 9, SMA N 10, and six private HSs were SMA Muhammadiyah 1, SMA Muhammadiyah 2, SMA Stella Duce 1, SMA Stella Duce 2, SMA BOPKRI 1, and SMA BOPKRI 2. The samples were taken by cluster random sampling through draw two classes in each of these HSs. The total sample administered in this work was 694 students.

2.3. Ethics

All of the study samples in this research were informed that their participation was voluntary, anonymous, no affect their academic or professional records, and they were free to withdraw from this study at any time.

2.4. Research Design

This research was an ex-post facto study adapted from Ary et al. (2018), without any treatment to the object has been tested; the research was carried out in accordance with existing circumstances. Data collection conducted by administered students who became sample study using developed critical thinking instrument models that had been validated by three reviewers. The qualitative data obtained as an answer patterns from each student represent critical thinking ability in daily life application of acid and base concept. The data was analyzed by count the number of the correct answers for each item and transformed it into percentage of correct answers (quantitative data, %) for each item represent a criterion of each HS. Flow chart of the methodology was presented in Figure 1.

Table 2. Quantitative data (%) conversion guidance

| Percentage (Quatitative) | Category (Qualitative) |
|-----------------------------|---------------------------|
| $X \geq 80\%$ | Excellent |
| $60\% \leq X < 80\%$ | Good |
| $40\% \leq X < 60\%$ | Moderate |
| $20\% \leq X < 40\%$ | Deficient |
| $X < 20\%$ | Poor |

The percentage of correct answers was converted from quantitative data (%) into qualitative category adapted from Johnson & Christensen (2019) (Table 2). The critical thinking skill's grade (CTSG) obtained from conversion of correct answers to a 100 scale, by formula: $CTSG = 10 \times (\text{correct answer})/1.5$.

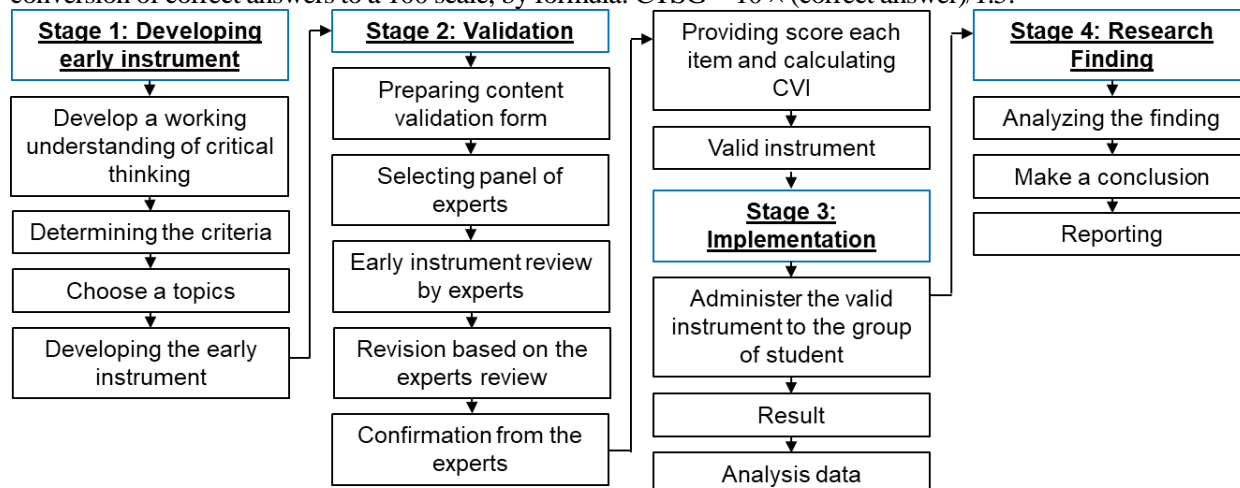


Figure 1. Flow chart of research design

2.5. Item Analysis

The items were one-best type, having a single stem and five answer options, one of them being correct answer and the other being distractors. The students were required to select the correct choice and fill in the answer sheet given separately. Each correct response was awarded 1 mark. No mark was given for blank response or incorrect answer and there was no negative marking. Thus, the maximum possible score of the overall test was 50 and the minimum 0.

The item difficulty index is calculated as a percentage of the total number of correct responses to the test items. It is calculated using the formula $p\text{-value} = R/T$ where $p\text{-value}$ is the item difficulty index, R is the number of correct responses, and T is the total number of responses (which includes both correct and incorrect responses). When multiplied by 100, $p\text{-value}$ converts to a percentage, which is the percentage of students who got the correct item. The higher the $p\text{-value}$, the easier the items. The $p\text{-value}$ between 20 and 90% are considered as good and acceptable. Among these, items with $p\text{-value}$ between 40 and 60% are considered excellent, because discrimination index (DI) is maximum at this range. Items with $p\text{-value} < 20\%$ (too difficult) and more than 90% (too easy) are not acceptable and need modification. It needs to be conceptualized that a $p\text{-value}$ is basically a behavioral measure. Instead of explaining difficulty in terms of some intrinsic substantial characteristic of the item, difficulty is defined in terms of the relative frequency with which those taking the test choose the correct response (Quaigrain & Arhin, 2017).

