

Developing Science Virtual Test to Measure Students' Critical Thinking on Living Things and Environmental Sustainability Theme

M N Akbar, H Firman and L Rusyati

International Program on Science Education-Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi No. 229, Bandung 40154, Indonesia
Email : muhamad.nur@student.upi.edu

Abstract. Critical thinking is skill and ability to use of risk taking creativity to make a decision and knowledge as a result, analysis and synthesis that, evaluation, to acquire, information search, to develop thinking, as an individual aware of his or her own thinking. The aim of this study is to develop the science virtual test to measure students' critical thinking on living things and environmental sustainability theme. The research method that is used in this research was descriptive research. The development of science virtual test item consist of five steps: (1) content analysis; (2) constructing the instrument (multiple choice) based on the elements of critical thinking by Inch; (3) validity judgment of the instrument by the expert; (4) legibility test of the instrument; (5) conducting the large field test. On the large field test was gained the results of validity and reliability of the test, difficulty index, discriminating power, and quality of distractor. The subjects of research were 8th grade students at International Junior High School in Bandung with 125 total of respondents. The coefficient alpha (α) was 0.747, the reliability of the test was categorized as 'high' and value of RXY correlation was 0.63 which mean that the validity of the test was categorized as 'high'. These means that science virtual test can be used to measure student's critical thinking with a good consistency. It is expected for other researcher to take this description as one of the basic information to be considered in developing science virtual test for improving students' critical thinking by various kind of topic.

1. Introduction

Modern technology offers educators a variety of new tools that can be used in the classroom. Technology can help teachers track and assess their students' as well as their own performance in the classroom. This research introduces a virtual test to measure the critical thinking of the students. Critical thinking is the most important skill for problem solving, inquiry and discovery. The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused on inquiry, and persistent in seeking results which are as precise as the subject and the circumstances the inquiry permit¹.

Various examination methods are used in higher education institutions to assess the academic progress, for example paper-pencil-based examinations, assignments, and presentations². More than fifty varied techniques used within higher education for assessment purposes; the most commonly method used are examinations³. The rapid advancement of Information and Communication Technologies (ICT)



in teaching and learning has shifted the paradigm from paper-pencil-based to computer-based system of examinations which are usually termed as Computer Assisted Testing, Computer Based Testing (CBT), Computer Based Assessment (CBA), Online Assessment, E-Assessment and Web-Based assessment. This test uses computer based test method known as CBT⁴.

Nowadays knowledge always develop, that's why the test should be developed because one of the test purpose is to develop students' achievement, ability and skill. The item tests should be selected because it has the highest fidelity with the target domain of task representing an important ability, such as reading, writing, or mathematical problem solving and it should be developed in consistently with the most rigorous guidelines⁵. The quality of test questions or items can be measured with the help of "Item Analysis Process". The major concepts related to item analysis including validity, reliability, item difficulty, and item discrimination⁶. As item development is a major step in test development, validity can be greatly affected by a sound, comprehensive effort to develop and validate test items⁵. The test item should be fieldly tested, and test takers should be appropriately prepared, then they should know what is the expected of item. Item analysis uses statistics and expert judgment to evaluate tests based on the quality of individual items, item sets, and entire sets of items, as well as the relationship of each item to other items⁶. Therefore, the quality of test will be increased.

2. Research methodology

The research method that used in this research was descriptive research. According to best, descriptive research is concerned with how what is or what exists is related to some preceding event that has influenced or affected a present s condition or event⁷. Furthermore, the descriptive studies describe a given state of affairs as fully and carefully as possible⁸. In this method the research focus to develop inch critical thinking multiple choice question from 8th grade on living things and environmental sustainability. It was consist of 26 items with 4 options each number.

The subjects in this research were 8th grade students from International Junior High School in Bandung. The sample was drawn from 125 respondents of the 8th grade students from three schools of International Junior High School in Bandung. The sampling technique that used for this research was a purposive sampling. For purposive sampling the sample has been chosen for a specific purpose⁷. The consideration was because the researcher choose sample based on the requirement such as the school implemented the curriculum 2013 or student has been learned about the concept that will be tested, have a good multimedia facilities e.g. personal computer, and located in Bandung, due to limited time and resources.

