Developing computer based test to assess students' problem-solving in physics learning

Developing computer based test to assess students' problem-solving in physics learning

E istiyono[1], W S B Dwandaru[1], A K Permatasari[1] and Aristiawan[1]

[1]Physics Education, Universitas Negeri Yogyakarta, Sleman, Yogyakarta

Corresponding author: edi_istiyono@uny.ac.id

Abstract. This study was developmental research. The study was development of CBT based IRT to assess students'problem-solving(PhysTePSoS-CBT). PhysTePSoS-CBT was developed based on 4D model. The PhysTePSoS-CBT was assessed by three experts to know the feasibility of PhysTePSoS-CBT. The feasibility of PhysTePSoS-CBT was assessed in the aspectsof correctness, reliability, integrity, usability, interface, and navigation. Based on the assessment from the experts, all aspects of feasibility were in excellent category with an average percentage of 98%. So, it can be concluded that PhysTePSoS-CBT was valid and has good quality.

Keywords: *CBT, problem solving skills, physics test, PhysTePSoS-CBT, reasoning multiple choice*

1. Introduction

Problem solving skills (PSS) becomes an important aspect in physics learning and Indonesia's Curriculum . One skill that is required in the 21st century is PSS [1] - [4]. In Indonesia, student are required to have problem solving skills [5]. In addition, PSS is included in the Curriculum of Indonesia, ecspecially in the Core Competency in physics subject. PSS is important in physics. Physics learning contains problems from daily life. PSS is a component needed by students to understand the concept of physics in real situations [6], [7]. Physics learning not only masters concepts, but also applies concepts in solving physics problems [8]. PSS is needed by students to understand physics in real situations through equations and correct concepts to be used to solve physics problems. The aspects of PSS according to Polya[9] are identification the problem, planning a solution, conductethe solution, and evaluation.

PSS is related to assessment education [10]. Assessment is part of the plan and implementation of the learning process and to determine the effectiveness and efficiency of the learning process . Teacher needs truly assessment that can assess problem solving skills [5]. Problem solving skills assessment is needed to know the effectiveness and efficiency of the physics learning.

The assessment used by teachers so far is only limited to the assessment of low-level cognitive domains. The high-level cognitive domains concerning students' high order thinking skills (HOTS), such as Bloomian and Marzonian HOTS, critical thinking, and problem solving [2], [11] need to be done nowadays, which is is important in Indonesia's Curriculum. Hence, developing PSS must be done.

Technology plays an important role in education. The use of technology in learning is to improve the effectiveness of learning [1]. One of the utilization of technology in education is using

computer to assess students' ability. This is called computer based test (CBT). The CBT in educational assessments have been widely used. This relates to the benefits of CBT. CBT has a feature that can process data at high speed without errors making the computer as an assessment tool in education [12]. CBT also helps to accelerate in meeting the needs of feedback in education [1].

Assessment of students' high order thinking skills (HOTS) requires the test form that can measure students' ability accurately. This is due to many shortcomings ofmultiple choice, namely students answering randomly and cheating. The development of a reasoned multiple choice format to cover multiple choice shortcomings has been investigated in [13] and the study was used for scoring students' answer and reasoning. An assessment score can be seen in table 1.

Table 1. Assessment score in PhysTePSoS-CBT.

Score	Category
4	Answer and reason are right
3	Reason is right but answer is wrong
2	Reason is wrong, but answer is right
1	Answer and reason are wrong

The use of item response theory (IRT) in the assessment also increases the accuracy of the measurement result. It is because IRT is done to cover the weakness of classical test theory (CTT). One of the weakness of CTT is it cannot calculate the difficulty level of each step of completion [14]. The modern test theory is IRT. Two postulates are the basic of IRT, i.e.: 1) the test of student's ability on items is predicted with latent traits (θ) and 2) the relationship between student's ability on the test item and the underlying ability is related to the item characteristic curve (ICC) [15]. ICC is useful for removing the weaknesses of CTT because it shows the interaction of the test with the ability of students [14]. Students who have high ICC have higher latent abilities because ICC shows the opportunity to answer correctly. The equation is used for calculating the probability of answering true [P (θ)]. The result of ability (θ) is in the range of -3 to +3. Based on the explanation of IRT application it can be concluded that the analysis to get the right PSS uses IRT.

