Physics Creative Thinking Measurement using Two-Tier Multiple Choice to Support Science, Technology, Engineering, and Mathematics

Muhammad Megawan*, and Edi Istiyono
Yogyakarta State University, Indonesia
Email: muhammad.megawan2016@student.uny.ac.id

Abstract. The Importance of STEM as an implementation to Indonesian Curriculum has to be backed with proper assessment. This due to assessment take the most important part in evaluation. This research develops creative thinking assessment, which using 3 stages including test design, trials, and assembling. The design of the test will be tested in two-tier multiple choice and scored using item response theory (IRT) with daily problem in Physics. The answer will be varied in divergent answer, consisting of fluency, flexible, original and elaborate idea as a creative thinking aspect. From 5 expert’s judgements, the test passed with 7.2 value using Aiken’s V index. After revised, the test was reach trial stage with 270 students in several schools in Gunung Kidul. The Reliability of the test described in information function and SEM (Standard error measurement) are valid result and fit to PCM. The test is ready to assemble since its valid and reliable.

Keywords: Creative thinking; STEM; Measurement; Physics; Senior high school.

1. Introduction

Rather than teach in science, technology, engineering, and mathematics (STEM) separately they do better if teaching together as an application. STEM was successful in developing higher order thinking students in United States, it then spreads globally [1]. In Indonesia, STEM will help education because both of education purposes and STEM develop the students with higher order thinking skills as creative and critical thinking.

In science, physics have a good position to apply the skills to think creatively. Physics is a subject consisting of basic theory and formula about nature. From that part of nature, the students can think creatively by processing the theory and formula to solve the problems [2]. The Process that lead into creating more than one solution is called divergent thinking [3]. Divergent thinking let s solution to be varied by thinking fluently, flexibly, originally, and elaborate [4]. Developing divergent answer as a part of test is a way to understand creative thinking in physics.

After some learning periods, teacher has to evaluate their learning by measuring the students to keep improving in order to reach those purposes [5]. The problem in Indonesia is that the availability of creative measurement tool in physics is lack options. Teacher mostly uses lower order thinking instrument to measure higher order thinking skills in Physics [6]. Teacher can’t evaluate their learning process if the measurement is not suitable for education purposes.
Evaluation is essential in our daily life, including in our education. By doing evaluation, teacher is able to give judgement and evaluate against education purposes. Teacher unable to perform any of evaluation if they did not know the quality or value of the students. Activity to measure how much value their student had is called assessment [7]. Teacher must assess their student in order to evaluate and improve.

Develop assessment for higher order thinking skills is not easy as lower order thinking skills. Brookhart [8] stated that the test for higher order thinking cannot be done in usual multiple choice. Higher order thinking skills ask students to analyze, synthesize and create, while usual test ask them to think which one is not right or which one is right. Two tier multiple choice can accommodate higher order thinking assessment.

Activity to create something from scratch of knowledge can be utilized by thinking creative. Bott[9] stated that thinking creative must be able to synthesize information to produce some solution. Several solution can be classified into thinking divergently. The skills to think divergently lead into (1) fluency, which means producing solution in large numbers, (2) flexibility, which means flexible in producing varied ideas and answers, (3) originality, which means the ability to produce different and unique ideas, and (4) elaboration to produce ideas with details [10]. In Physics learning, the solution can be processed by thinking creatively. Physics as a subject consist of basic formula and concept about STEM. Hence, this paper presents physics creative thinking measurement using two-tier multiple choice to support science, technology, engineering, and mathematics.

The rest of this paper is organized as follow: Section 2 describes the proposed research method. Section 3 presents the obtained results and following by discussion. Finally Section 4 concludes this work.

2. Research Method

This paper uses Research and Development 4-D (Define, Design, Development, Disseminat) method as a method to develop computer adaptive test (CAT) combined with test development planning by Dallo Antonio & Oriondo [11] that consist of several step (a). Design (2). Trials, and (3). Assembling the test. This research will conduct in several class in several school in Gunung Kidul, Yogyakarta. The test that developed needed to get tested by many students to test the reliability.