3. Result and discussion

As the result of the research there are three part that has been founded: (1) The general characteristic of the virtual test to measure students' critical thinking, (2) Legibility of virtual test by the students, and (3) Test item analysis such as validity, reliability, difficulty level, discriminating power, and distractor.

3.1 Critical thinking test

The designing of science virtual test to measure students' critical thinking on living things and environmental sustainability consist of three steps: (1) content analysis; (2) constructing the instrument (multiple choice test); (3) validity judgment of the instrument by expert. Before constructing the test item, content analysis was needed to determine the content that used to measure students' critical thinking on Junior High School. The theme of this research about the living things and environmental sustainability in 8th grade that limited by Core Competency and Basic Competence based on 2013 Curriculum. Then, there were five major topic that appropriate with the theme used such as structure and function of plant, reproduction system, developing of citizen and the impact, sun radiation, and climate change.

The instrument that has been constructed are in multiple choice form. It was constructed to measure the students' critical thinking based on the elements of critical thinking by Inch. The detailing the eight elements of critical thinking is a function that is interconnected, as follows: purpose, question at issue,

assumptions, point of view, information, concepts, interpretation and inference, implication and consequences⁹. It contains 26 sub elements, which Paul and Elder used as a guide for student in the critical thinking¹⁶. Then, author adjust it for constructing the instrument of science virtual test based on that criteria. In order to test the validity of the instrument the questions that have been constructed then judged by the 3 experts of education, 3 experts of content, and 3 experts of media. The science virtual test made by Adobe Flash 9.0. The example appearance of science virtual test can be seen on Figure 1.

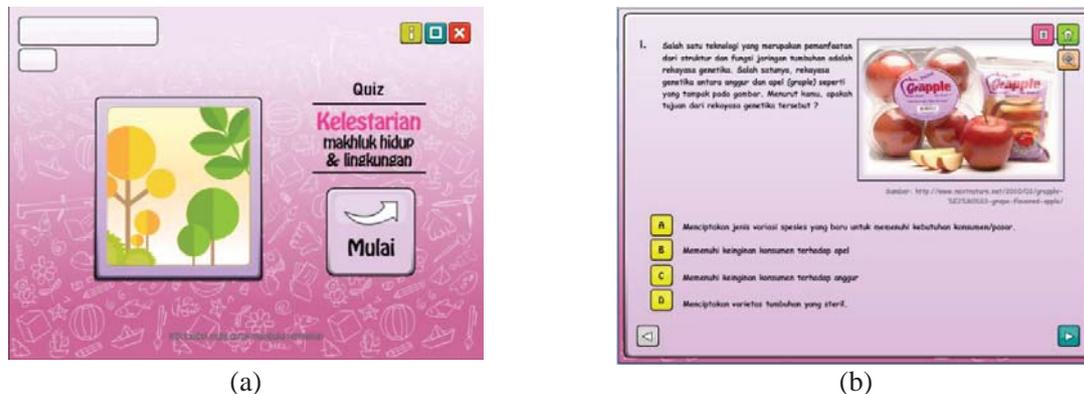


Figure 1. Appearance of Science Virtual Test. (a) Opening and (b) Test Item.

3.2. Legibility test of science virtual Test

Legibility is ease and accuracy of reading of written or printed material¹⁰. The test item that has been made based on the sub-element of critical thinking by Inch, then the researcher do the legibility test of science virtual test. This test was done by five students of 8th grade in private junior high school. There are five criteria that has been test such as: (1) the description of article / picture / comic / video / tables easily understandable, (2) question easily understandable, (3) the option of the answer easily understandable, (4) questions and answers are interrelated, and (5) there are no words or terms that are difficult to understand or give rise to ambiguous meanings.