From the above explanation, the problem statement here is "What is the assessment theory to assess students' problem solving skills correctly?" Hence, the aim of this research is developing item response theory-based computer based test to measure the students' ability to solve problems in physics (PhysTePSoS-CBT) that is valid and has good quality.

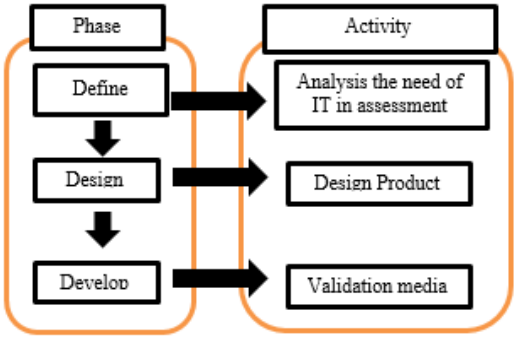


Figure 1. Phase and Activity of developing PhysTePSoS-CBT.

2. Research method

The study was a research and development study. The study was developing CBT based IRT (PhysTePSoS-CBT). PhysTePSoS-CBT was developed based on 4D model that was developed by Thiagarajan, Semmel&Semmel[16].There was four phases of development of PhysTePSoS-CBT, viz.: 1) define; 2) design; 3) develop; and 4) desseminate the media. However the development of PhysTePSoS-CBT in this study only reached the third stage, i.e.: developing the media. Procedure and activities of developing the PhysTePSoS-CBT can be seen in figure 1.

PhysTePSoS-CBT had been developed.The feasibility of the media also had been examined. The feasibility of the PhysTePSoS-CBT had been examined by the media expert. The media expert of PhysTePSoS-CBT consists of 3 experts. The aspects of the feasibility of PhysTePSoS-CBT are based on Pressman [17]. Softwere as media must fulfiill the aforementioned aspects. These aspects can be seen in table 2.

Table 2. Feasibility aspects of PhysTePSoS-CBT.

No.	Aspect	Indicators
1.	Correctness	Completeness
2.	Reliability	Accuracy
		Tolerance of error
3.	Integrity	Instrumentation
		Safety
4.	Usability	Ease use of program
5.	Interface	Ease use of menu and button
		Layout of Navigation
		Visibility of text
		Aesthetic and ease using
6.	Navigation	Correctness link of navigation
		Ease in Seeking the Content
		Correctness link and direction in navigation system
		Ability to go back in the page

The feasibility of the media was done to reveal the feasibility of PhysTePSoS-CBT. The questionnaire of the media uses the Guttman scale. Data response of the media by the media expert was analyzed using Equation (1), which was calcuted from the total score in all item questionnaires, i.e.:

$$N = \frac{k}{Nk} \times 100\%, \qquad (1)$$

With N isthe percentage of the response, k is thescore of the result of the expert media, and $N_k$isthe maximum score. Moreover, the percentage of the result was converted tobecome category of the feasible media to PhysTePSoS-CBT. The PhysTePSoS-CBT category from the percentage of the feasible media can be seen table 3.

Table 3. The category of the feasible media.

Interval of Criteria	Category
86%≤N<100%	Excellent
72%≤N<85%	Good
58≤N<71%	Sufficiently
44≤N<57%	Bad
N<44%	Worst

3. Results and Discussion

The development of IRT-based CBT is required for physics assessment. It is obtained from the define phase. In the define phase,it is produced that IRT-based CBT is really needed in measuring the PSS. The result of the design phase is the product of PhyTePSoS-CBT. The design phase of PhyTePSoS-CBT starts with making the storyboards and determiningthe algorithms that are in accordance with IRT.