2.1 Build Test

2.1.1 Design the Test

This test purposes suited in curriculum, as the test will help teacher in develop students to act creative and effective in physics subject. The subject based on 2013 curriculum (2016 revised) physics subject in odd semester (see Table 1).

| Subject                        | Matter on the Subject                                      |
|--------------------------------|------------------------------------------------------------|
| Rotational move and inertia    | Force Moment, Inertia, Rigid body, angular momentum law on rotation |
| Elasticity                    | Elasticity on solid body, Hooke’s law, Strain and Stress  |
| Static Fluid                  | Fluid pressure, Hydrostatic pressure, Pascal’s law, Archimedes’s law, Meniscus, capilarity, viscosity, stokes law |
| Dynamic Fluid                 | Ideal fluid, continuity, Bernoulli’s law                   |
| Heat and Calor                | Heat and thermal, Black’s law, Conduction, convection, radiation, State of matter. |

After deciding the subject, construct matrixes which similar to test blueprint contain information about the number of matter distribution. The test will developed in 200 problems in 4 different model called A, B, C, and D.
As stated above that creative thinking will be noted as divergent thinking. The test will be build by using aspect and sub aspect from divergent thinking. Aspects (1) fluency, which means producing answers in large numbers, (2) flexibility, which means flexible to produce varied ideas and answers, (3) originality, which means the ability to produce different and unique ideas, and (4) elaboration to produce ideas with details. Researcher then concludes indicator from aspects and sub-aspects [12] (see Table 2).

| Aspects       | Sub-Aspects                      | Indicators                                      | Bloom’s Cognitive level |
|---------------|----------------------------------|-------------------------------------------------|-------------------------|
| Fluency       | Summarizing the answers          | Answer Questions with facts                     | C5                      |
|               | Critiquing the objects or situation | Able to see the mistakes of a objects or situation | C5                      |
|               | Producing Ideas                  | Fluent to generate ideas                        | C6                      |
|               | Attributing situation            | Giving point of view about some situation       | C4                      |
| Flexibility   | Organizing situation             | Organizing situation into different category    | C4                      |
|               | Planning different answers       | Think and find some problem solutions            | C6                      |
| Originality   | Producing new things             | Finishing new kind of problem                   | C6                      |
|               | Differentiating ideas            | Enrich the idea from previous discoveries        | C4                      |
| Elaboration   | Planning to solve problem with procedural ways | Seek deeper meaning from a solution of a problem | C6                      |
|               | Generating something new         | Create something different                      | C6                      |

The scoring guidance of this test is using Partial Credit Model (PCM) for polythomous item. Student will get 4 if both answer and reason is correct, 3 if only reason is correct, 2 if only answer is correct and 1 if none of the answer and reason is right. Then the value will be put in the formula formula to get the value of creative thinking ability for each students.

The purposes of using PCM was to get better in scoring students ability [13]. PCM can be used to interpret the science and critical thinking ability [14]. It is a scoring system teacher needed to measure the ability not only score of the student. The PCM is a developed Rasch model of scoring system, every item has different difficulty depend on the person trait. Muraki and Bock has mentioned the formula of PCM [15].

\[
p_{ig}(\theta) = \frac{\exp \left[ \sum_{n=1}^{m} (I_i - h_{ig}) \right]}{\sum_{n=1}^{m} \exp \left[ \sum_{n=1}^{m} (I_i - h_{ig}) \right]}, \quad g=1, 2, 3, ..., m-1
\]

Where:
- \( I_i(\theta) \) = information about i item
- \( P_{i g}(\theta) \) = probability student with ability \( \theta \) to answer correctly
Those formula can be arranged with scores 1, 2, 3, and 4. The differences of the score have interpretation of difference probability of each student. After construct at least 200 question for test trials, the next step was getting validation from expert judgements. Validity Rubric consist of 3 scale, from 1 until 3. Scale (1) for Not Suited, scale (2) for Suit and then (3) for Suit very much. Those scale will be analyzed by Aiken’s V indexs by Azwar [16].

\[
V = \frac{s}{n(c-1)}
\]

Where, \(V =\) Aiken’s index; \(I_0 = \) Lowest scale point; \(r = \) from \(I_0+1\) until \(I_0+(c-1)\); \(s = r- I_0\)

Next step by converting the data from Aikens value of all validation into interval of Aiken’s value form 0 until 1. Aiken’s value can be accepted if scores from 0.037 until 1.00. The test should get revised after this judgements. After the test designed and got validated, the example would be like in below:

\textit{Cerobong asap didesain memiliki 2 bagian, yakni atas dan bawah. Bagian bawah cerobong asap berhubungan dengan sebuah ruangan sedangkan bagian atasnya berhubungan dengan lingkungan yang luas. Bagaimana kondisi kecepatan dan tekanan pada cerobong asap, apabila simbol a untuk cerobong diatas ruangan, dan b untuk cerobong dibawah ruangan?}

