

The Development and Validation of Science Virtual Test to Assess 7th Grade Students' Critical Thinking on Matter and Heat Topic

Yustika Sya'bandari^{1*}, Harry Firman², Lilit Rusyati¹

¹International Program on Science Education, Faculty of Mathematics and Science Education, Universitas Pendidikan Indonesia

²Department of Chemistry Education, Faculty of Mathematics and Science Education, Universitas Pendidikan Indonesia

*Corresponding Author. yustika.sya'bandari@student.upi.edu

ABSTRACT An efficient way to improve the quality of education in critical thinking is developing the better tests. The test has been shifted towards the use of computer-based procedures. Lack of specific topic of critical thinking tests in science and the advanced of technology made the researcher intended to develop and validate a test to measure students' critical thinking in the matter and heat topic based on computer for seventh grade junior high school. The method that used in this research was descriptive. Generally, the process of developing and validating the test consists of 5 steps: (1) content analysis; (2) constructing multiple choice items; (3) readability test and expert validation; (4) limited tryout; and (5) larger application. Based on larger application, it is obtained the reliability value, difficulty level, discriminating power and distractor quality. The subject was 117 students of public Junior High School in Kabupaten Bandung. The instrument validation resulted 30 items that represent 8 elements and 21 sub-elements to measure students' critical thinking based on Inch in matter and heat topic. The instrument is further called as Science Virtual Test Matter and Heat (SVT-MH). The alpha Cronbach (α) is 0.642 which means that the instrument is sufficient to measure students' critical thinking matter and heat topic. The result also shows the profile of students in critical thinking is 73% categorized as 'moderate', while students respond the positive impression towards the use of SVT-MH.

Keywords Critical thinking, matter and heat, reliability, science virtual test, students' impression, validity.

1. INTRODUCTION

In the 21st century, there have been shifts in the way we live as well as in how we get and evaluate information. These shifts mean we need to look critically at critical thinking itself and determine what approaches are most effective to meet the challenges and opportunities (Anne & Kreitzberg, 2010). Critical thinking is the skill to prevent people from making bad decisions and helps to solve the problems (Inch, Warnick & Endres, 2006). Critical thinking makes us think not just about the world around us but also about the thought process itself (Harpen, 2003). Consequently, teaching critical thinking skills is important for students, as they need to adjust to such change by actively and skillfully conceptualizing, applying, analyzing, synthesizing, evaluating the information gathered from, or generated by, observation, experience, reflection, reasoning, or communication (Paul & Elder, 2004).

Teaching critical thinking at schools is the main topics in the discussion regarding 21st Century skills (Greenhill, 2009). Students can be taught to critically examine different viewpoints on issues concerning the impact of science and technology on everyday life and evaluate these issues

through critical thinking (Mapeala & Siew, 2015). Teaching critical thinking in science and technology also helps to develop students' analytical skills, as well as their ability to make informed choices in their everyday lives (Mapeala & Siew, 2015). According to (UNESCO, 2000) measuring the improvements of critical thinking skills will essentially improve the education quality. If tests are understood to shape both the curriculum and teaching, an efficient way to improve the quality of education in critical thinking is developing the better tests (Yeh, 2001). Tiruneh et al. (2016) stated that critical thinking tests which have been developed such as the Cornell Critical Thinking Test–Level Z (CCTT), the California Critical Thinking Skills Test (CCTST), the Ennis-Weir Critical Thinking Essay test, the Watson-Glaser Critical Thinking Appraisal, and the Halpern Critical Thinking Assessment (HCTA) were mostly general content-based and reviewed based on the following criteria: (a) is the test based on a clear

Received: 28 October 2017

Revised: 29 November 2017

Published: 30 November 2017

definition/conception of critical thinking, (b) are the targeted critical thinking skills common across tests, and (c) do the test items appear to sufficiently measure the critical thinking skills targeted on a test. All the general content of critical thinking tests are diverse in terms of their formats, scope, and psychometric characteristics.

Recognizing the lack of specific content of critical thinking tests, it is required to develop critical thinking test in specific subject area. The accompanying expectation has been embedding critical thinking skills within a subject matter instruction in various specifics content will facilitate the acquisition of critical thinking skills that are applicable to a wide variety of thinking tasks within the content in question and that it will facilitate their transfer to other problems in everyday life (Lawson, 2004). The present study addressed this concern by developing and validating a critical thinking test based on Inch Critical Thinking in science content for junior high school student.

The higher order thinking and complex information in the critical thinking test can be supported by the advanced technology in its development. The assessment also has been shifted towards the use of computer-based procedures (Jurecka 2008). The Computer Based Test (CBT) is a test that the presentation and selection are computerized which can simulate real-world problems (ill-structured and complex in nature) (Rusyati & Firman, 2017). Computer based-test enables teachers and researchers to collect different types of data. Hence, researchers have the possibility to design meaningful and motivating real-life scenarios, where students can solve complex and interactive problems (Greiff et al. 2013). The information in the computer based-test can be presented not only using picture, the item which is presented also can use the interactive video. Even in the field of science education, there are different approaches to use computers in order to facilitate the work on complex problems (Jonassen, 2004). Computer based test is effective solution for education evaluation (Flagbola, Adigun & Oke, 2013). In addition, to traditional scores on multiple-choice or constructed-response items, data on the time needed to perform a number of interactions and the sequence of operations are accessible (Wirth, 2008). The major advantage of computer-based tests is in the assessment of new content areas and constructs (Scherer, Koppelt & Tiemann, 2014). The use of computer based assessment is advantageous due to test economics, improvements in objectivity, and test reliability (Scherer, Koppelt & Tiemann, 2014).