The item DI is the point biserial correlation between getting the item right and the total score on all other items. Then, the total number of students in the upper 27% (UG) who obtained the correct responses and the lower 27% (LG) who obtained the correct responses were counted. The DI was calculated using the formula $DI = 2(UG - LG)/n$, where n is the number of people in the two groups (Hingorjo & Jaleel, 2012). The higher the DI the better the test item discriminates between the students with higher test scores and those with lower test scores. Guidelines on classical test theory item analysis items were categorized in their discriminating indices: (1) if $DI \geq 0.40$, then the item is functioning satisfactorily; (2) if $0.30 \leq DI \leq 0.39$, then little or no revision is required; (3) if $0.20 \leq DI \leq 0.29$, then the item is marginal and needs revision; (4) if $DI \leq 0.19$, then the item should be eliminated or completely revised.

Interpretation of items is based upon the distribution of responses among the correct answers and distractors. Nonfunctional-distractors (NF-Ds) were those selected by less than 5% of students (Towns, 2014). Distractor efficiency ranged from 0 - 100% and was determined on the basis of the number of NFDs in an item. Four NF-D: DE = 0%; 3 NF-D: DE = 25%; 2 NF-D: DE = 50%; 1 NF-D: DE = 75%; No NFD: DE=100%.

3. RESULTS AND DISCUSSION

3.1. Developing Research Instrument

This study was aim to measure the level of critical thinking ability in the acid-base topic of 2nd-grade students of six public HS, six private HS, and combination of both of them in Yogyakarta for the 2018/2019 academic year based on the answer patterns of the students toward the developed critical thinking questions.

A literature review has been performed especially from the establish instrument: CCTST (Assessment Insight, 2013), WCGTA (AssessmentDay, 2019), the Watson-Glaser Critical Thinking Appraisal Short Form (Pearson, 2018), CCTT-Z (The Critical Thinking Co, 2019), EWCTET (Ennis, 1993), and the HCTA (Halpern, 2010) focused on searching for various phenomena in daily life related to the chemistry concept. The result was the seven criteria and their indicators represent the critical thinking ability criteria in daily life acid-base concept.

The questions were made in the form of multiple-choice questions with five choice options characterized by literacy about chemical phenomena related to the acid-base topic. Before being implemented, critical thinking questions were reviewed first by the three reviewers consisting of content, an evaluation, and an education expert. The results of the review from the three experts were in the form of input and recommendation for question improvement, both in the terms of the correctness of the questions according to the material substance in the questions (content expert), the quality of the questions in terms of construction and language aspects (evaluation expert), and the function and meaning of the questions as a test of critical thinking skills of students (chemistry education expert).

All input and recommendations from the three experts were followed up by improving the critical thinking questions according to the input/ recommendation given, except for input from the education expert for

question number 10 (Table 3). This was because option E in question number 10 is intentionally given so that the student with good reasoning can correlate this option with the statement on the subject matter (question stem construction). Thus, this question can be answered properly if the students are critical. After revision, the three experts declared that the questions of critical thinking ability were excellent and worthy as a research instrument. The CVI calculations on the item scale by three experts were tabulated in Table 4. Based on the calculation, it can conclude that item-CVI (I-CVI), Scale-CVI based on the average method (S-CVI/Ave) and Scale-CVI based on the Universal Agreement method (S-CVI/UA) meet satisfactory level, and thus the scale of questionnaire has achieved satisfactory level of content validity, so all of the items were acceptable (valid) for implementation step. The valid instrument was presented in Table S1 of supplementary materials.

Table 3. Input/Recommendation from three experts (Topic, Evaluation, and Education)

| Reviewers | Question Number | Input/Recommendation |
|-------------------|------------------|---|
| Topic Expert | 1 | Key B is wrong because the research results show that the pH of apples is around 4, which is acidic. |
| | 5, 8, 10, and 13 | Editorial sentences. |
| | 7 | The writing of K_b . |
| | 12 | The use of the non-standard term “where” (because it is not an interrogative sentence) |
| Evaluation Expert | 2 | Options A and C do not correlate with chemistry, should be replaced |
| | 3 | The answer has no relation with the prohibited info, so the info in the stem does not work |
| | 6 | The answer key is not related to information, preferably not because of the “forget” element |
| | 9 | Option E is used in Perusahaan Air Minum (PAM), changed to be false. |
| | 11 | Mylanta dan Promag contain the same substance, option E is not correct |
| Education Expert | 1 | Options D and E are not related to reading with a food focus, so students can get confused |
| | 3 | In the reading, it is stated that gas was produced, so students will point options C and E (there is a word gas) |
| | 11 | The reading teaches children to take matters into their own hands by treating others. This should not be conducted. Reading settings should be changed with the same purpose. |
| | 10 | Do the students even think about option E? |
| | 12 | Should not fructose be written as sugar in fruits? |
| | 13 | If asked for a tool, the answer options are only A and B. |