- \(a.\) \(V_a > V_b\) sehingga \(P_a > P_b\)
- \(b.\) \(V_a > V_b\) sehingga \(V_a < V_b\) \textbf{Jawaban}
- \(c.\) \(V_a > V_b\) sehingga \(P_a = P_b\)
- \(d.\) \(V_a < V_b\) sehingga \(P_a > P_b\)
- \(e.\) \(V_a = V_b\) sehingga \(P_a > P_b\)

\textbf{Alasan}

- \(a.\) Udara yang ada dibawah cerobong asap bergerak keseluruh ruangan sehingga tekanannya menyebab
- \(b.\) Udara yang ada dibawah cerobong asap memiliki kecepatan yang lebih tinggi daripada diatas, hal ini dikarenakan udara didalam ruangan bebas bergerak
- \(c.\) Udara yang ada diatas cerobong asap yang memiliki lubang kecil, memiliki tekanan yang besar
- \(d.\) Udara yang ada diatas cerobong asap kecil karena pengaruh lingkungan \textbf{Jawaban}
- \(e.\) Tekanan diatas dan dibawah cerobong asap bernilai sama, hal ini dikarenakan udara yang

\subsection{2.1.2 Trial Test}

The researcher need to do limited test which needed 250 students at least. The test was hold in 5 schools in Gunung Kidul, Yogyakarta with PBT (Papper Based Test) to test the question level and the reliability of the test. The overall reliability value of the instrument is in accordance with the interpretation of the reliability value with the Rasch model is shown in Table 3.

| No | High School Name     | Packet A | Packet B | Packet C | Packet D | Student Total |
|----|----------------------|----------|----------|----------|----------|---------------|
| 1  | SMAN 1 Wonosari      | 9        | 8        | 8        | 8        | 33            |
| 2  | SMAN 2 Wonosari      | 20       | 18       | 17       | 20       | 75            |
| 3  | SMAN 1 Tanjungsari   | 7        | 6        | 6        | 7        | 26            |
| 4  | SMAN 2 Playen        | 20       | 20       | 20       | 15       | 75            |
| 5  | SMAN 1 Karangmojo    | 15       | 19       | 18       | 14       | 62            |
|    | Student Total        | 73       | 73       | 71       | 66       | 269           |
2.1.3 Test Assembly
Analyzing the test results for quality information such as reliability of the items and the test.

3. Results and Discussion
This section presents the results obtained and following by discussion.

3.1. Test Validation
The validation taken by 5 expert judgements. The values from judgement then calculated with aiken’s V formula above. The results produce point at 0.67-0.74. Based on the theory above, the result is valid.

3.2. Item Estimation and Reliability
After instrument got validated and revised, the test hold in five school to reach at least minimum requirement to get the reliability value (see Table 4).

Table 4. Item Estimation and Reliability in Empirical PBT

| No | Detail                          | Estimation for Item | Estimation for Test |
|----|--------------------------------|---------------------|---------------------|
| 1  | Average and Std.Deviation      | -0.01± 0.61         | -0.49±0.36          |
| 2  | Reliability                    | 0.71                | 0.73                |
| 3  | Average and Std.Deviation INFIT MNSQ | 1.00±0.05         | 1.01±0.25          |
| 4  | Average and Std.Deviation OUTFIT MNSQ | 1.04±0.26         | 1.06±0.66          |
| 5  | Average and Std.Deviation INFIT t | 0.00±0.57          | -0.10±1.11         |
| 6  | Average and Std.Deviation OUTFIT t | 0.14±1.10          | 0.06±0.66          |
| 7  | 0 Score in Test                | 0                   | 0                   |
| 8  | Perfect Score in Test          | 0                   | 0                   |

By looking at INFIT Mean of Square (Mean INFIT MNSQ) is shown in Table 5. An Item is fit if INFIT MNSQ between 0.77 and 1.30. also by using INFIT t between -2.0 and 2.0. After analyzed by Quest, the 170 item including the anchors item is fit with the test [17].

Table 5. Infit MNSQ Example with 10 items sample

| MNSQ | .53 | .56 | .59 | .63 | .67 | .71 | .77 | .83 | .91 | 1.00 | 1.10 | 1.20 | 1.30 |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|
| 1.40 |     |     |     |     |     |     |     |     |     |      |      |      |      |
| 1.50 |     |     |     |     |     |     |     |     |     |      |      |      |      |
| 1.60 |     |     |     |     |     |     |     |     |     |      |      |      |      |
| 1.70 |     |     |     |     |     |     |     |     |     |      |      |      |      |
| 1.80 |     |     |     |     |     |     |     |     |     |      |      |      |      |
| 1.90 |     |     |     |     |     |     |     |     |     |      |      |      |      |

3.3. Item Theory Characteristic
The character of items will be able to look by ICC item characteristic curve and threshold item index (see Figure 1). Using Parscale as a program to help analyzing the results.
3.4. Item Reliability

Information function and SEM recognize how item theory can describe trait ability of students. The higher information function in the test, resulting minimum mistakes. The connection between Parscale output of IF and SEM connection is shown in Figure 2 below:

![Figure 2. Information Function and Standard Error Measurement (SEM)](image)

From the Table 5 above, this test is suit for students with higher ability in physics (θ) which between $-1.9 \leq \theta \leq 2.7$.