The ministry of Education and Culture of Indonesia also has launched the National Examination based on computer in 2014 in SMP Indonesia Singapura dan SMP Indonesia Kuala Lumpur (SIKL). Ujian Nasional Berbasis Komputer (UNBK) also called Computer Based Test (CBT) is a national examination using the computer as a media. In the implementation, UNBK is different with

paper national examination or Paper Based Test (PBT) which has been running out (Kemendikbud, 2016). It means, developing a test should follow the current technology in order to have the better benefits.

According to Agah, Ogbeche & Okorie (2016), there was a positive and direct relationship between computer anxiety, operation skills, attitude and students' preparedness for computer-based assessment. By using this developed test, students can practice before they conduct the national examination to prevent the influences of their computer anxiety, operation skills and attitude in using computer.

The author intended to develop the instrument in the topic of matter and heat because in 2013 curriculum, this topic integrates each other which is connected from basic competence 3.3 and 3.4. Based on Rahayu & Kita (2009) students have great difficulties with and hold some alternative conceptions of the concepts of matter (homogeneous mixtures, phase changes from solid to liquid, and phase changes from solid to gas). So, matter and heat topic is also important to be presented virtually because the content which is abstract for students is easier to be presented in the form of animation or interactive video.

Thus the present study, first, proposes to develop and validate test to assess 7th grade students' critical thinking based on Inch critical thinking on the topic of matter and heat. Author defines critical proficiency as the ability to reasonably respond to critical thinking tasks that do not necessarily require specific content knowledge, but rather application of content knowledge of everyday life. The topic which is included is based on the basic competence of 2013 curriculum, matter and heat topic that consists of mixture, element, compound, physical and chemical properties, physical and chemical change, state of matter and its change, temperature, expansion, heat and heat transfer sub-topic. Second, the design and characteristics of the assessment tool which use the computer based test are described. The design of critical thinking test using the computer based-test will be in the form of "Science Virtual Test" which not only can assess students' critical thinking but also can used as preparation for students' before conducting another computer based examination.

2. METHOD

The research method that used in this research was descriptive research. According to McMillan and Schumacher (2001) descriptive research is concerned with the current or past status of something and simply describe achievements, attitudes, behaviors or other characteristics of a group of subject. Gall, Gall and Borg (2003) stated that descriptive research is a research that involves making careful descriptions of educational phenomena.

Descriptive research involves gathering data that describe events as organizes, tabulates, illustrate and

describe data collection. Therefore, the research method in this research is descriptive because appropriate with the objectives of the research which is describing the development and validation of critical thinking instrument based on Inch to describe 7th grade students' critical thinking on matter and heat topic.

The location of this research was Public Junior High School "X" in *Kabupaten* Bandung. The school uses *Bahasa Indonesia* in teaching learning process. 2013 Curriculum is implemented for 7th and 8th grade, while 9th grade is still using KTSP Curriculum. This school appropriate because the instrument intended for 7th grade which uses 2013 curriculum. This school has 540 students of 7th grade, 551 students of 8th grade and 533 students of 9th grade in 2015.

This school was chosen because it has been accredited "A" by Indonesia Ministry of Education and also supported by multimedia facilities e.g. personal computer. The subject was students from three classes of 7th grade students who have learnt about matter and heat topic. The average age is about 12-13 years old. For the descriptive research, the subject is better more than 100 students. In this research, researcher involved 117 seventh grade junior high school student consist of 60 female students and 57 male students.

Critical thinking was measured by Science Virtual Test that has been developed based on Inch elements of critical thinking. The elements are purpose, question at issue, assumption, point of view, information, concepts, interpretation and inference, implication and consequences. These eight elements also developed to be 26 sub elements. These elements can lead student the students to the level of critical thinking. The science virtual test (SVT) is the test that use computer as media. It was developed using Macromedia Flash Professional 8.0. The test is multiple choice questions with four option of critical thinking instrument. The content validity was measured by expert judgment through rubric assessment and students responds that eventually was analyzed through IBM Statistic 23 and ANATES 4.1.0.

There are two types of instrument used in this research. There are rubric and questionnaire. **First**, the rubric was used to validate the content of the instrument. The validation was divided into content, media and education which have its own aspects to be assessed. **Second**, the questionnaire was used to know students' impression after using the science virtual test. It consists of four aspects which are experience, technical, preference and media. Students' impression uses the Likert scale analysis. In Likert, subject reads every statement in the questionnaire and evaluates the question based on the categorized answer (Suherman, 2003). There are 4 answers that can be chosen by students and it has its own score. For the favorable statement (positive statement), the strongly agree has 5 score, agree has 4 score, disagree has 2 score, strongly disagree has 1 score, and neutral score is 3.

3. RESULT AND DISCUSSION

The result and discussion founded in this research is divided into 5 parts: (1) The development of Science Virtual Test, (2) The validity of Science Virtual Test, (3) The reliability of Science Virtual Test, (4) The characteristics of Science Virtual Test, (5) The profile of students' critical thinking and students' impression toward SVT.

3.1 The Development of Science Virtual Test

The importance of developing measurement for students to think critically especially on specific subject of science continues to grow. Researchers need to have valid and reliable tests to evaluate the effectiveness of various instructional efforts. In this study, the author has developed the instrument to measure students' critical thinking on the matter and heat topic which implements the computer based test. This instrument is further called "Science Virtual Test Matter and Heat or SVT-MH". The development of science virtual test to measure students' critical thinking on matter and heat topic consists of 5 steps: (1) content analysis of 26 elements Inch critical thinking framework and the matter and heat topic; (2) constructing the multiple choice items; (3) readability test by 3 junior high school students and 2 science teacher, validity test by 2 education expert, 2 content expert and 2 media expert; (4) conducting limited tryout by 40 students of junior high school; (5) conducting a larger test by 117 students of junior high school.

