Development of Critical Thinking Test Instruments with Problem Solving Context on The Salt Hydrolysis Material

Budi Susetyo, Wiwi Siswaningsih*, Fizky Rizki Oktavi

Department of Chemistry Education, Universitas Pendidikan Indonesia, Bandung, Indonesia.

*Email: wiwisiswaningsih@upi.edu

Abstract. The assessment tools must be developed to measure the competencies in 21st century skills, such as critical thinking and problem-solving. Critical thinking is seen as a prerequisite for problem-solving abilities. Therefore, this study aims to produce critical thinking test with problem-solving context on the salt hydrolysis material, also measure students' mastery of the critical thinkings sub-indicators and problem-solving indicators. The research method used the development and validation. Participants involved in the test were 42 students of class XI who had studied salt hydrolysis at a high school in Bandung. The product from this research is critical thinking skills test instrument in the context of problem-solving as many as 10 multiple choice with an open reason on the salt hydrolysis material. Based on results of test quality, the developed test has good and decent criteria. The instruments declared valid based on the content validity at the CVR value and the empirical validity on Pearson's product-moment correlation coefficient, and reliable with Cronbach Alpha value of 0.823. Criteria for empirical validity and reliability test as a whole are included in high and very high. The results analysis quality of the items as a whole have medium difficulty level, very good distinguishing power, and good distractor effectiveness. The results of level student mastery in the critical thinkings sub-indicators and problem-solving indicators that are most mastered are to express problems and mention facts related to the problems. Meanwhile, the least mastered are drawing conclusions according to facts and checking the feasibility of the solutions made.
Keywords: Critical thinking skills, Problem-solving, Salt hydrolysis, Test Development.

Introduction

The competencies of students and graduates that have to be achieved according to 21st century skills identified by US-based Partnership for 21st Century Skills are critical thinking skills, creative thinking skills, communication skills, and collaboration skills (Zubaidah, 2018). These thinking skills that students need to develop are included in higher order thinking skills (HOTS). Moreover, HOTS-oriented learning is a program developed by the Ministry of Education and Culture through the Directorate General of Teachers and Education Personnel (Ditjen GTK) in an effort to improve the quality of learning and the quality of graduates. This program was developed following the policy direction of the Ministry of Education and Culture which in 2018 also has been integrated as Strengthening Character Education and learning oriented towards Higher Order Thinking Skills (HOTS) (Ariyana, et al., 2018).

There are four types of higher-order thinking, there are critical thinking, creative thinking, problem solving, and decision making (Tawil & Liliasari, 2013). Problem solving is closely related to critical thinking. Problem-solving skills require the ability to think

critically in exploring various alternative ways or solutions, as well as providing problematic situations that trigger the development of students' critical thinking potential. This similarity needs to be explored as a basis for developing these two capabilities (Cahyono, 2015). One way to develop and train these skills is to develop instrument test that can measure critical thinking skills and problem solving (Ningsih, et al., 2018).

Assessment is one item that must continue to develop to measure the real conditions of all national students in a valid and reliable (Khotimah, 2019). As constructive alignment suggests, good teaching requires assessment which effectively assesses the outcomes an educator wants students to be able to demonstrate. Traditional methods of assessment, such as summative examinations and the outputs from expository laboratories, such as reports, are limited in their ability to evaluate students' critical thinking skills (Danczak, 2018). Currently, the assessment of critical thinking skills and problem solving was done separately (Sadhu & Laksono, 2018). Thus, an integrated instrument needs to be developed that aim to assess critical thinking skills and problem solving in a single test only. To determine the level of student success in developing critical thinking and problem solving, an evaluation tool is needed that can measure this ability (Kartimi, et al., 2012). The test questions that are made must require answers to the results of critical thinking in problem solving, so it is very good to use in the form of essay, because it can train students to formulate answers from the results of their own thoughts and demand various methods of solving and answers (Rosbiono, 2007). An open reasoned multiple choice test is a multiple choice test accompanied by reasons so that students must write down the reasons for the answers they choose (Suwarto, 2012). Threfore, the open reasoned multiple choice test type is more suitable to be used to measure students' critical thinking skills in the context of problem solving.

Tests to measure critical thinking skills are usually very much needed in science, such as in chemistry subjects (Yuanita & Yuniarita, 2018). Salt hydrolysis is a chemical subject that contains concepts related to everyday events (Dina, et al., 2015) and requires critical thinking skills for problem solving (Nurfitriana, et al., 2018).