Table 4. Calculations on the item scale by three experts

| Item | Expert-1 | Expert-2 | Expert-3 | Expert in Agreement | I-CVI | UA |
|--|----------|----------|----------|---------------------|----------|----|
| Q1 | 1 | 1 | 1 | 3 | 1 | 1 |
| Q2 | 1 | 1 | 1 | 3 | 1 | 1 |
| Q3 | 1 | 1 | 1 | 3 | 1 | 1 |
| Q4 | 1 | 1 | 1 | 3 | 1 | 1 |
| Q5 | 1 | 1 | 1 | 3 | 1 | 1 |
| Q6 | 1 | 1 | 1 | 3 | 1 | 1 |
| Q7 | 1 | 1 | 1 | 3 | 1 | 1 |
| Q8 | 1 | 1 | 1 | 3 | 1 | 1 |
| Q9 | 1 | 1 | 1 | 3 | 1 | 1 |
| Q10 | 1 | 1 | 1 | 3 | 1 | 1 |
| Q11 | 1 | 1 | 1 | 3 | 1 | 1 |
| Q12 | 1 | 1 | 1 | 3 | 1 | 1 |
| Q13 | 1 | 1 | 1 | 3 | 1 | 1 |
| Q14 | 1 | 1 | 1 | 3 | 1 | 1 |
| Q15 | 1 | 1 | 1 | 3 | 1 | 1 |
| Proportion Relevance | 1 | 1 | 1 | S-CVI/ave | 1 | |
| Average proportion of items judged as relevance across the three experts | | | | 1 | S-CVI/UA | 1 |

Furthermore, the revised questions started to be reorganized and two versions of the questions were made, namely A and B by simply randomizing the question numbers without changing the contents of the questions. This was conducted considering the critical thinking skill test was imposed on students in two classes for each HS in sequence, so that if the order of the question numbers was the same, there was a concern that the questions would be biased. In other words, randomizing the number of the questions into two types of questions aims to anticipate question leakage because the questions consist of only 15 question items.

Two versions of the questions that had been made were ready to be implemented for data collection at the 12 HS that had been determined as the study sample. After ensuring the acid-base topic had been taught by the chemistry teacher at the HS sample, coordination with the schools was then carried out to determine the class and the day of data collection (critical thinking ability test). The results of the answer sheets from each student in each HS were then entered into the basic data table that had been prepared so that the average percentage per criteria and the overall percentage of criteria for each HS sampled could be calculated.

3.2. Measurement of Critical Thinking Ability

The internal consistency (reliability) of the instrument was satisfactory with the Cronbach alpha 0.704 in the total participant and in each item. The reliability testing by referring to the coefficient of Cronbach's alpha shows the instrument was reliable (>0.600). Based on the recapitulation provided in Table 5, it indicates that for each criterion tested the highest percentage for those who answered correctly did not always occupy by the students from the same school, both for public and private HS. As an example, for criteria 1 (ability to distinguish facts, non-facts, and opinions) represented by questions number 1, 2, and 3, the highest percentage for public school was occupied by SMA N 6 Yogyakarta (code C), which was 55.8%, while it had the smallest percentage in criteria 6 (ability to identify cause and effect) as represented by question number 12 and 13.

Table 5. Recapitulation of average percentage of each criterion for public and private HS