3.3. Test item analysis

The data obtained from the test were used in the statistical analysis. The analysis of the result was using SPSS Program Version 16 and ANATES Program Version 4.1.0 with the result show the reliability and validity of the test, item difficulty, item discrimination, and distractor analysis.

Table 1. Result of the reliability by SPSS 16.

Cronbach's Alpha	N of Items
.747	26

Based on the SPSS Program Version 16 the result of the reliability of the test $\alpha = 0.747$. Interpretation reliability coefficient of the test was categorized as 'high'¹¹. Then, the result of the validity (using ANATES version 4.1.0) correlation of $R_{xy} = 0.63$ means that the validity of the instrument was categorized as 'high' according to interpretation value of R_{xy} (correlation)¹¹. For item difficulty, item discrimination, and distractor analysis of each item have been measured by the ANATES version 4.1.0 analysis. The result of analysis per each item are:

3.3.1 *Item difficulty (difficulty index)*. The indices of item difficulty level of each item are presented in Table 2. Test items were classified into three categories in terms of level of difficulty¹² as indicated in the Table 3.

Table 2. Item difficulty of comprehensive test items.

Item No	Difficulty index (p)	Item No	Difficulty index (p)	Item No	Difficulty index (p)
1	0.824	10	0.784	19	0.696
2	0.768	11	0.744	20	0.784
3	0.464	12	0.616	21	0.320
4	0.808	13	0.472	22	0.832
5	0.496	14	0.728	23	0.600
6	0.368	15	0.688	24	0.816
7	0.256	16	0.544	25	0.720
8	0.168	17	0.888	26	0.440
9	0.680	18	0.576		

Table 3. Classification of difficulty level.

Value of Difficulty Index (P)	Classification
0.00 – 0.30	Difficult
0.31 – 0.70	Moderate
0.71 – 1.00	Easy

3.3.2 *Item discrimination*. The indices of item discrimination of each item are presented in Table 4. The items were classified accordingly to their discrimination index¹² on Table 5. If the value of the discriminating power is negative, the item test should be rejected¹².

Table 4. Item discrimination of comprehensive test items.

Item No	DP	Item No	DP	Item No	DP
1	0.260	10	0.176	19	0.588
2	0.529	11	0.411	20	0.353
3	0.441	12	0.206	21	0.206
4	0.294	13	0.558	22	0.353
5	0.353	14	0.500	23	0.588
6	0.176	15	0.323	24	0.617
7	0.530	16	0.529	25	0.382
8	0.530	17	0.206	26	0.500
9	0.265	18	0.706		

Table 5. The Classification of Discriminating Power (DP).

Discriminating Power (DP)	Interpretation
0.00 – 0.20	Poor
0.21 – 0.40	Satisfactory
0.41 – 0.70	Good
0.71 – 1.00	Excellent

3.3.3 Distractor analysis

In the section on item analysis, the choice of distractors is critically important. Studies have shown that it is rare to find items for which more than three or four distractors operate efficiently. The distractor is usable if the distractor was chosen by 5% of all the respondent¹². Result analysis shown in the Table 6.

Table 6. Distractor analysis of comprehensive test items.

Item Number	Option			
	A	B	C	D
1	Answer	1	0	21
2	8	7	13	Answer
3	Answer	55	10	1
4	Answer	4	9	11
5	5	45	Answer	11
6	Answer	25	2	32
7	66	13	Answer	14
8	Answer	22	65	17
9	5	6	27	Answer
10	Answer	9	2	16
11	Answer	18	4	10
12	29	Answer	8	9
13	Answer	29	24	12
14	20	3	Answer	5
15	6	Answer	10	22
16	13	39	Answer	4
17	8	Answer	1	3
18	5	45	Answer	11
19	Answer	30	4	3
20	10	Answer	9	8
21	Answer	63	3	19
22	3	15	Answer	2
23	13	Answer	29	8
24	Answer	6	2	15
25	Answer	14	13	8
26	Answer	28	32	7