In the salt hydrolysis material, students will study the properties of the salt solution, the concept of hydrolysis, and calculate the pH value of the salt solution. The material of salt hydrolysis has a concept that is not enough to memorize, but there are concepts that need to be observed through practicum and discussion in groups. In the hydrolysis material there are also calculations, so students must first understand the concept in order to be able to apply the calculation formula (Arini & Saputro, 2017).

Based on the above problems, this study was conducted to produce a critical thinking skill test instrument with a problem solving context (HOTS) on salt hydrolysis material that meets the eligibility criteria for a test, and is expected to be used as an alternative evaluation tool to determine students' mastery of sub-indicators of critical thinking skills and problem solving indicators.

The development carried out is developing the type of critical thinking skills in the context of problem solving in the salt hydrolysis material. The indicator is developed to become a slice indicator between the sub indicators of critical thinking skills that are appropriate and relevant to the problem solving indicators, so it is expected that the test instrument can measure critical thinking skills and problem solving. The test developed for critical thinking skills in the context of problem solving is in the form of an open reasoned multiple choice test. The item questions developed in this study were based on the critical thinking skills sub-indicators of Ennis 1985, as many as 6 sub indicators were selected and problem solving indicators Mourtos, Okamoto, & Rhee 2004, as many as 7 indicators were selected.

Methods

This research method refers to the development and validation method according to Adams & Wieman in 2010. Broadly, this method consists of: 1) the development stage and 2) the validation and testing phase.

Participants in this study were high school students who had studied salt hydrolysis at a high school in Bandung. Two classes were chosen to be given a critical thinking skill test with the context of the problem solving, in class XI MIPA 3 as many as 20 students as trial 1 and class XI MIPA 4 as many as 22 students as trial 2.

In this study, two research instruments were used, they are the test instrument validation sheet and the test items for critical thinking skills with the context of problem solving on the salt hydrolysis material as many as 10 multiple choice with an open reasons. The test instrument validation sheet is used to obtain an assessment from the validator regarding the validity of the content, in terms of the content of the test instrument being developed. The validators involved were 7 experts in the field of education and/or chemistry, 4 lecturers of chemistry education and 3 chemistry teachers. Meanwhile, the instrument test were used to obtain student score data from the results of the two trials which processed to determine the empirical validity, reliability, difficulty level, differentiation power, and distractor effectiveness.

The data collection technique was carried out at the validation stage and at the trial stage. The data collection technique at the validation stage uses a test instrument validation sheet. Accomplish by a validator of 7 experts in the field of education and/or chemistry to determine the validity of the content of the test instrument. The content validity is determined by the experts to consider the "content" aspect of the test instrument being developed. Calculated using CVR (Content Validity Ratio) (Wilson, et al., 2012).

$$CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}} \tag{1}$$

Remark:
CVR   : content validity ratio
n_e : the number of respondents who stated Yes
N   : number of respondents

The data collection technique at the trial stage uses a test items for critical thinking skills with the context of problem solving on the salt hydrolysis material. The data obtained from the results of trials with students as many as two trials. The data processing technique at this stage aims to measure the feasibility of the test instrument. The data were processed regarding empirical validity, reliability, difficulty level, differentiation power, distractor effectiveness, and analysis mastery level of sub indicator of critical thinking skills and indicator of problem solving.

Internal validity (empirical validity) can be seen from the results of the correlation coefficient between the item scores and the total test scores. The calculation is done using Pearson's Product Moment correlation technique with data processing using the IBM SPSS Statistics 22 program (Arikunto, 2003).

Processing and analysis of the reliability test was carried out with the IBM SPSS Statistic 22 program and then the correlation coefficient of reliability was known, using the Cronbach Alpha method. A test is said to be reliable if it has a Cronbach Alpha value greater than 0.60 (Ghozali, 2009).