| Code | HS Name | Class | % Average Each Criterion | | | | | | | % Average All Criteria |
|------------------------------------|-------------------------------|--------|--------------------------|------|------|------|------|------|------|------------------------|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| A | SMA N 3 Yogyakarta | IPA-5 | 54.2 | 45.3 | 29.7 | 37.5 | 34.4 | 29.7 | 45.3 | 40.6 |
| | | IPA-6 | 46.7 | 35.0 | 25.0 | 38.3 | 36.7 | 28.3 | 31.7 | 35.3 |
| | % Average per Criteria | | 50.5 | 40.2 | 27.4 | 37.9 | 35.6 | 29.0 | 38.5 | 38.0 |
| B | SMA N 4 Yogyakarta | IPA-1 | 43.8 | 28.1 | 42.2 | 48.4 | 39.1 | 40.6 | 46.9 | 41.5 |
| | | IPA-5 | 49.0 | 45.3 | 43.8 | 40.6 | 43.8 | 51.6 | 32.8 | 44.2 |
| | % Average per Criteria | | 46.4 | 59.4 | 43.0 | 44.5 | 41.5 | 46.1 | 39.9 | 42.9 |
| C | SMA N 6 Yogyakarta | IPA-2 | 54.0 | 39.7 | 41.4 | 27.6 | 36.2 | 37.9 | 34.5 | 39.8 |
| | | IPA-3 | 57.5 | 39.7 | 53.4 | 39.7 | 41.4 | 17.2 | 29.3 | 40.9 |
| | % Average per Criteria | | 55.8 | 39.7 | 47.4 | 33.7 | 38.8 | 27.6 | 31.9 | 40.4 |
| D | SMA N 8 Yogyakarta | IPA-5 | 52.1 | 59.4 | 39.1 | 39.1 | 20.3 | 34.4 | 26.6 | 39.6 |
| | | IPA-6 | 45.8 | 39.6 | 35.4 | 29.2 | 33.3 | 41.7 | 35.4 | 37.8 |
| | % Average per Criteria | | 48.9 | 49.5 | 37.3 | 34.2 | 26.8 | 38.1 | 31.0 | 38.7 |
| E | SMA N 9 Yogyakarta | IPA-2 | 24.2 | 40.9 | 11.4 | 50.0 | 52.3 | 50.0 | 27.3 | 35.8 |
| | | IPA-6 | 28.4 | 29.4 | 39.7 | 20.6 | 39.7 | 44.1 | 25.0 | 32.2 |
| | % Average per Criteria | | 26.3 | 35.2 | 25.6 | 35.3 | 46.0 | 47.1 | 26.2 | 34.0 |
| F | SMA N 10 Yogyakarta | IPA-4 | 47.0 | 59.1 | 59.1 | 54.5 | 36.4 | 52.3 | 36.4 | 49.1 |
| | | IPA-5 | 52.4 | 42.9 | 33.3 | 59.5 | 33.3 | 35.7 | 54.8 | 45.1 |
| | % Average per criteria | | 49.7 | 51.0 | 46.2 | 57.0 | 34.9 | 44.0 | 45.6 | 47.1 |
| % Average criteria (six public HS) | | | 38.2 | 45.8 | 37.8 | 40.4 | 37.3 | 38.7 | 35.5 | 40.2 |
| Code | School name | Class | % Average each criterion | | | | | | | % Average all criteria |
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| G | SMA Muhammadiyah 1 Yogyakarta | IPA-1 | 32.3 | 34.4 | 28.1 | 31.3 | 53.1 | 34.4 | 26.6 | 34.2 |
| | | IPA-2 | 25.5 | 29.4 | 26.5 | 42.6 | 50.0 | 22.1 | 35.3 | 32.5 |
| | % Average per Criteria | | 28.9 | 31.9 | 27.3 | 37.0 | 51.6 | 28.3 | 31.0 | 33.4 |
| H | SMA Muhammadiyah 2 Yogyakarta | IPA-4 | 41.4 | 32.8 | 19.0 | 29.3 | 31.0 | 22.4 | 32.8 | 30.6 |
| | | IPA-5 | 38.5 | 34.4 | 18.8 | 32.8 | 35.9 | 34.4 | 18.8 | 31.0 |
| | % Average per Criteria | | 40.0 | 33.6 | 18.9 | 31.1 | 33.5 | 28.4 | 25.8 | 30.8 |
| I | SMA Stella Duce 1 Yogyakarta | MIPA-1 | 46.2 | 30.6 | 30.6 | 17.7 | 43.5 | 21.0 | 50.0 | 35.1 |
| | | MIPA-2 | 46.4 | 44.6 | 26.8 | 28.6 | 39.3 | 30.4 | 39.3 | 37.1 |
| | % Average per Criteria | | 46.3 | 37.6 | 28.7 | 23.2 | 41.4 | 25.7 | 44.7 | 36.1 |
| J | SMA Stella Duce 2 Yogyakarta | MIPA-1 | 48.1 | 9.7 | 56.9 | 51.4 | 29.2 | 52.8 | 47.2 | 42.6 |
| | | MIPA-2 | 53.8 | 29.5 | 37.2 | 21.8 | 37.2 | 46.2 | 43.6 | 39.5 |
| | % Average per Criteria | | 51.0 | 19.6 | 47.1 | 36.6 | 33.2 | 49.5 | 45.4 | 41.1 |
| K | SMA BOPKRI 1 Yogyakarta | MIPA-2 | 18.5 | 53.7 | 20.4 | 25.9 | 38.9 | 24.1 | 42.6 | 31.1 |
| | | MIPA-3 | 17.2 | 41.4 | 17.2 | 19.0 | 29.3 | 31.0 | 24.1 | 25.1 |
| | % Average per Criteria | | 17.9 | 47.6 | 18.8 | 22.5 | 34.1 | 27.6 | 33.4 | 28.1 |
| L | SMA BOPKRI 2 Yogyakarta | MIPA-1 | 26.3 | 15.8 | 57.9 | 39.5 | 68.4 | 57.9 | 26.3 | 40.7 |

| | | | | | | | | | |
|-------------------------------------|--------|------|------|------|------|------|------|------|------|
| | MIPA-2 | 40.4 | 73.7 | 57.9 | 25.0 | 51.3 | 39.5 | 23.7 | 39.0 |
| % Average per Criteria | | 33.4 | 44.8 | 57.9 | 10.5 | 34.2 | 21.1 | 21.1 | 37.2 |
| % Average criteria (six private HS) | | 36.3 | 35.9 | 33.1 | 29.3 | 30.4 | 33.2 | 34.0 | 34.8 |
| % Average All Criteria (twelve HS) | | 37.3 | 40.9 | 35.5 | 34.9 | 33.9 | 36.0 | 34.8 | 37.5 |

Likewise for private HS, SMA Stella Duce 2 (code J) had the highest percentage in criteria 1 (51.0%), however in criteria 2 (ability to distinguish definitive and temporary conclusions) represented by questions number 4 and 5, it had the lowest percentage (19.6%). This result also applies to other HS which indicate that students from each school show the prominent critical thinking ability on one criterion only but shows a relatively low ability at the other criteria.

Conversely, the smallest percentage was occupied by SMA N 9 Yogyakarta (Code E) in deficient category (34.0%). Likewise, for private HS, the highest percentage was occupied by SMA Stella Duce 2 Yogyakarta (Code J), in the moderate category (41.1%). On the other hand, the smallest percentage was occupied by SMA BOPKRI 1 (Code K) in the deficient category (28.1%).

When viewed from the average percentage of all criteria for public HS (40.2%) and private HS (34.8%) it showed a relatively moderate difference with an overall average (12 HS) of 37.5%. If the value was converted into qualitative criteria, then the average for public HS was in the moderate category, while private HS was in the deficient category, and the overall average was in the deficient category.

Based on these results, it shows that public and private HS, which have been known as favorite HS, was not always have students with better critical thinking ability than non-favorite school. Although this critical thinking test is only limited to the acid-base topic, at least it becomes empirical data that can illustrate that so far students were still relatively rarely accustomed to facing questions that can reveal students' critical thinking ability. So, when faced with this question, they may not be ready to digest the direction of each question they were working on (Martensson & Hansson, 2018).