After the content was analyzed, the items were constructed, it resulted the preliminary design of science virtual test. The preliminary design tested its readability, validity and reliability (limited and larger test) result the final design of SVT-MH.

3.1.1 The Preliminary Design of Science Virtual Test Matter and Heat

SVT-MH consists of 41 items based on two basic competences of 2013 Curriculum: (1) explain the concept of mixture and pure substance (element and compound), physical and chemical properties, physical and chemical changes in daily life; (2) analyze the concept of temperature, expansion, heat, heat transfer and its implementation in daily life including the mechanism in maintaining the stability of human and animal body temperature. These two basic competences interpreted into the items that represent 8 elements and 26 sub-elements of Inch critical thinking. The story board developed into the virtual test and result the preliminary design of SVT-MH.

The topic of matter and heat consists of eight sub-topic which is matter (element, compound, mixture), physical and chemical properties, physical and chemical change, state of matter and its change, temperature, expansion and heat transfer. The researcher intended to develop 40 items

Table 1 The blueprint of SVT-MH

Critical Thinking Element	Matter	Physical and Chemical Properties	Physical and Chemical Change	State of Matter and Its Change	Temperature	Expansion	Heat and Heat Transfer	Item's Number	
								Number of Item	Percentage
Purpose (4 sub-element)	1,5		3	6	2,4		8	7	23%
Question at Issue (3 sub-element)		15	10,13	14	9			5	17%
Assumption (2 sub-element)					17	16		2	6.5%
Point of View (2 sub-element)			19		18			2	6.5%
Information (3 sub-element)	24,27	21,26		22				5	17%
Concepts (3 sub-element)	29,31			32		28	30	5	17%
Interpretation and Inference (2 sub-element)		34	37					2	6.5%
Implication and Consequences (2 sub-element)		41			40			2	6.5%
Total	6	5	5	4	6	2	2	30	
Percentage	20%	17%	17%	13%	20%	6.5%	6.5%		100

where every topic has 5 items representation (14%). In constructing the item, the researcher collaborated the topic with 26 Inch critical thinking sub-elements. It is resulted 41 item which consist of 7 items of matter (17%), 6 items of physical and chemical properties (15%), 8 items of physical and chemical change (19%), 4 items of state of matter and its change (10%), 8 items of temperature (19%), 3 items of expansion (7%) and 5 items of heat transfer (12%).

3.1.2 The Final Design of Science Virtual Test Matter and Heat

SVT-MH has been revised from its preliminary design resulted the final design of the instrument that can measure students' critical thinking on matter and heat topic. The item number 7, 11, 12, 20, 23, 25, 33, 35, 36, 38 and 40 were deleted. The blueprint of the items SVT-MH can be seen in Table 1.

The final design and its item analysis specifically is in the form of item card, consists of 30 items that represent 8 elements and 21 sub-elements of Inch critical thinking. In this research, author also find several constraints that occurs in developing and validating the SVT-MH such as: (1) difficulties in selecting the video, image, comic, articles that provide the relevant information with Inch critical thinking framework and the matter and heat topic; (2) in making the instrument, it's hard to proportionally

distribute the items to the 26 sub-elements of Inch critical thinking and 8 sub-topic of matter and heat, and the result of expert validation will possibly resulted disproportional distribution of the items; (3) the low spec of computer which is used that made the program hang and loading while the SVT-MH is installed; (4) the earphone is not available in the computer laboratory that make the video can't be open in each computer because the music will disturb other students.

3.2 The Validity of Science Virtual Test

Before the instrument conducted its validity test, it was tested its readability by three 7th grade students and two science teachers of public junior high school. There are 5 aspects which should be assessed by students and teachers: (1) The description of the article/figure/comic/video/table/graph was easily to comprehend, (2) The question was easily to comprehend, (3) The option was easily to comprehend, (4) There was correlations between the question and the answer, (5) There was no ambiguous word/term which made student find difficulties in comprehending the question.

Most of the items are easily to understand. The recommendation of readability test on multiple choice questions of SVT-MH are: (1) The word "paraffin wax" elusive; (2) Students can assume all the options are right, make the words "tujuan utama" in bold or italic; (3) For

JHS students, there should be a common name for the chemical gas. Example: CO₂ (Carbon dioxide); (4) There is mistyping in several words.

The validation of SVT-MH used content validity by experts. The items judged by 2 experts of education to validate the critical thinking element, 2 experts of content to validate the matter and heat topic, and 2 experts of media to validate the virtual test. The analyses of the expert judgment result used the average congruency percentage by Popham (1978). The item will be valid if the percentage is 90% or higher. Based on the expert validation, there are 30 items that have been valid. The item number 7, 11, 12, 20, 23, 25, 33, 35, 36, 38 and 40 were deleted. The 30 items cover 8 elements and 21 sub-element of Inch critical thinking. The difficulty level, discriminating power and distractor of SVT-MH are analyzed using ANATES version 4.1.0.

3.2.1 Discriminating Power of SVT-MH

The discriminating power is a measurement to compare the number of people with high test scores who answered that item correctly with the number of people with low scores who answered the same item correctly. The higher the discriminating power, the better the item because such a value indicates that the item discriminates in favor of the upper group, which should get more items correct (Hetzl, 1997).

The item that categorized into poor discriminating power are the items number 2, 4, 7, 8, 9, 12, 15, 24 satisfactory discriminating power are the items number 1, 3, 11, 16, 17, 18, 20, 21, 26, 27, 29, and good discriminating power are the items number 5, 6, 10, 13, 14, 19, 22, 23, 25, 28, 30. From the result, the item has a good quality to differentiate between high and low achiever with 22 satisfactory and good discriminating power or 73% from the whole items. Additional validation studies that involve a larger and diverse group of respondents representing the target population should be conducted to further strengthen the quantitative data set and related measures (Tiruneh et al., 2016).