The level of difficulty of a test item is determined by dividing the number of items correctly answered by the number of respondents. Therefore, the item difficulty level (p) = the proportion of correct answers. The formula for calculating the level of difficulty is as follows (Susetyo, 2015).

$$P_i = \frac{f_i}{M} \tag{2}$$

Remark:

P_i : the difficulty level of the i th test item
f_i : participants who answered correctly
M   : number of participants

      The differentiation power (D) of test items is the ability of how much a test item can distinguish (discriminate) between high-ability test participants and low-ability test participants. The differentiation power is indicated by a discrimination index number abbreviated as D. The following formula determines the differentiation power (Susetyo, 2015).

$$D = \frac{f_{Ti}}{f_{Ti} + f_{Ri}} \qquad\qquad (3)$$

Remark:
D   : differentiation power
f_{Ti} : the frequency with which the i th test item answered correctly for the high group
f_{Ri} : the frequency which the i th test item answered correctly for the low group

      Distractor analysis aims to find out which options are not functioning properly. distractor analysis is carried out because each distractor must really function as a distractor, in the sense of attracting the attention of students who do not master the subject matter related to the subject of the test (Firman, 2013). Distractor that classified as function properly is minimum has proportion 5%. One way of calculating to check the functioning of a distractor is through the calculation of the percentage proportion using the following formula (Susetyo, 2015).

$$px_i = \frac{fx_i}{M} \times 100\% \qquad\qquad (4)$$

Remark:
pxi   : the proportion of each answer choice
fxi   : the frequency of each choice of answers to a test item
M   : number of respondents

      To determine the percentage value of the critical thinking skills sub indicator mastery and problem solving indicators, can obtain through their scores test. The percentage value is obtained using the following formula (Purwanto, 2006).

$$\text{Percentage} = \frac{R}{SM} \times 100\% \qquad\qquad (5)$$

Remark:
R   : student scores from each indicator
SM   : the maximum score of each indicator

Results and Discussion

      In this developed test instrument, the sub-indicators of critical thinking-Ennis 1985 are used, were (1) express problems, (2) identify/formulate criteria to consider, (3) considering the use of appropriate procedures, (4) choose the criteria for considering possible solutions, (5) involves little prediction, and (6) draw conclusions according to facts.

      Meanwhile, seven indicators of problem solving by Mourtos et al., 2004 are used, were (1) mention facts related to the problem, (2) define a concept or category, (3) check out previous solutions to solve the problem, (4) choose theories, principles, and approaches to solving related problems, (5) estimate the results that will be obtained through the solutions that have been made, (6) Check the feasibility of the solutions created, and (7) determine information/data related to the problem given.

      Compilation of test instruments from predetermined the framework of question test, there were from 10 indicators of slice results from sub indicators of critical thinking skills Ennis 1985 and indicators of problem solving Mourtos, et al., 2004, then made 10 multiple choice questions with open reasons. The questions are arranged according to problem solving steps for each given discourse of the problem. The 10 questions are divided into 3

texts (discourse) on the problems raised, the first text for items number 1 to 5, the second text for items number 6 to 7, and the third text for item number 8 to 10. After the preparation of the instrument, validation was carried out to test the content validity.

The contents validity of the test instrument was obtained from the consideration of 7 validators. 10 items were validated and there were 3 texts also validated. There are four aspects of validation, three aspects for item validation and one aspect for text validation. The aspects that were validated on the items were the suitability of the text with the items, the suitability of the indicators with the items, and the suitability of the answers to the reasons with the rubric. Meanwhile, the validated aspect of the text is the suitability of the accuracy of the text content. Following are the CVR results for each item and text on each aspect of the validation.

Table 1. Results of the validity of the item questions content.

No.	The suitability of the text with the items				The suitability of the indicators with the items				The suitability of the answers to the reasons with the rubric				Avrg. CVR
	Validator that states		CVR value	Rmk.	Validator that states		CVR value	Rmk.	Validator that states		CVR value	Rmk.	
	Yes	No			Yes	No			Yes	No			
1	7	0	1,00	Valid	7	0	1,00	Valid	7	0	1,00	Valid	1,00
2	7	0	1,00	Valid	7	0	1,00	Valid	7	0	1,00	Valid	1,00
3	7	0	1,00	Valid	7	0	1,00	Valid	7	0	1,00	Valid	1,00
4	7	0	1,00	Valid	6	1	0,71	Valid	7	0	1,00	Valid	0,91
5	7	0	1,00	Valid	7	0	1,00	Valid	7	0	1,00	Valid	1,00
6	7	0	1,00	Valid	7	0	1,00	Valid	7	0	1,00	Valid	1,00
7	7	0	1,00	Valid	7	0	1,00	Valid	7	0	1,00	Valid	1,00
8	7	0	1,00	Valid	6	1	0,71	Valid	7	0	1,00	Valid	0,91
9	7	0	1,00	Valid	7	0	1,00	Valid	7	0	1,00	Valid	1,00
10	7	0	1,00	Valid	7	0	1,00	Valid	7	0	1,00	Valid	1,00

Table 2. Results of the validity of text content.