Science virtual test was constructed to measure students' critical thinking on living things and environmental sustainability. It is a new way to test the level of student's critical thinking by Inch especially for 8th grade students in junior high school. Advantages of virtual science test are: (1) it was in the form of multiple choice, in contrast to the type of test used to measure students' critical thinking is commonly in essay form. With this test, measurements made are more objective than the essay form and it easy to know the profile or level of the students' critical thinking based on the students answer. Multiple choice questions are quick and easy to score, it is can be written so that they test a wide range of higher order thinking skills, and cover lots of content on a single test¹³; (2) compared to the base paper test, science virtual test can use multimedia which is cannot used in paper test, such as video, animation,

and music. In this science virtual test is provided the interesting video that given an information that give more visualization to student, then student can see the phenomena clearly; (3) the result of students answer directly shown in the last session of test, so student will know the score that students gained; (4) paper less and it is very simple to use.

The item analysis conducted in this study implicates three statistics to assists in analyzing the effectiveness of each of the comprehensive test questions specifically validity and reliability of test, item difficulty, item discrimination and distractor analysis. The statistical analysis was done by SPSS version 16 and ANATES version 4.1.0 in the following:

- Reliability and validity of the test, based on the SPSS version 16 the number of Cronbach's alpha or coefficient alpha was used to determine the reliability of the test. If alpha is high, then we might assume that the reliability of the test is acceptable. Reliability coefficient was employed to evaluate the performance of the test as a whole. According to the result of the coefficient alpha (α) is 0.747. One should attempt to generate a KR20 reliability coefficient of .70 and above to acquire reliable score⁸. So, the reliability of the 26 question of the test was categorized as 'high', means that the question has a good consistency and accuracy to measure students' critical thinking. The number of the validity of whole of the test based on the ANATES version 4.1.0 from value of XY correlation is 0.63 which mean that the overall question or item was valid with significant result that categorized as 'high'. Beside that result, the validity that used is from the expert judgment with the result that all of the question can be used by some revision.
- Difficulty index, difficulty index in one of important aspect to determine that the instrument of test is feasible or not. The difficulty index is important because it helps to know how well students have acquired the knowledge they are supposed to¹⁴. Difficulty index of this test is resulted from the analysis by ANATES version 4.1.0. Test items were classified into three categories in terms of level of difficulty as indicated in the Table 3 such as easy, moderate, and difficult. Based on the result the item number of test which are classified as easy level are number 1, 2, 4, 10, 11, 14, 17, 20, 22, 24, 25. Item number which classified as moderate level are 3, 5, 6, 9, 12, 13, 15, 16, 18, 19, 21, 23, 26 and classified as difficult are 7 and 8. The total number which classified as easy 11 items, moderate 13 items, and difficult 2 items. From the result most of the item number have item difficulty in categorized as 'moderate'. While, two items which are categorized as 'difficult'. Analysis from the researcher that the one of factor that influence to the difficulty index is the level of students understanding to the question of the items and the information that provided. Another factor is when the instrument is tested in the class, there are students that still confuse with the term that used in the question and the information that has provided.
- Item discrimination, item discrimination as a measure used to discriminate between students in the top with that of the low group who obtained the correct responses. A highly discriminating item indicates that the students who had high tests scores (high achiever) got the item correct whereas students who had low test scores (low achiever) got the item incorrect¹⁵. The item number that has a categorized by poor items are number 6 and 10, satisfactory items are number 1, 4, 5, 9, 12, 15, 17, 20, 21, 22, 25, good items are number 2, 3, 7, 8, 11, 13, 14, 16, 19, 23, 24, 26, and excellent item is number 18. While, the other item such as item number 6 and 10 are categorized by poor item. From the result the number of item that has a good quality to differentiate between high respondent or achiever and low respondent or low achiever are 24 test items or 92% of test items.
- Distractor analysis, a distractor analysis assists in distinguishing plausible distractors from implausible ones. One way to study responses to distractors is with a frequency table that tells the proportion of students who selected a given distractor. Remove or replace distractors selected by a few or no students because students find them to be implausible. Distractors are classified as the