3.2.2 Difficulty Level of SVT-MH

Difficulty level expresses the proportion or percentage of students who answered the item correctly. It one of the most useful, and most frequently reported, item analysis statistics (Testing, 2006). The item which categorized into very easy level are item number 26, 30, easy level are items number 1, 3, 4, 13, 15, 16, 19, 20, 24, 27, 29, moderate level are items number 2, 5, 6, 7, 8, 9, 11, 14, 17, 18, 21, 22, 23, 25, 28, difficult level is item number 10 and very difficult is item number 12. The total number of very easy level is 2, easy level is 11, moderate level is 15, difficult level is 1 and very difficult level is 1.

Table 2 Reliability result of SVT-MH

N of Student	N of Item	Cronbach's Alpha
40	30	0.638
117	30	0.642

It appears that most of the SVT-MH items were rather moderate to easy to our test participants. About 43% of the items were found to have easy level. 2 items included into difficult and very difficult which are item number 10 and 12. Therefore, from the researcher analyses these two items included to the question at issue element and this critical thinking element had lowest achievement from students' result. Students are confused to the form of the questions where they have to raise the question based on the information presented.

3.2.3 Distractor of SVT-MH

One important element in the quality of a multiple choice item is the quality of the item's distractors. A distractor analysis addresses the performance of these incorrect response options. Distractor is measured and analyzed to know whether it is plausible or not. Distractors used in the items need to be plausible so that they attract at least 5% students (Chavda, Misra & Duttaroy, 2015).

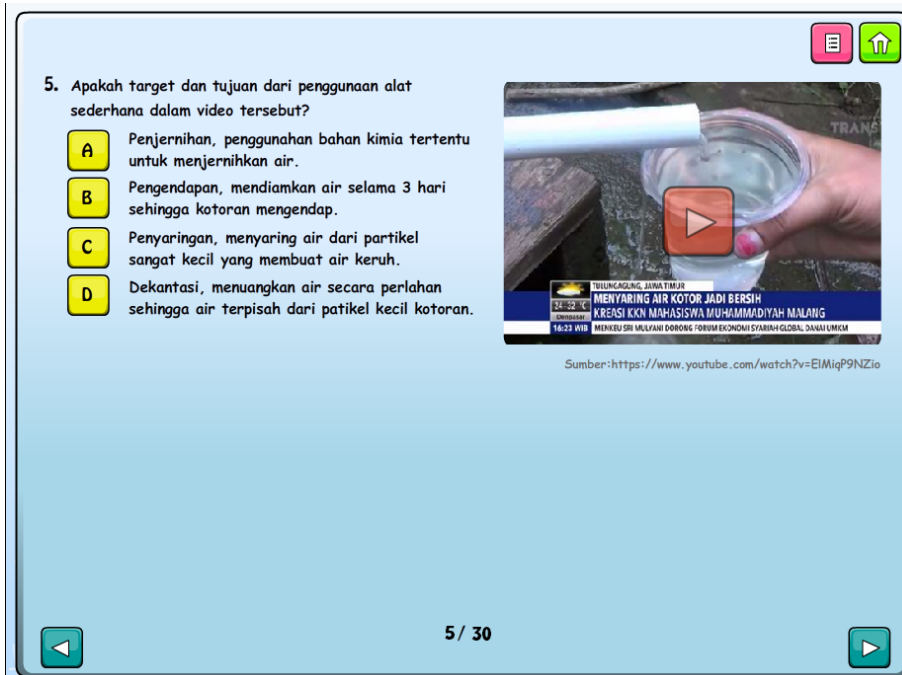
There are 120 distractors from 30 items. In this research, the distractors are computed by ANATES Version 4.1.0. The result shows that there are 18 distractors which are not usable because it is less than 5% chosen by total respondent. There should also be the revision for the usable distractor because some of them are included to the poor distractor. It has bigger respondent rather than its answer key such as the item number 10. 39 students chose distractor D, while the correct answer is distractor C which chosen by 32 students. Most of students assumed that the answer key is D. So, we have to recheck the correct answer key or revise another distractors. In general, the distractor that should be revised is 30% from the total distractors.

3.3 The Reliability of Science Virtual Test

After SVT-MH is revised based on readability and validity test, the test continue to analyze its reliability by a group of student. The respondents were 7th grade students of public junior high school that consists of 40 students. The data obtained analyzed using statistical analyses by IBM SPSS Statistics Version 23.

The result showed in the Table 2 that reliability of the test of 40 students $\alpha = 0.638$. It is common to see the journal articles where one or more scale alphas which the range was 0.60 - 0.69 (Leech, Barrett & Morgan, 2005). It interpreted as 'acceptable'. So, the instrument is reliable to measure students' critical thinking on heat and matter topic.

SVT-MH that has been revised tested to the 7th grade public junior high school student. The respondent consists of 117 students. The Table 2 shows that the Cronbach's



Number of Item	5
Critical Thinking Element	Purpose
Critical Thinking Sub-Element	Stating objectives on target
Topic	Mixture Separation
Basic Competence	<i>Menjelaskan konsep campuran dan zat tunggal (unsur dan senyawa), sifat fisika dan kimia, perubahan fisika dan kimia dalam kehidupan sehari-hari</i>

Difficulty Level		Discriminating Power	
0.58		0.44	
Distractor			
A (12.8%)	B (6.9%)	**C (58.1%)	D (22.2%)

**Answer Key

Figure 1 Item Card Represent Video as the Information

Alpha 117 students, $\alpha = 0.642$. The value is sufficient because it is common to see the journal articles where alpha were in the 0.60 – 0.69 range (Leech, Barrett & Morgan, 2005). The interpretation of reliability coefficient of the test categorized as ‘acceptable’ to measure students’ critical thinking on heat and matter topic.