Text	The suitability of the accuracy of the text content			
	Validator that states		CVR value	Rmk.
	Yes	No		
1	7	0	1,00	Valid
2	7	0	1,00	Valid
3	7	0	1,00	Valid

The minimum value of CVR for 7 validators with a significance level of a one-sided test of 0.05, is 0.622. If the CVR value is higher or equal to the minimum CVR value, then the item is valid (accepted), whereas if the CVR value is below the minimum CVR value then the item is invalid (rejected) (Wilson et al., 2012). Based on the results validation of the items and text, all CVR values have a value of >0.622 so that they are declared valid, item and text accepted. Of the 10 validated items, eight items had a CVR = 1 and two items had a CVR = 0.71. Meanwhile, the results of the validation on the text are all those with a CVR = 1. Based on the results of the validity value from the validation test to the

expert, it can be said that all items and text are declared valid (accepted) and can be used for trial 1.

Quality Analysis of Test Instrument from Trial 1

Trial 1 was carried out after the test instrument was revised according to the suggestions contained in the validation results from the validators. The results of the students' scores were then analyzed to determine the quality of the test, there were empirical validity, reliability, difficulty level, differentiation power, and distractor effectiveness. The following table shows the results of test instrument quality of trial 1.

Table 3. Results of test instrument quality from trial 1.

No.	Empirical validity		Reliability	Difficulty level		Differentiation power		Distractor effectiveness
	Value	Criteria		Value	Criteria	Value	Criteria	
1	0,64	High		0,75	Easy	0,63	Good	Bad (1)
2	0,67	High		0,55	Medium	0,80	Very good	Bad (2)
3	0,57	Medium		0,65	Medium	0,83	Very good	Good (4)
4	0,86	Very high		0,55	Medium	1,00	Very good	Bad (1)
5	0,70	High	0,855 Reliable (Very high reliability criteria)	0,65	Medium	0,83	Very good	Good (4)
6	0,62	High		0,65	Medium	0,71	Very good	Good (4)
7	0,49	Medium		0,30	Medium	0,75	Very good	Bad (1)
8	0,89	Very high		0,35	Medium	1,00	Very good	Good (4)
9	0,54	Medium		0,15	Hard	1,00	Very good	Good (4)
10	0,60	Medium		0,50	Medium	0,67	Good	Bad (1)

The empirical validity determined in this study is the internal validity seen from the results of the correlation coefficient between the item score and the total test score, which is intended to determine how far the test being developed can measure what is being measured. The results of the empirical validity analysis in trial 1 showed that all items were declared valid. The analysis was obtained by comparing the Pearson correlation (r count) with r 5% significance table, that is, the items were declared valid if r count> r table (Sudarmanto, 2005). In the first trial followed by 20 test participants, the value of r table of 5% significance was 0,44. Based on empirical validity criteria of Arikunto 2003, there are two items that have very high validity criteria (4 and 8), four items with high validity criteria (1, 2, 5, and 6), and four items with medium criteria (3, 7, 9, and 10). These results mean that all the items can measure what is being measured or declared empirically valid.

The reliability value of the test was calculated using the IBM SPSS Statistics 22 application and using the Cronbach Alpha coefficient, to determine the degree of consistency of the test. The value of the test reliability coefficient in trial 1 is 0.855 which is said to be reliable because it is >0.60 (Ghozali, 2009). Based on the interpretation of the reliability value according to Jacob & Chase 1992, the reliability value obtained in trial 1 is included in the very high reliability criteria.

The level of difficulty analysis is carried out to determine the degree of difficulty of an item. The items are said to be good if they have a balanced (proportional) level of difficulty (Sukardi, 2009). The results of the difficulty level test of ten items, one item included in the easy criteria (1), eight items included in the medium criteria (2, 3, 4, 5, 6, 7, 8, and 10), and one item included in the difficult criteria (9). The average value of the difficulty level on all items from the results of trial 1 was 0.51 which was included in the medium criteria. The proportion of difficulty level of a good test is 30% easy, 40% medium, and 30% difficult (Sukardi, 2009), judging from the proportion of the difficulty level of the items in trial 1, the criteria for the items were said to be quite good.