Therefore, the results of this study were expected to be an input for teachers to plan a better teaching, associate subject matter with the events or phenomena that occur in the daily life or natural surroundings, so that the ability of students to criticize was well-practiced. This can be achieved if the teachers want to continuously learn, broaden horizons, and improve their quality and their professionalism by providing learning innovations so that students were not only fed with theoretical concepts but more directed at the application of the concept in life. The low learning outcomes of the students can be caused by learning chemistry only in the classroom, focuses on the theory without being directed to link the learning content with everyday life (Cano et al., 2014).

Thus, the subject matter provided is beneficial for students in solving life problems, because it provides a provision for deep concepts concerning the explanation of various natural phenomena that may occur around their lives. Parallel with that, Espinosa et al. (2013) reported that students which learning chemistry in the classroom have many hopes, one of which is a willingness to understand the benefits of chemistry in everyday life. Through thinking critically, students can find various ways for solving problems (Moore, 2001).

Table 6. Recapitulation of average score of critical thinking ability test results

| Code | Class | Average score | Total average | Code | Class | Average Value | Total Average |
|-----------------------------------|-------|---------------|---------------|--------------------------------|--------|---------------|---------------|
| A | IPA-5 | 41.5 | 47.55 | G | IPA-1 | 41.5 | 33.40 |
| | IPA-6 | 53.6 | | H | IPA-2 | 53.6 | |
| B | IPA-1 | 40.5 | 37.90 | H | IPA-4 | 40.5 | 30.85 |
| | IPA-1 | 35.3 | | I | IPA-5 | 35.3 | |
| C | IPA-2 | 39.8 | 40.40 | I | MIPA-1 | 39.8 | 36.10 |
| | IPA-3 | 41.0 | | J | MIPA-2 | 41.0 | |
| D | IPA-5 | 39.5 | 38.65 | J | MIPA-1 | 39.5 | 40.95 |
| | IPA-6 | 37.8 | | K | MIPA-2 | 37.8 | |
| E | IPA-2 | 35.8 | 33.95 | K | MIPA-2 | 35.8 | 28.05 |
| | IPA-6 | 32.1 | | L | MIPA-3 | 32.1 | |
| F | IPA-4 | 49.1 | 47.10 | L | MIPA-1 | 49.1 | 39.00 |
| | IPA-5 | 45.1 | | | MIPA-2 | 45.1 | |
| Total | | | 245.55 | Total | | | 208.35 |
| Average score (six public HS) | | | 40.93 | Average score (six private HS) | | | 34.73 |
| Overall average score (twelve HS) | | | | | | | 37.83 |

Observing from the scores obtained by students both from public and private HS, it showed that the highest average score was occupied by SMA N 3 Yogyakarta (Code A, 47.55), while the lowest average value was

occupied by SMA BOPKRI 1 Yogyakarta (Code K, 28.05) (Table 6). Interestingly, some of the public HS had an average score below private HS i.e. average score of SMA N 9 Yogyakarta (Code E, 33.95) was below from SMA Stella Duce 1 (Code I, 36.10), SMA Stella Duce 2 (Code J, 40.95), and SMA BOPKRI 2 (Code L, 39.00) (Note: the average score of six public HS, six public HS, and twelve HS in Table 6 and Table 5 were actually equal, however the value was quietly different because of the rounds of decimal digit).

Based on these results, the problem of critical thinking ability was not determined by where the students came from, however it depends on how students get used to and were practiced in using their critical attitudes in the learning situations. The possibility of chemistry teachers in private HS may be more concerned with questions that are sometimes a bit strange as asked by students, even though such questions are questions that can only cross the minds of the students who have a critical attitude (Robinson, 2011).

A student categorized as an unsuccessful student if have score less than 60 (Sunarti, 2014). This means that the students cannot master the topic in the questions. In this case, because all average value was less than 60, it means that the critical thinking ability of students from 12 HS was still relatively low (deficient). These results were not the only one problem indicator regarding the low critical thinking ability of the students which is need higher-order thinking skills. Previous studies also show this condition (Khasanah et al., 2017; Pursitasari et al., 2020).

The low critical thinking ability of students may cause by several issues. Firstly, students tend to be passive and do not respond the questions given by the teacher. The condition makes students difficult to understanding the concept (Angelo, 1995; Basuki, 2020; Basuki et al., 2018; Fitriani et al., 2019) Therefore, the teachers were necessary to have learning innovations to improve the critical thinking ability of students (Afandi et al., 2019).

Secondly, students have low critical attitude. Critical attitudes is a disciplined process that is intellectually active and skilled at conceptualizing, applying, analyzing, synthesizing, and/or evaluating information collected from various processes (Moore, 2001). One of the characteristics that students have a critical attitude is when they get information in the form of new concepts from the learning process; they quickly evaluate them in their cognitive structure so that when the concept is unclear, they will immediately ask questions. With the questions asked, students' critical thinking would be utilized in exploring problems that demand immediate answers. Relatively, many teachers do not like students who asked too many questions. Even though such children were what we expect many to be in the classroom so that the class becomes alive and teachers' horizon becomes broad, which of course affects other students in gaining knowledge more than just written in a book.

Thirdly, the teacher presented a low quality of learning process (Moore, 2001). If a teacher delivers a subject matter that students feel was less profound and then they ask questions, that were a sign that the student thinks critically, and the teacher must respond immediately so that this critical thinking continues to grow, not the other way, dying slowly because the teacher breaks the spirit by not responding to their questions. Therefore, the results of this study were expected to trigger the chemistry teachers to improve their learning quality by designing innovative lessons so that students were able to think critically. The low quality of learning process also make student tend to be passive and low participation in discussion (Oktaviana et al., 2016).