The most commonly used type of internal consistency reliability is Cronbach’s coefficient alpha. This measure indicates the consistency of a multiple item scale (Leech, Barrett & Morgan, 2005). It means if the instrument is used for the next measurement, the result of the measurement will be quite the same with the previous result.

However, it has to be noted that the coefficient alpha was not as large as expected. The relatively low alpha value can be explained by at least two factors:

The SVT-MH was intended to elicit students’ ability in eight elements of Inch critical thinking. But according Tiruneh et al. (2016) it is possible that cognitive processes required to respond to the items were multifaceted. For

instance, a student who performed well in an item that focuses on purpose element, may not have done well on a different item that focuses on question at issue element as these two components slightly vary terms of the required cognitive process.

The lower coefficient alpha may have to do with the composition of the test group and number of participants. Responding accurately and consistently to the SVT-MH items require students who have an adequate mastery of the matter and heat content. The present test takers clearly show the low average score and may have possibly responded to some of the randomly item. This may have influenced the internal consistency of the test.

3.4 The Characteristics of Science Virtual Test

The characteristics of science virtual test found after the SVT-MH was constructed. The item is in the form of multiple choice items and the information presented in each items is in the form of video, article, comic, figure or

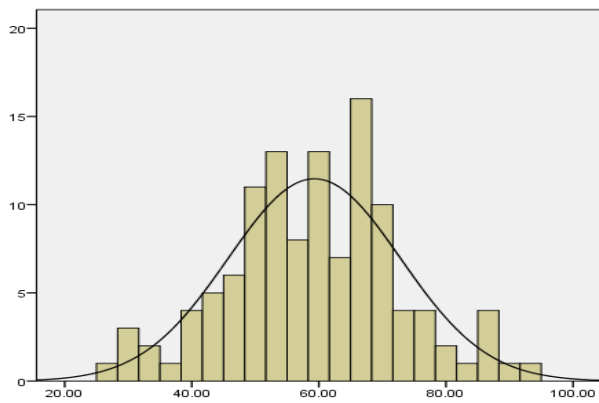


Figure 2 Frequency Distribution of Students' Critical Thinking

graph. The information also related with daily activity. This instrument not only change the media form paper based test to computer based test but also change virtually the information to be interactive and the concept is applied in daily life.

The examples of the item card that represents the format test with the video as information presented can be seen in Figure 1. It shows that difficulty level of the item no.5 is categorized in 'moderate' level (0.58). The discriminating power is categorized into 'good' discriminating power (0.44). The item no.5 can differentiate the high and low achiever of the students. The distractor A is chosen by 12.8% students, B is chosen by 6.9% students and D is chosen by 22.2% students. It means that all distractor are usable.

Analyzing the item card, science Virtual Test Mater and Heat (SVT-MH) resulted several features of SVT-MH which make it different with other instrument.

SVT-MH is the instrument to measure students' critical thinking skill which is developed in specific subject topic.

The topic is based on the two basic competence of 2013 curriculum that integrates each other. The connected topic is matter and heat. Similar with the research (Tirunch et al., 2016) which develops and validates the measurement of critical thinking skills in specific science domains. This research targeted content of an introductory Electricity and Magnetism course and resulted the design of CTEM (Critical Thinking Electricity and Magnetism) test;

The items are in the form of the multiple choices. Multiple choice items are more objective rather than essay to analyze the profile of students because every single item of critical thinking sub-element has the exact answer. According to Weimer (2015) multiple choice items are quick and easy to score by hand or electronically, can be written so that they test a wide range of higher-order thinking skills, can cover lots of content areas on a single exam and still be answered in a class period. Similarly with the research Mapeala and Siew (2015) which has developed the multiple choice items measuring critical thinking ability among fifth graders in primary science The other commonly administered general critical thinking tests that used the multiple choice item format are Cornell Critical Thinking Test–Level Z (CCTT), the California Critical Thinking Skills Test (CCTST), the Watson-Glaser Critical Thinking Appraisal (Tirunch et al., 2016).

SVT-MH uses the computer based test which follow the advanced of technology nowadays. The utilization of technology in educational assessment is aimed at the effectiveness and efficiency of the implementation of the test (Chee and Wong, 2003). SVT-MH also can be used by students to practice their computer operational skill and anxiety for computer preparedness before they conduct the national examination that has been implemented computer based test. In the research of Agah, Ogbeche & Okorie (2016) there was a positive and direct relationship between