Diffetention power analysis on test items is carried out to determine how much test items can distinguish between high-skilled test takers and low-ability test takers. (Susetyo, 2015). Before analyzing the differention power of the test instruments, students were grouped into high, medium, and low groups as seen from the results of the multiple choice test scores. The group division of students is 27% of the high group (which has the highest test score), 27% of the low group (which has the lowest test score), and the remaining 46% belongs to the medium group (Susetyo, 2015). The results analysis of the differention power of the items obtained that eight items had excellent differention power (2, 3, 4, 5, 6, 7, 8, and 9) and as many as two items had quite good differention power (1 and 10). Based on the interpretation of differention power criteria of the items according to Susetyo 2015, the average differention power of all test items was 0.82 which included in the very good criteria.

Analysis of the distractor or confounder in multiple choices is to find out whether all possible answer choices have been selected by the test taker, a distractor can be said to be good if minimal of 5% is chosen by the test taker (Susetyo, 2015). The results of the distractor effectiveness analysis from a total of 40 distractors, there were 34 distractors are included in the good criteria and 6 distractors including the bad criteria contained in items 1, 2, 4, 7, and 10, so a revision must be carried out then the revision results are used for trial 2.

Quality Analysis of Test Instrument from Trial 2

Trial 2 was carried out after the instrument was repaired, it was on items that had bad criteria or were less than the results of data of trial 1. The corrected items were item number 1, 2, 4, 7, and 10 which based on the results of the distractor effectiveness analysis there were bad distractors. The following is a table that shows the results of the test quality data of test instrument in trial 2.

The results of the empirical validity analysis in trial 2 showed that all items were declared valid. The test participants were 22 students, the value of r table of 5% significance was 0.43. Based on the validity criteria according to Arikunto 2003, there is one item with very high criteria (4), three items with high criteria (1, 6, and 8), and six items with medium criteria (2, 3, 5, 7, 9, and 10). The average of all items was declared valid with high validity criteria. These results mean that all the items can measure what is being measured or declared empirically valid.

The reliability value of the second trial test obtained a Cronbach Alpha coefficient of 0.823 which is said to be reliable because it is >0.60 (Ghozali, 2009). Based on the interpretation of the reliability value according to Jacob & Chase 1992, the reliability value obtained is included in the very high reliability criteria.

Table 4. Results of test instrument quality from trial 2.

No.	Empirical validity		Reliability	Difficulty level		Differentiation power		Dictractor effectiveness
	Value	Criteria		Value	Criteria	Value	Criteria	
1	0,66	High		0,77	Easy	0,60	Good	Good (4)
2	0,57	Medium		0,14	Hard	0,33	Little revision or not	Good (4)
3	0,56	Medium	0,823 Reliable (Very high reliability criteria)	0,68	Medium	0,56	Good	Good (4)
4	0,85	Very high		0,32	Medium	0,83	Very good	Good (4)
5	0,45	Medium		0,68	Medium	0,86	Very good	Good (4)
6	0,73	High		0,41	Medium	1,00	Very good	Good (4)
7	0,46	Medium		0,23	Hard	0,67	Good	Good (4)
8	0,78	High		0,45	Medium	1,00	Very good	Good (4)
9	0,49	Medium		0,27	Medium	1,00	Very good	Good (4)
10	0,57	Medium		0,36	Medium	1,00	Very good	Good (4)

The results of analysis the difficulty level of ten items, one item included in the easy criteria, it was item number 1, seven items included in the medium criteria, it was items number 3, 4, 5, 6, 8, 9, and 10, and two. The items are included in the difficult criteria, it was items number 2 and 7. The average value of the level of difficulty on all items from the results of trial 2 was 0.43 which was included in the medium criteria. The proportion of difficulty level of a good test is 30% easy, 40% medium, and 30% difficult (Sukardi, 2009), judging from the proportion of the difficulty level of the items in trial 2, the criteria for the items were said to be quite good.

The results of analysis the differentiation power of the items obtained six items had very good criteria (4, 5, 6, 8, 9, and 10), three questions had good criteria (1, 3, and 7), and one question had revised criteria little or not (2). The differentiation power for item number 2 is 0.33, it can be said that it is included in the good criteria, because the differentiation power of the item is adequate if it has a D value ≥ 0.25 (Firman, 2013). The average differentiation power on all test items was 0.78 which was included in the very good criteria, in other words, the developed critical thinking skills test in the context of problem solving was able to distinguish the abilities of high-group students and low-group students well.