Based on the overall results obtained in this study, it implies that senior HS chemistry teachers, in particular, have to create learning that accustom and practice the critical thinking ability of their students. If learning on his own is considered too difficult, a teacher can share his knowledge with fellow friends who share the same field of knowledge, either in official forums (MGMP) or in small discussion forums through direct meetings or networks. The teacher also can apply the various methods i.e. Problem/Project Based Learning (Sumarni & Kadarwati, 2020; Yuliyanto & Rohaeti, 2013) Contextual Learning (Susanti et al., 2018), STEM (Mutakinati et al., 2018), and so forth which possible to linking the chemistry topics with the phenomena and events in real life of students can improve their critical thinking ability.

Teachers must realize that the era of globalization is an era full of challenges and competition, so that students in the future not only require simple intellectual thinking skills, but also critical thinking, communication, creative, and collaborative skills (Sipayung et al., 2018).

Hope the results of this study can be followed up in the form of community service activities, especially the education community: senior HS chemistry teachers who are packaged in a training workshop on the preparation of critical thinking ability question for each subject matter contained in chemistry subjects in senior HS as well direct practice so that the results of this study useful in providing provisions for teachers in developing critical thinking skills questions.

3.3. Item Analysis

The test consisted of 50 items. The scores of 247 students ranged from 11 to 42 (out of 50). After getting the result, students were ranked in order of merit from the highest score the lowest score. The first 27% students made the high group and the last 27%, the low group. The *p*-value, DI and DE were analyzed for each item (Table 7). Majority of the items (11 items, 73%) were of average (recommended) difficulty. Similarly, majority of items (73%) had good DI in no revision required and satisfactorily criteria. A combination of the two indices revealed that 5 (33%) items could be called 'ideal' having a *p*-value from 30 to 70, as well as a DI > 0.24. The total number of distractors were 60 (4 per item) out of which 13 (21.6%) were NF-Ds. Ten (67.7%) items had maximum two NF-Ds, while 5 (33.3%) items had effective distractors. Items with 2 NF-Ds had a high *p*-value and good DI; items number 1, 4, 6, and 14 with one or no NF-D were too difficult (*p*-value<20), and with poor DI (DI<0.19) that required to revision the items.

Table 7. Item analysis of the instruments

| Item No. | <i>p</i> -value | DI | DE (%) |
|----------|--------------------|---------------------------------------|--------|
| 1 | 13 (too difficult) | 0.14 (eliminated/ completely revised) | 100 |
| 2 | 49 (acceptable) | 0.56 (satisfactory) | 50 |
| 3 | 62 (acceptable) | 0.40 (satisfactory) | 75 |
| 4 | 17 (too difficult) | 0.18 (eliminated/ completely revised) | 100 |
| 5 | 58 (acceptable) | 0.31 (no revision required) | 75 |
| 6 | 15 (too difficult) | 0.20 (revision required) | 100 |
| 7 | 54 (acceptable) | 0.48 (satisfactory) | 75 |
| 8 | 49 (acceptable) | 0.46 (satisfactory) | 75 |
| 9 | 20 (acceptable) | 0.26 (revision required) | 100 |
| 10 | 25 (acceptable) | 0.21 (revision required) | 100 |
| 11 | 52 (acceptable) | 0.31 (no revision required) | 50 |
| 12 | 27 (acceptable) | 0.18 (eliminated/ completely revised) | 75 |
| 13 | 45 (acceptable) | 0.36 (no revision required) | 50 |
| 14 | 19 (too difficult) | 0.14 (eliminated/ completely revised) | 75 |
| 15 | 51 (acceptable) | 0.42 (satisfactory) | 75 |

3.4. Recommendations and Future Outlook

Several recommendations can be made as follows: (1) for instructor, the results of this study indicated that the critical thinking skill of students was relatively low, especially in chemistry subject. Therefore, it is necessary to organize training or workshop for teachers of chemistry subjects in particular and teachers of other subjects in general in creating learning tips that can reveal critical thinking skills of students, so that teachers can apply it in the learning process in their respective schools; (2) for teachers, the results obtained from this study are preliminary findings that should be followed up with activities to improve the learning process, self-development, increase professionalism, utilized the internet to enrich knowledge, and increase sharing with peers to be able to create familiar learning and sharpen the critical thinking skills of the students; (3) for other researchers, this study can be used as a reference and data source for use in research with different materials, larger population coverage and a wider area, and research on other high-order thinking skills, such as the ability to think creatively, innovatively, and scientifically; (4) for observers of chemistry education, in particular, considering the results of this study indicated that the critical thinking skills of students were relatively low, it is necessary to research critical thinking skills for teachers. This is maybe due to the low critical thinking skill of students because the teacher himself does not necessarily have adequate critical thinking skills.

4. CONCLUSION

The critical thinking questions have been developed based on the results of the reviews conducted by three reviewers as the topic, evaluation, and chemistry education experts. After improved, the quality of critical thinking questions was excellence (CVI=1.00) and suitable for use as a research instrument to measure critical thinking ability of the students. The reliability of the instrument was satisfactory with the Cronbach alpha 0.704 in the total participant and in each item. The level of critical thinking skills in the acid-base topic of 2nd-grade students of senior HS in Yogyakarta at the 2018/2019 academic year showed six public SMA were 40.2% (moderate), six private SMA were 34.8% (deficient), and a combination of both (public and private HS) was 37.5% (deficient). Item analysis showed that majority of the items (11 items, 73%) were of average (recommended) difficulty and 73% of items had good DI (no revision required and satisfactorily criteria) with DE 67.5% of all distractors.