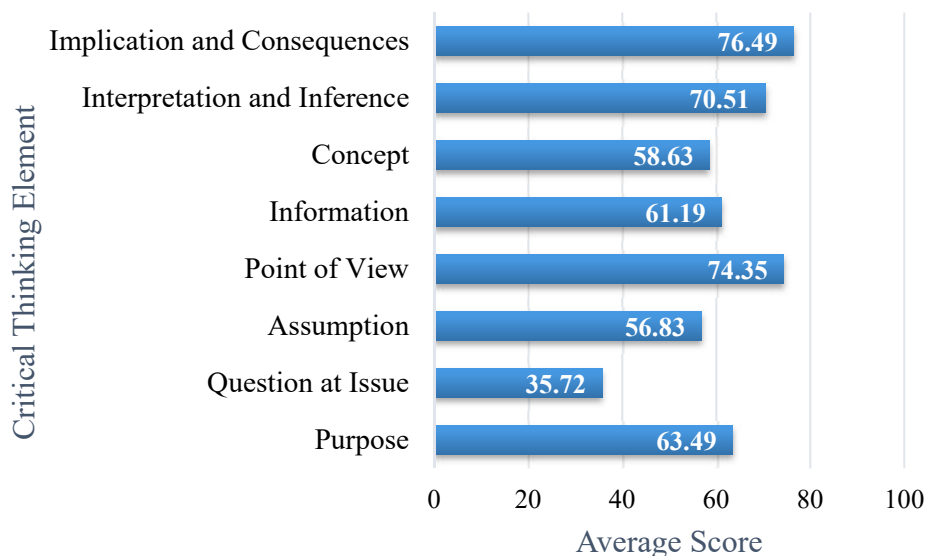


Figure 3 The Average Score of Each Element

Table 3 Descriptive Statistics Test of Students' Profile

	N	Mean	Range	Std. Deviation
Value	117	59.29	66.67	13.57

Table 4 Students' Critical Thinking Level

Level	Number of Student	Percentage
Low	17	14%
Moderate	85	73%
High	15	13%

computer anxiety, operation skills, attitude and students' preparedness for computer-based assessment. So, students need to practice in using computer based-assessment;

SVT-MH is paper less and easy to use. Paperless make it effective and efficient in economic especially for the test with higher number of test taker. This benefit have been corroborated by Scherer, Koppelt & Tiemann (2014), who stated that computer based-test is advantageous due to test economics, improvements in objectivity, and test reliability.

The information presented in the item of SVT-MH can be in the form of interesting animation, video with music compared with paper based test. It is due to the complexity of the critical thinking problem, so the virtual test (computer based-test) is needed in developing the instrument. The score also can be seen directly by students after they finish the test. Similar with the research of Scherer, Koppelt & Tiemann (2014) which stated that computer-based assessments is needed in instructional development due to the complexity of complex problem-

solving processes in Chemistry, multidimensionality of the construct could be assumed.

3.5 The Profile of Students' Critical Thinking and Students' Impression Toward SVT-MH

3.5.1 Students' Critical Thinking on Matter and Heat Topic

SVT-MH was used to measure students' critical thinking on mater and heat topic. The data collected have been analyzed using IBM SPSS Statistics 23. It computed the descriptive statistics test that resulted in the Table 3. The result shows that the respondent consists of 117 students. The average score of students' critical thinking is 59.29 in the scale 0-100. The standard deviation is 13.57. The range is 66.67 and other frequency distribution can be seen in the Figure 2. It shows that data has normal distribution. The minimum score achieved by student is 26.67 while the maximum score achieved by student is 93.33. Most students have a score 66.67 which are 16 students and the middle score of students is 60. From the 117 students, the authors categorized students' achievement in critical thinking into low, moderate and high level based on its mean and standard deviation. Table 4 shows that 14% students are categorized into low level of critical thinking, 13% students are categorized into high level and most of students are categorized into moderate level of critical thinking which is 73%.

The author calculated the students' score based on the eight elements of Inch critical thinking. The data based on 117 students can be seen in the Figure 3. It shows that the

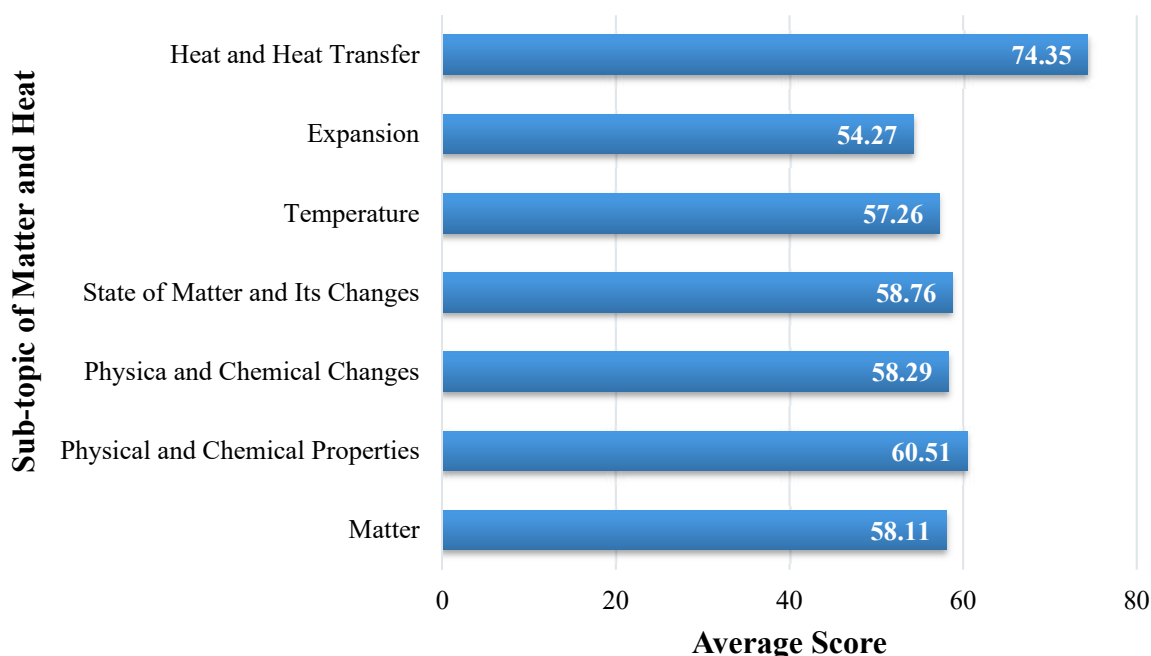
**Figure 4** The average score of each topic

Table 5 Students impression after using SVT-MH

Aspect	Description	Average Value (1-5)
Experience	Students' opinion after experiencing science virtual test.	4.02 (+)
Technical	Students' impression in accessing science virtual test technically.	4.10 (+)
Preference	Students' preference between paper-based test and science virtual test.	4.17 (+)
Media	Students' impression to the appearance of science virtual test in term of position and size of text, background, information (picture, video, table, graph), and navigation button.	3.98 (+)

highest average score of the element is implication and consequence which is 76.50. While, the lowest average scores of the element is question at issue which is 35.73. Students are rare to get the questions to train their critical thinking. From 8 elements, the highest average score is 76.50 for implication and consequences element. It means that most of students can reason and think which carries with potential outcome from the process. While, the lowest average scores of the element is question at issue which is 35.73. Most of the students can't identify an issue or problem that needs to be addressed.