The results of the distractor effectiveness analysis of trial 2, all distractors are included in the good criteria. Therefore, it can be said that the test of critical thinking skills with problem solving has a criterion for the effectiveness of a well-functioning item distractor, in the sense of attracting the attention of students who do not master the subject matter related to the subject of the test.

Analysis Mastery Level of Sub Indicator of Critical Thinking Skills and Indicator of Problem Solving

Analysis the level mastery of the indicators was carried out from the results of trial 2 because trial 2 was an application trial. In this study, the item indicators were obtained from the results of the slice between the sub indicators of critical thinking skills Ennis 1985 with the indicators of problem solving Mourtos et al., 2004. The following table shows the sub indicators critical thinking and problem solving indicators used for item slice indicators

Table 5. The pair of sub indicator critical thinking and problem solving indicators for slice indicators of items.

CTS sub indicators	Problem solving indicators	Items
Express the problems [1]	State facts related to the problem [1]	1
Identify/formulate criteria to consider [2]	Define a concept or category [2]	2 and 6
	Determine information/data related to the problem given [7]	8
Consider using proper procedures [3]	Checking the previous solutions to solve the related problem [3]	3
Choose criteria for considering possible solutions [4]	Choose theories, principles, and approaches to solving related problems [4]	4 and 9
Involves a little predictions [5]	Estimating the results that will be obtained through the solutions that have been made [5]	5
Draw conclusions according to facts [6]	Checking the feasibility according to the solutions made [6]	7

Of the 7 pairs selected critical thinking skills (CTS) sub indicators & problem solving indicators (1 and 1, 2 and 2, 2 and 7, 3 and 3, 4 and 4, 5 and 5, and 6 and 6), a slice indicator of 10 item indicators was made. Because the item indicators are slice indicators between the CTS sub-indicators and the problem solving indicators, it is possible to measure the level of mastery of the CTS sub indicators and problem solving indicators. The following graph shows the level of student mastery of each pair indicators.
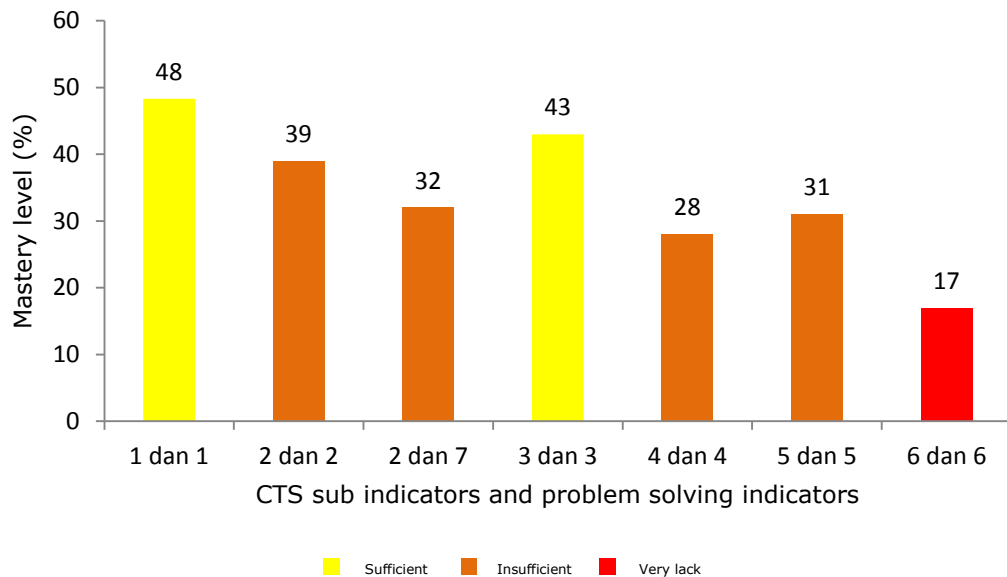

Figure 1. Graph of mastery level (%) sub indicator of critical thinking skills and indicator of problem solving for all students on the trial 2.