3.5. Item Difficulty

The difficulty able to analyze using Quest as a program. From this test, the test reach interval between -0.9 and 1.37. Hambleton & Swaminathan [18] state as long it is not under -2 or above 2, the item is fit.

3.6. Creative Measurement

From the test, each students answer will be analyzed with polythomus scoring system. The results was varied for each students with results from $\Theta=0.15$ with score 53, to 1.6 with score 73. Based on the information above, this test is valid and reliable as a creative thinking measurement in Physics.
4. Conclusion

This paper has presented physics creative thinking measurement using two-tier multiple choice to support science, technology, engineering, and mathematics. From the results obtained that this creative thinking measurement in Physics test is valid and reliable.

References

[1] Rehmat, A.P. Engineering the path to higher-order thinking in elementary education: A problem-based learning approach for STEM integration. UNLV Theses, Dissertations, Professional Papers, and Capstones 2496, 1 (2015).
[2] Ramankulov, S., Berkimbaev, K., Bekbayev, S., Abdumanapov, U., Ormanova, G., & Sarybaeva, A. Formation of Creative Thinking of Future Teachers on Physics by Means of Information Technologies. Cmes, 492–495. (2015)
[3] Runco, M. A., Acar, S., Runco, M. A., & Acar, S. Divergent Thinking as an Indicator of Creative Potential Divergent Thinking as an Indicator of Creative Potential. Creativity Research Journal, 24(1), 66-75. 2012.
[4] Tahar V. A., Tej J., & Sirkova M. Creative management techniques and methods as a part of the management education: analytical study on student’s perceptions. Procedia-Social and Behavioral Sciences, 197, 1918-1925, 2015.
[5] Cheng, V. M. Y. Teaching creative thinking in regular science lessons: Potentials and obstacles of three different approaches in an Asian context. In Asia-Pacific Forum on Science Learning & Teaching, Vol. 11, No. 1, 2010.
[6] Istiyono, E., Mardapi, D. & Suparno. Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PhysTHOTS) PesertaDidik SMA. Jurnal Penelitian dan Evaluasi, 2014.
[7] Mardapi, D. Pengukuran, Penilaian, dan Evaluasi Pendidikan. Nuha Medika, 2012.
[8] Brookhart, S. How to Assess Higher Order Thinking Skills in Your Classroom, ASCD 2011.
[9] Bott, N., Quintin, E. M., Saggar, M., Kineitz, E., Royalty, A., Hong, D. W. C., Liu, N., Chien, Y. H., Hawthorne, G., & Reiss, A. L. Creativity training enhances goal-directed creativity, 2014.
[10] Zeng, L., Proctor, R. W., Salvendy, G., Zeng, L., Proctor, R. W., & Salvendy, G. Can Traditional Divergent Thinking Tests Be Trusted in Measuring and Predicting Real-World Creativity, 2014.
[11] Oriondo, L. L. & Antonio, E. M.D. Evaluating educational outcomes (Test, measurement and evaluation). Florentino St: Rex Printing Company, Inc., 1998.
[12] Rahayu, F. & Istiyono, E. “Development of Creative Thinking Skills Tests Using Computerized Adaptive Test (CAT) on High School Physics Subjects Class X”. Unpublished Thesis. UniversitasNegeri Yogyakarta, 2017.
[13] Wahyu, W. PengembanganSkalaPsikologi”. Universitas Gajah Mada, 2010.
[14] Istiyono, E. The analysis of senior high school students’ physics HOTS in Bantul District measured using PhysReMChoTHOTS. 2017.
[15] Muraki, E. & Bock, R. D. PARSCE: IRT item analysis and test scoring for rating scale data. Chicago: Scientific Software, 1997
[16] Aiken, L. R. Three Coefficients for Analyzing the Reliability and Validity of Ratings. Educational and Psychological Measurement. 1985.
[17] Adams, R. & Khoo, S. T. Quest: the interactive test analysis system quest. 1996.
[18] Hambleton R. K. & Swaminathan, H. Fundamentals of Item Response Theory. SAGE Publications, Inc., 1991.