Besides calculating the average score based on the eight element of Inch critical thinking, the authors also calculated the score based on the sub-topic of matter and heat. The result shown in the Figure 4, it shows the achievement of students in each topic. The highest average score is the topic of heat and heat transfer which is 74.36. While, the lowest average score is the topic of expansion which is 54.27. The matter and heat topic of the instrument consists of 7 subtopics that author analyzed its each score. The Figure 4 shows that students' achievement in each topic is almost the same. But, there is the high score which is 74.36 in the topic heat transfer. Most of students already

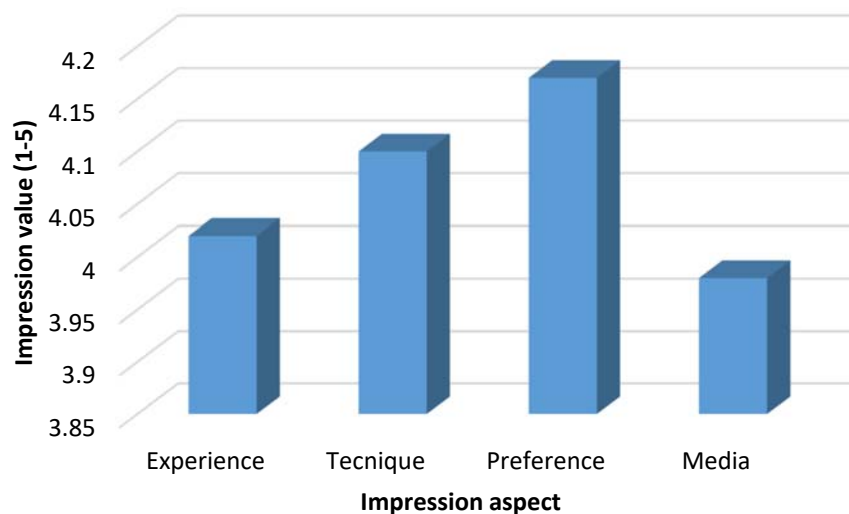
mastered this topic is very daily life topic that the example is easy to find by students.

3.5.2 Students' Impression After Using Science Virtual Test

The students have been given the questionnaire to know their impression towards critical thinking. It consists of four aspects which are experience, technical, preference and media. Each aspect has several statements that should be chosen by student through indicator strongly agree (5), agree (4), disagree (2) and strongly disagree (1).

Table 5 and Figure 5 show the students' impression in average value 1-5. The result shows that the value of every aspect is more than neutral score which means that it has positive (+) impression towards SVT-MH. The highest impression is in preference and the lowest impression is in media aspect.

Students' impression after using science virtual test is used to know students' opinion in experiencing SVT-MH, in using SVT-MH technically, the preference and the evaluation to the media. The result shows that the students' impression is good (positive) to science virtual test because

**Figure 5** Students' Impression after Using SVT-MH

the average value for experience, technical, preference and media are more than 3 (neutral score) from the average range 1-5. The preference shows the highest value which is 4.17. It means that students prefer to use science virtual test rather than paper based test. The research is similar with Joshua, Joshua & Ikiroma (2015) which stated that the level of readiness for using computer based-test is high and acceptability is moderate, with a relatively higher preference by students' federal government owned and privately owned schools.

4. CONCLUSION

Research in developing and validating the Science Virtual Test in the topic of matter and heat has been done systematically. It produced 30 items that can assess 7th grade students' critical thinking on matter and heat topic which is further called 'Science Virtual Test Matter and Heat (SVT-MH)'. Based on the result and analyses, it is acquired some conclusions as follows.

1. The development of science virtual test to measure students' critical thinking on matter and heat topic consists of 5 steps: (1) content analysis of 26 elements Inch critical thinking framework and the matter and heat topic; (2) constructing the multiple choice items; (3) readability test by 3 junior high school students and 2 science teacher, validity test by 2 education expert, 2 content expert and 2 media expert; (4) conducting limited tryout by 40 students of junior high school; (5) conducting a larger test by 117 students of junior high school.
2. In the content validity by experts that analyzed using the average congruency percentage, it is produced 30 valid items that covers eight elements and 21 sub-element of Inch critical thinking.
3. Based on the larger test with 117 junior high school students, it is acquired the reliability of the test where the coefficient alpha (α) is 0.642. It means that the instrument has sufficient reliability to measure students' critical thinking on matter and heat topic.
4. The items of SVT-MH are in the form of multiple choice item which has specific characteristic: (1) the information presented not only in the form of figure, article, graph but also in form of video and animation, (2) The information is related with daily activity, (3) The information of some items is in the form of interesting comic.
5. The profile of students' critical thinking in public junior high school is 73% moderate level. The element that mastered by students is implication and consequences, while the element that less mastered by students is question and issue. The topic that mastered by students is heat transfer, while the topic that less mastered by students is expansion. Most of students also respond the positive impression to the SVT-MH in all aspects (experience, technical, preference and media).