Based on the graph, the level of mastery in the sufficient criteria according to Riduwan (2009), there are only two CTS sub indicators and two problem solving indicators. The two CTS sub indicators and the two problem solving indicators are the CTS 1 sub-indicator (revealing problems), the CTS 3 sub-indicator (considering the use of proper procedures), problem solving 1 indicator (stating facts related to the problem), and

problem solving 3 indicator (checking solutions that have been done to resolve related problems).

The level of mastery in the insufficient criteria, there are as many as three CTS sub indicators and four problem solving indicators. There are the CTS sub-indicators: 2, 4, and 5, the problem solving indicators: 4, 5, and 7. The mastery level of the CTS sub-indicators and problem solving indicators for all students in trial 2 in very lack criteria is found in the CTS 6 sub-indicator (drawing conclusions according to facts) and problem solving indicator 6 (checking the feasibility of the solutions made). This indicator is found in item 7. The problem solving aspect of item number 7 is the fifth step, is evaluating, if associated with critical thinking steps in problem solving according to Facione, this indicator is included in the step of researching carefully (fifth step) (Facione, 2015). This item requires in-depth calculation and analysis, and if it is seen from the level of difficulty, this item is included in the difficult criteria. In these items students are asked to determine the sodium benzoate content of a beverage brand A so that a conclusion can be drawn whether the drink is safe for consumption or not.

From the results of analysis the level of critical thinking skills and problem solving on all students in trial 2, Obtained an average mastery of the CTS sub-indicator and problem solving indicators by 34% which are included in the poor criteria according to Riduwan 2009 of students mastery level criteria. This achievement is thought to be due to the students not having a complete understanding of the material on salt hydrolysis. Another factor that is the cause is the situation when doing tests online using the Google Form platform, so filling in the reason column for items that require calculation steps or writing down reaction equations becomes difficult. This can cause the student's score to be not optimal.

Conclusion

The findings of this study ascertained that the developed integrated assessment for the critical thinking instrument test with the context of problem solving on the salt hydrolysis material has relatively high validity and reliability. The results of the content validity test on each item were declared valid and the average value of empirical validity was valid under high criteria. The reliability value of the test was 0.823 which declared reliable with very high criteria. The instrument test has the proper and good test quality criteria seen from the results of difficulty level, differentiation power, and effectiveness distractor. Thus, the researchers believe that the instrument test could be used to assess critical thinking skills and problem solving of the salt hydrolysis material on high school students. The mastery of the critical thinking sub-indicator most mastered by students in the salt hydrolysis material is to reveal problems, and problem solving indicator is mentioning facts related to problems. Meanwhile, the sub-indicator of critical thinking skills that are least mastered is drawing conclusions according to facts, and problem solving indicator is checking the feasibility of the solutions made.

Accknowledgment

References

Adams, W.K. & Wieman, C.E. 2010. Development and validation of instrumens to measure learning of expert-like thinking. *International Journal of Science Education*, 33(9):1–24. doi: 10.1080/09500693.2010.512369.

Arikunto, S. 2003. *Dasar-dasar evaluasi pendidikan*, Bumi Aksara, Jakarta.

Arini, H.S. & Saputro, S. 2017. Upaya peningkatan kemampuan berpikir kritis dan prestasi belajar siswa dengan menggunakan model pembelajaran learning cycle 5e pada materi pokok hidrolisis garam kelas XI MIA 1 SMA negeri 1 Banyudono semester genap tahun pelajaran 2015/2016. *Jurnal Pendidikan Kimia*, 6(2):161–170.

Ariyana, Y., Pudjiastuti, A., Bestary, R., & Zamroni. 2018. *Buku pegangan pembelajaran berorientasi pada keterampilan berpikir tingkat tinggi*, Direktorat Jenderal Guru dan Tenaga Kependidikan, Kementerian Pendidikan dan Kebudayaan, Jakarta.

Cahyono, B. 2015. Korelasi pemecahan masalah dan indikator berfikir kritis. *Jurnal Pendidikan MIPA*, 5(1):15–24.

Danczak, S.M. 2018. Development and validation of an instrument to measure undergraduate chemistry students' critical thinking skills, *M.S. Thesis*, Monash University, Melbourne Australia.

Dina, Setiabudi, A., & Nahadi. 2015. Pembelajaran berbasis masalah untuk meningkatkan keterampilan berargumentasi siswa SMA pada konsep hidrolisis garam. *Jurnal Pendidikan Matematika dan Sains Tahun III*, 3(2):133–141.