The students' answers are recorded by CBT. Using the score guidelines, the CBTs' algorithm used categories of the students' answers. It used Equations (2), (3), (4), and (5), i.e.:

Category 1 $$P_{ih}(\theta) = \frac{\exp(\theta - b_{i1})}{\psi} \quad (2)$$

Category 2 $$P_{ih}(\theta) = \frac{\exp(2\theta - b_{i1} - b_{i2})}{\psi} \quad (3)$$

Category 3 $$P_{ih}(\theta) = \frac{\exp(3\theta - b_{i1} - b_{i2} - b_{i3})}{\psi} \quad (4)$$

Category 4 $$P_{ih}(\theta) = \frac{\exp(4\theta - b_{i1} - b_{i2} - b_{i3} - b_{i4})}{\psi} \quad (5)$$

In the PCM model, analyzing the students' response concerned is the item and the ability parameters of students. In this estimation, it is known as the likelihood function. The likelihood function for cases with *N* students can be stated in Equation (6), viz.:

$$\boldsymbol{L}(\boldsymbol{\theta}, \mathbf{b}; \mathbf{u}) = \prod_i \prod_h \boldsymbol{P}_i\,(\boldsymbol{\theta}_h; \mathbf{b}_i)\,[\mathbf{1} - \mathbf{P}_i(\boldsymbol{\theta}_h; \mathbf{b}_i)], \quad (6)$$

Where $P_{ih}(\Theta)$ is the probability of the student(Θ) to get score *h* categoryon *i* item, Θ is the students' PSS, b_{ih} is the difficulty index of item *i* in category *h*, *m*+1 is the category, *L* is the maximum likelihood estimation, and *u* is the category on item *i*. Moreover, the value of θ can be converted in the range of 0 to 100 using Equation (7), i.e.:

$$\text{Score}(100) = 50 + \frac{50}{3}\theta. \quad (7)$$

Algorithm is applied in CBT to got θ (the students' PSS). The result of θ in PhysTePSoS-CBT can be seen in figure 2.

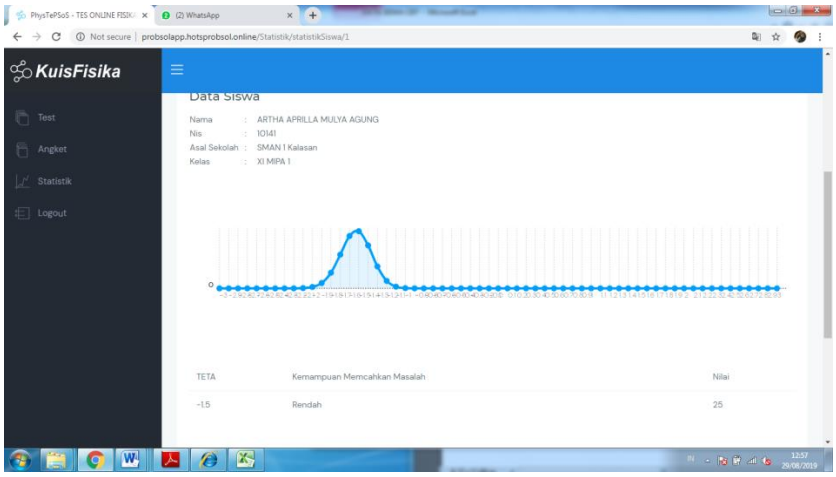
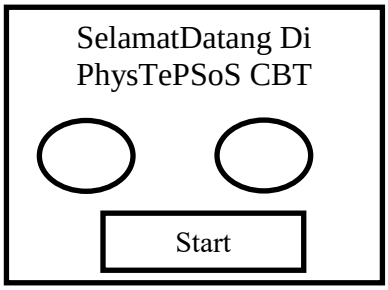
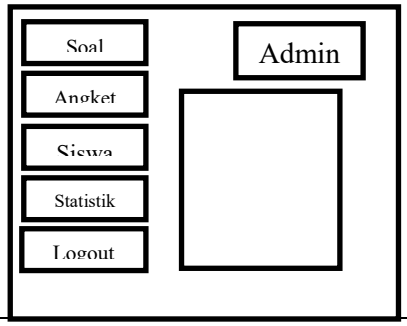
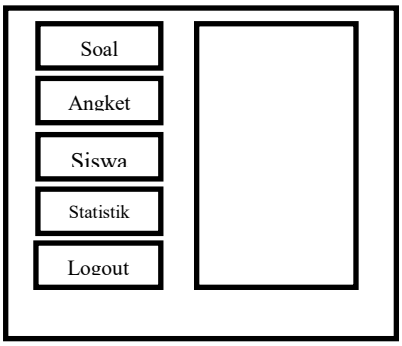
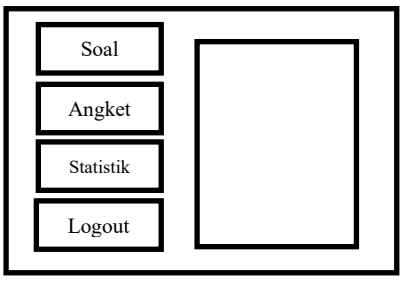


Figure 2. Display of the students' PSS.