REFERENCES

- Agah, D. J., Ogbache, A. T. & Okorie, E. U. (2016). Computer Anxiety, Operation Skills and Attitude as Correlate of Students Preparedness for Computer Based Assessment. *International Journal of Education and Research*, 4(2), 71-74.
- Anne & Kreitzberg, C. (2010). *Critical Thinking for The Twenty-First Century: What It Is and Why It Matters to You*. [Online]. Retrieved from www.agilecriticalthinking.com. [Accessed on November 15, 2016].
- Chavda, P., Misra, S. & Duttaroy, B. (2015). Item Analysis of Multiple Choice Questions based Undergraduate Assessment in Community Medicine. *South East Asian Journal of Medical Education*, 9(1), 66-69.
- Chee, T. S. & Wong, A. F. (2003). *Teaching and Learning with Technology*. Singapore: Prentice Hall.
- Flagbola, T. M., Adigun, A. A. & Oke, A. O. (2013). Computer-Based Test (CBT) System for University Academic Enterprise Examination. *International Journal of Scientific & Technology Research*, 2(8), 1.
- Gall, M., Gall, J. P. & Borg, W. R. (2003). *Educational Research*. Boston, MA: Pearson Education.
- Greenhill. (2009). *P21 Framework Definitions Document*. [Online]. Retrieved from www.21stcenturyskills.org [Accessed on October 20, 2016]
- Greiff, S, Holt, DV, Wüstenberg, S, Goldhammer, F, & Funke, J. (2013). Computer based Assessment of Complex Problem Solving: Concept, Implementation, and Application. *Educational Technology Research & Development*, 61, 407-421.
- Harpen. (2003). *Thought and Knowledge: An Introduction to Critical Thinking*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hetzl, S. M. (1997). Basic Concepts in Item and Test Analysis. *Clearinghouse on Assessment and Evaluation*, 1-24.
- Inch, E. S., Warnick, B. & Endres, D. (2006). *Critical Thinking and Communication The Use of Reason in Argument*. Boston, MA: Pearson.
- Jonassen, DH. (2004). *Learning to Solve Problems: An Instructional Design Guide*. San Francisco: John Wiley & Sons.
- Joshua, M., T., Joshua, A., M. & Ikiroma, B. (2015). Computer-Based Testing in Nigeria's University Entrants' Matriculation Examination: Readiness and Acceptability Of Critical Stake-Holders, 1-10.
- Jurecka, A. (2008). Introduction to the Computer-Based Assessment of Competencies. *International Journal Hartig, E Klieme, & D Leutner (Eds.), Assessment of Competencies in Educational Contexts (193–214)*. Cambridge/Göttingen: Hogrefe & Huber Publishers
- Kemendikbud. (2016). *Kementerian Pendidikan dan Kebudayaan: Ujian Nasional Berbasis Komputer (UNBK)*. [Online]. Retrieved from http://unbk.kemdikbud.go.id. [Accessed October 20, 2016].
- Lawson, A. (2004). The Nature and Development of Scientific Reasoning: A Synthetic View. *International Journal of Science and Mathematics Education*, 2(3), 307–338.
- Leech, Barrett & Morgan. (2005). *SPSS for Intermediate Statistics: Use and Interpretation*. Mahwah, NJ: Lawrence Erlbaum Associates
- Mapeala, R. & Siew, N. M. (2015). The Development and Validation of a Test of Science Critical Thinking for Fifth Graders. *Spinger Plus*, 4(741), 1-13.
- McMillan, J. H. & Schumacher, S. (2001). *Research in Education A Conceptual Introduction*. New York, NY: Priscilla McGeehon.
- Paul, R. & Elder, L. (2004). *The Nature and Functions of Critical and Creative Thinking*. Dillon Beach: The Foundation for Critical Thinking.
- Popham, W.J. (1978). *Criterion-Referenced Measurement*. Englewood Cliffs, NJ: Prentice Hall.
- Rahayu, S. & Kita, M. (2009). An Analysis of Indonesian and Japanese Students' Understandings of Macroscopic and Submicroscopic Levels of Representing Matter and Its Changes. *International Journal of Science and Mathematics Education*. 8(4), 667–688.

- Rusyati, L & Firman, H. (2017). Validation of Science Virtual Test to Assess 8th Grade Students' Critical Thinking on Living Things and Environmental Sustainability Theme. Bandung: *American Institute of Physics*, 1-7.
- Scherer, R., Koppelt, J. M & Tiemann, R. (2014). Developing A Computer-based Assessment of Complex Problem Solving in Chemistry. *International Journal of STEM Education*, 1(2), 1-6.
- Suherman, E. (2003). *Evaluasi Pemelajaan Matematika*. Bandung: FPMIPA UPI
- Testing, P. (2006). *Step 9. Conduct the Item Analysis*. [Online]. Retrieved from <http://www.proftesting.com>. [Accessed on July 9, 2017]
- Tiruneh, D. T., Cock, M. D., Weldeslassie, A. G., Elen, J. & Janssen, R. (2016). Measuring Critical Thinking in Physics: Development and Validation of a Critical Thinking Test in Electricity and Magnetism. *International Journal of Science and Mathematics Education*, 1-20.
- UNESCO. (2000). *Dakar Framework for Action. Education for All: Meeting Our Collective Commitments. In Adopted by the world education forum, Dakar*. [Online]. Retrieved from <http://unesdoc.unesco.org>. [Accessed on October 23, 2016]
- Weimer, M. (2015). *Different Type of Test Question*. [Online]. Retrieved from <http://www.facultyfocus.com/>. [Accessed on July 7, 2017]
- Wirth, J. (2008). Computer-Based Tests: Alternatives for Test and Item Design. *International Journal Hartig, E Klieme, & D Leutner (Eds.), Assessment of Competencies in Educational Contexts (235–252)*. Cambridge/Göttingen: Hogrefe & Huber Publishers.
- Yeh, S. S., (2001). Tests worth teaching to: constructing state-mandated tests that emphasize critical thinking. *SAGE Journals*, Vol 30(9), 12-17.