Ennis, R.H. 1985. A logical basis for measuring critical thinking skills. *Educational Leadership*, 43(2):44–48.

Facione, P.A. 2015. *Critical thinking: What it is and why it Counts*, Measured Reasons LLC and distributed by Insight Assessment, Hermosa Beach-California.

Firman, H. 2013. *Evaluasi pembelajaran kimia*, Jurusan Pendidikan Kimia FPMIPA UPI, Bandung.

Ghozali, I. 2009. *Aplikasi analisis multivariate dengan program SPSS*, Badan Penerbit Universitas Diponegoro, Semarang.

Kartimi, Liliasari, & Permanasari, A. 2012. Pengembangan alat ukur berpikir kritis pada konsep senyawa hidrokarbon untuk siswa SMA di Kabupaten Kuningan. *Jurnal Pendidikan MIPA*, 13(1):18–25.

Khotimah, K. 2019. Pengembangan evaluasi pembelajaran berbasis higher order thinking skill di sekolah dasar. *Proceeding Seminar Nasional Pendidikan FKIP UNMUL I*, Samarinda, 24 Agustus 2019.

Jacob, L.C. & Chase, C.I. 1992. *Development and using test effectively*, Jossey-bass Publisher, San Fransisco.

Mourtos, N.J., Okamoto, D., & Rhee, J. 2004. Defining, teaching, and assessing problem solving skills. *Prosiding UICEE annual conference on engineering education*, Mumbai India, 9-13 Februari.

Ningsih, D.R., Ramalis, T.R., & Purwana, U. 2018. Pengembangan tes keterampilan berpikir kritis berdasarkan analisis teori respon butir. *Jurnal Wahana Pendidikan Fisika*, 3(2):45–50.

Nurfitriana, R.A., Fuldiaratman, & Yusnaidar. 2018. *Analisis keterlaksanaan model pembelajaran problem solving dan pengaruhnya terhadap keterampilan berpikir kritis siswa pada materi hidrolisis garam di kelas XI MIPA SMA Negeri Titian Teras H. Abdurahhman Sayoeti*, Program Studi Pendidikan Kimia, FKIP, Universitas Jambi, Jambi.

Purwanto, N. 2006. *Prinsip-prinsip dan teknik evaluasi pengajaran*, Remaja Rosdakarya, Bandung.

Riduwan. 2009. *Belajar mudah penelitian untuk guru-karyawan dan peneliti pemula*, Alfabeta, Bandung.

Rosbiono, M. 2007. *Teori problem solving untuk sains*, Direktorat Jenderal Peningkatan Mutu Pendidik dan Tenaga Kependidikan Depdiknas, Jakarta.

Sadhu, S. & Laksono, E. W. 2018. Development and validation of an integrated assessment for measuring critical thinking and chemical literacy in chemical equilibrium. *International Journal of Instruction*, 11(3):557–572.

Sudarmanto, R.G. 2005. *Analisis regresi linier ganda dengan SPSS*, (edisi pertama), Graha Ilmu, Yogyakarta.

Sukardi. 2009. *Evaluasi pendidikan prinsip dan operasionalnya cetakan ketiga*, Bumi Aksara, Jakarta.

Susetyo, B. 2015. *Prosedur penyusunan & analisis tes*, PT. Refika Aditama, Bandung.

Suwarto. 2012. *Pengembangan tes diagnostik dalam pembelajaran*, Graha Ilmu, Yogyakarta.

Tawil, M. & Liliasari. 2013. *Berpikir kompleks dan implementasinya dalam pembelajaran IPA*, Badan Penerbit Universitas Negeri Makasar, Makasar.

Wilson, F.R., Pan, W., & Schumsky, D.A. 2012. Recalculation of the critical value for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development*, 45(3):197–210.

Yuanita & Yuniarita, F. 2018. Pengembangan petunjuk praktikum IPA berbasis keterampilan proses untuk meningkatkan keterampilan berpikir kritis siswa sekolah dasar. *Jurnal Profesi Pendidikan Dasar*, 5(2):139–146. doi: https://doi.org/10.23917/ppd.v1i2.6608.

Zubaidah, S. 2018. Mengenal 4c: Learning and innovation skills untuk menghadapi era revolusi industri. *2nd Science Education National Conference*. Universitas Trunojoyo, Madura.