ACKNOWLEDGMENT

D.S. thanks to Universitas Negeri Yogyakarta for financing this work through Research Group Program fiscal year 2019. The authors acknowledge reviewers (Dr. Isana Supiah Y.L., Dr. Sri Yaminah, and Dr. Maria Paristiwati) for validating instrument in this work.

REFERENCES

- Afandi, A., Sajidan, S., Akhyar, M., & Suryani, N. (2019). Development Frameworks of the Indonesian Partnership 21st-Century Skills Standards for Prospective Science Teachers: A Delphi Study. *Jurnal Pendidikan IPA Indonesia*, 8(1), 89–100.
- Angelo, T. A. (1995). Classroom assessment for critical thinking. *Teaching of Psychology*, 22(1), 6–7.
- Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2018). *Introduction to research in education*. Cengage Learning.
- Assessment Insight. (2013). *California critical thinking skills test (CCTST)*. <https://www.insightassessment.com/Products/Products-Summary/Critical-Thinking-Skills-Tests/California-Critical-Thinking-Skills-Test-CCTST>
- AssessmentDay. (2019). *Watson Glaser critical thinking appraisal, Critical thinking tests are high-level tests, with the Watson-Glaser being the most common*. <https://www.assessmentday.co.uk/watson-glaser-critical-thinking.htm>
- Bain, K., Moon, A., Mack, M. R., & Towns, M. H. (2014). A review of research on the teaching and learning of thermodynamics at the university level. *Chemistry Education Research and Practice*, 15(3), 320–335.
- Basuki, R. (2020). Conceptual Difficulties Experienced by First Year Undergraduate Chemistry Students to Assigning Oxidation Number: A Case Study Of High School Chemistry Textbooks. *Indonesian Journal of Chemistry*, 1(223–236). <https://doi.org/10.22146/ijc.36695>
- Basuki, R., Amanda, H., Bemis, R., Lisma, A., & Yusnadar, Y. (2018). Incomplete explanation in determining oxidation number: A case study on chemistry program students. *Jurnal Pendidikan IPA Indonesia*, 7(3), 333–340.
- Behar-Horenstein, L. S., & Niu, L. (2011). Teaching critical thinking skills in higher education: A review of the literature. *Journal of College Teaching & Learning (TLC)*, 8(2).
- Cano, F., Garcí'a, Á., Berbén, A. B. G., & Justicia, F. (2014). Science learning: A path analysis of its links with reading comprehension, question-asking in class and science achievement. *International Journal of Science Education*, 36(10), 1710–1732.
- Carter, A. G., Creedy, D. K., & Sidebotham, M. (2015). Evaluation of tools used to measure critical thinking development in nursing and midwifery undergraduate students: a systematic review. *Nurse Education Today*, 35(7), 864–874.
- Danczak, S. M., Thompson, C. D., & Overton, T. L. (2020). Development and validation of an instrument to measure undergraduate chemistry students' critical thinking skills. *Chemistry Education Research and Practice*, 21(1), 62–78.
- Dilekli, Y., & Tezci, E. (2016). The relationship among teachers' classroom practices for teaching thinking skills, teachers' self-efficacy towards teaching thinking skills and teachers' teaching styles. *Thinking Skills and Creativity*, 21, 144–151.
- Eliasson, N., Karlsson, K. G., & Sørensen, H. (2017). The role of questions in the science classroom--how girls and boys respond to teachers' questions. *International Journal of Science Education*, 39(4), 433–452.
- Ennis, R. H. (1993). Critical thinking assessment. *Theory into Practice*, 32(3), 179–186.
- Espinosa, A. A., Monterola, S. L. C., & Punzalan, A. E. (2013). Career-oriented performance tasks in chemistry: Effects on students' critical thinking skills. *Education Research International*, 2013.
- Fitriani, H., Asy'ari, M., Zubaidah, S., & Mahanal, S. (2019). Exploring the prospective teachers' critical thinking and critical analysis skills. *Jurnal Pendidikan IPA Indonesia*, 8(3), 379–390.
- Garratt, J., Overton, T., Tomlinson, J., & Clow, D. (2000). Critical thinking exercises for chemists: Are they subject-specific? *Active Learning in Higher Education*, 1(2), 152–167.
- Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist*, 53(4), 449.
- Halpern, D. F. (2010). *Halpern critical thinking assessment*. SCHUHFRIED (Vienna Test System): <https://jurnal.unimus.ac.id/index.php/JPKIMIA/index>

Moedling, Austria.