incorrect answer in a multiple-choice question. According to Instructional Assessment Resources (IAR), student performance in an exam item are very much influence by the quality of the given distractors¹⁴. The distractor should be considered an important part of the item test. A total of distractor that used are 104 from 26 items. Based on the result the total distractor that not usable are 21 because it was less than 5% chosen by total of respondent. For the other distractor is useable but there are some of the distractor have a poor distractor because the total number of respondent who choose that distractor near from the total number of respondent that choose the correct option or answer such as the distractor of item number 3. The distractor of number 3 option B, correct answer option A chosen by 58 respondent and option B as distractor chosen by 55 respondents. It means that the distractor of option B is similar with the correct answer. For another example is item number 7, option A as distractor the number respondent who choose the option A is more than the correct answer (option C). It means that the respondent think that the correct answer is option A. In general the distractor that should be change or revise is 25% from overall of the distractor.

4. Conclusion

Based on the research results it is acquired some conclusions that are: (1) in general the development of science virtual test item to measure students' critical thinking on living things and environmental sustainability consist of five steps: content analysis, constructing the instrument (multiple choice), validity of the instrument by the expert, legibility test by students, and conducting the large test; (2) science virtual test item constructed to measure the students' critical thinking based on the elements of critical thinking by Inch. Detailing the eight elements of critical thinking is a function that is interconnected, as follows: purpose, question at issue, assumptions, point of view, information, concepts, interpretation and inference, implication and consequences; (3) there were five criteria that have been tested on legibility test such as the description of information, question, the option, interrelation of questions and answers, and terms that have been provided; (4) based on the large test field is gained the results of validity and reliability of the test, difficulty index, discriminating power, and quality of distractor. The coefficient alpha (α) was 0.747, the reliability of the test was categorized as 'high' and value of XY correlation was 0.63 which mean that the validity of the test was categorized as 'high'. These means that science virtual test item can be measure a students' critical thinking with good consistency.

5. References

- [1] Zhang L F 2003 *The Journal of Psychology* **137** 517-44
- [2] Jamil M, Tariq R H, Shami P A and Zakariys B 2012 *TOJET: The Turkish Online Journal of Educational Technology* **11**
- [3] Sim G, Holifield P and Brown M 2004 *Research in Learning Technology* **12**
- [4] Aliyu S, Ibrahim U E, Junaidu S B and Kana D F 2015 *International Journal of Advanced Research in Computer Science and Software Engineering* **5** 25-8
- [5] Haladyna T M and Rodriguez M C 2013 *Developing and Validating Test Items* (New York: Routledge)
- [6] McCowan R J and McCowan S C 1999 *Item Analysis for Criterion-Referenced Tests* (New York: Buffalo State College)
- [7] Cohen L, Manion L and Morrison K 2007 *Research Methods in Education* ed 6 (Companion Website)
- [8] Fraenkel J R, Wallen N E and Hyun H H 2011 *How to Design and Evaluate Research in Education* ed 8 (New York: McGraw-Hill)

- [9] Inch E S, Warnich B and Endres D 2006 *Critical Thinking and Communication: The Use of Reason in Argument* (Boston: Allyn and Bacon)
- [10] Page G T and Thomas J B 1979 *International Dictionary of Education* (London: Pitman)
- [11] Jacobs L C and Chase C I 1992 *Development and Using Test Effectively* (San Fransisco: Jossey – Bass Publisher)
- [12] Arikunto S 2012 *Prosedur Penelitian Suatu Pendekatan Praktik* (Jakarta: PT. Rineka Cipta)
- [13] Weimer M 2015 Advantages and Disadvantages of Different Type of Test Questions *Faculty Focus* (Mumbai: Magna Publication)
- [14] Adeleke A A and Joshua E O 2015 *Journal of Education and Practice* **6** 28-42
- [15] Zimmaro D M 2004 *Writing Good Multiple-Choice Exams* (The Division of Instructional and Assessment, University of Texas)
- [16] Paul R and Elder L 2006 *The Miniature Guide to Critical Thinking: Concepts and Tools* (The Critical Thinking Community)