- Hernández, G. E., Criswell, B. A., Kirk, N. J., Sauder, D. G., & Rushton, G. T. (2014). Pushing for particulate level models of adiabatic and isothermal processes in upper-level chemistry courses: a qualitative study. *Chemistry Education Research and Practice*, 15(3), 354–365.
- Hingorjo, M. R., & Jaleel, F. (2012). Analysis of one-best MCQs: The difficulty index, discrimination index and distractor efficiency. *Journal of the Pakistan Medical Association*, 62(2), 142–147.
- Huber, C. R., & Kuncel, N. R. (2016). Does college teach critical thinking? A meta-analysis. *Review of Educational Research*, 86(2), 431–468.
- Hwang, G.-J., & Chen, C.-H. (2017). Influences of an inquiry-based ubiquitous gaming design on students' learning achievements, motivation, behavioral patterns, and tendency towards critical thinking and problem solving. *British Journal of Educational Technology*, 48(4), 950–971.
- Jacob, C. (2004). Critical thinking in the chemistry classroom and beyond. *Journal of Chemical Education*, 81(8), 1216.
- Johnson, R. B., & Christensen, L. (2019). *Educational research: Quantitative, qualitative, and mixed approaches*. SAGE Publications, Incorporated.
- Kemendikbud. (2017). *Silabus Mata Pelajaran SMA/MA Mata Pelajaran Kimia*. Kemendikbud.
- Khasanah, A. N., Widoretno, S., & Sajidan, S. (2017). Effectiveness of Critical Thinking Indicator-Based Module in Empowering Student's Learning Outcome in Respiratory System Study Material. *Jurnal Pendidikan IPA Indonesia*, 6(1), 120425.
- Kipper, H., & Rüttemann, T. (2010). Strategies and techniques of questioning effectuating thinking and deep understanding in teaching engineering at Estonian centre for engineering pedagogy. *Problems of Education in the 21st Century*, 19, 36.
- Kogut, L. S. (1996). Critical thinking in general chemistry. *Journal of Chemical Education*, 73(3), 218.
- Martensson, P., & Hansson, H. (2018). Challenging teachers' ideas about what students need to learn. *International Journal for Lesson and Learning Studies*.
- McMillan, J. H. (1987). Enhancing college students' critical thinking: A review of studies. *Research in Higher Education*, 26(1), 3–29.
- Moore, B. N. (2001). *Critical thinking*. McGraw-Hill.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). TIMSS 2015 International Results in Mathematics. 2016. Retrieved from Boston College, TIMSS & PIRLS International Study Center Website: <http://Timssandpirls.bc.edu/Timss2015/International-Results>.
- Mutakinati, L., Anwari, I., & Kumano, Y. (2018). Analysis of students' critical thinking skill of middle school through stem education project-based learning. *Jurnal Pendidikan IPA Indonesia*, 7(1), 54–65.
- Oktaviana, I. A., Saputro, A. N. C., & Utami, B. (2016). Upaya Peningkatkan Kemampuan Berpikir Kritis Dan Prestasi Belajar Siswa Melalui Penerapan Model Pembelajaran Problem Based Learning (Pbl) Dilengkapi Modul Pada Materi Kelarutan Dan Hasil Kali Kelarutan Kelas XI SMA Negeri 1 Gondang. *Jurnal Pendidikan Kimia Universitas Sebelas Maret*, 5(1), 143–152.
- Pearson. (2018). *Watson-Glaser III Critical Thinking Appraisal (WG-III), Identify top performers and get critical thinkers into your workplace*. <https://www.assessmentday.co.uk/watson-glaser-critical-thinking.htm>,
- PISA. (2016). *Prograamme for International Student Assessement (PISA) Results From PISA 2015*. OECD.
- Pursitasari, I. D., Suhardi, E., Putra, A. P., & Rachman, I. (2020). Enhancement of Student's Critical Thinking Skill through Science Context-based Inquiry Learning. *Jurnal Pendidikan IPA Indonesia*, 9(1), 97–105.
- Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4(1). <https://doi.org/10.1080/2331186X.2017.1301013>
- Robinson, S. R. (2011). Teaching logic and teaching critical thinking: revisiting McPeck. *Higher Education Research & Development*, 30(3), 275–287.
- Sipayung, D. H., Sani, R. A., Bunawan, H., & others. (2018). Collaborative Inquiry For 4C Skills. *3rd Annual International Seminar on Transformative Education and Educational Leadership (AISTEEL 2018)*.
- Sumarni, W., & Kadarwati, S. (2020). Ethno-Stem Project-Based Learning: Its Impact to Critical and Creative Thinking Skills. *Jurnal Pendidikan IPA Indonesia*, 9(1), 11–21.

- Sunarti, S. R. (2014). Penilaian Dalam Kurikulum 2013. *Yogyakarta: CV Andi Offset*.
- Susanti, L. Y., Hasanah, R., & Khirzin, M. H. (2018). Penerapan Media Pembelajaran Kimia Berbasis Science, Technology, Engineering, and Mathematics (STEM) untuk Meningkatkan Hasil Belajar Siswa SMK/SMA pada Materi Reaksi Redoks. *Jurnal Pendidikan Sains (JPS)*, 6(2), 32–40.
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296.
- The Critical Thinking Co. (2019). *Cornell critical thinking tests*.
<https://www.criticalthinking.com/cornell-critical-thinking-tests.html>
- Towns, M. H. (2014). Guide to developing high-quality, reliable, and valid multiple-choice assessments. *Journal of Chemical Education*, 91(9), 1426–1431. <https://doi.org/10.1021/ed500076x>
- Yuliyanto, E., & Rohaeti, E. (2013). Pengembangan Majalah Kimia Untuk meningkatkan Motivasi Belajar Dan Kreativitas Peserta Didik Kelas X SMA N 1 MLATI. *Jurnal Pendidikan Sains (JPS)*, 1(1), 1–15.
- Yusoff, M. S. B. (2019). ABC of Content Validation and Content Validity Index Calculation. *Education in Medicine Journal*, 11(2), 49–54. <https://doi.org/10.21315/eimj2019.11.2.6>