The results of the PhysTePSoS-CBT in the design phase (see table 4) include: 1) home display; 2) administrator display; 3) teacher appearance; and 4) student display.

Table 4. The result of the design phrase of PhysTePSoS-CBT.

User	Display	Explanation
Home		Home display of PhysTePSoS-CBT contains the name of the program, the development team, the material to be tested, the definition of PSS and start button to log in.
Admin		Administration display contains: a) test; b) questionnaire; c) student; d) statistics;and e) logout. In the admin menu, admin can control all activities, either add and/or edit and/or delete the test and/or questionnaire and/or student and/or teacher.
Teacher		The teacher display is almost the same with admin display, but teacher cannot add and/or edit and/or delete the test and/or questionnaire and/or student and/or teacher.
Student		The student display is only able to file the test, questionnaire, and see the result of test and questionnaire.

The implementation of PhysTePSoS-CBT program can be seen in Figures 3 to 6. Figure3 shows the display of the PhysTePSoS-CBTs' home. Figure 4 shows the display of Admin of PhysTePSoS-CBT. Figure5 shows the display of teacher of PhysTePSoS-CBT. Figure 6 shows the display of student of PhysTePSoS-CBT.

Figure 3. The Display of PhysTePSoS-CBTs' home.

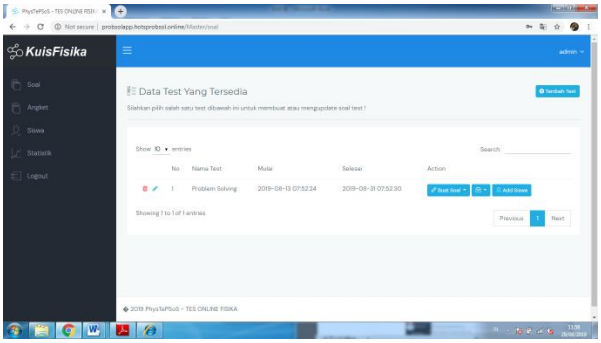


Figure 4. The display of admin of PhysTePSoS-CBT.

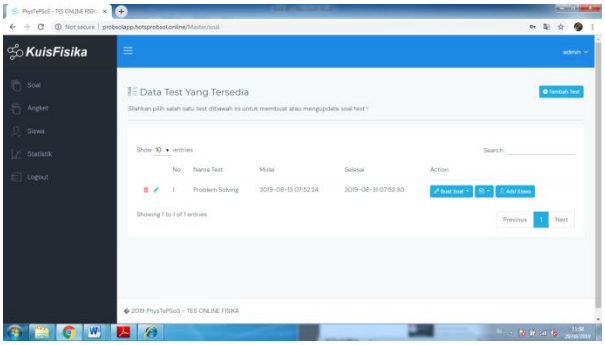


Figure 5. The display of teacher of PhysTePSoS-CBT.

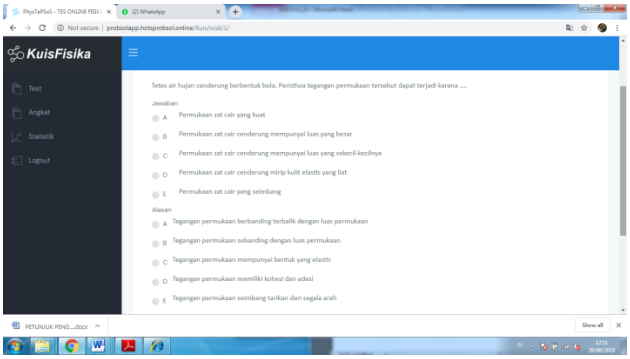


Figure 6. Display of student of PhysTePSoS-CBT.

The media validation to know the feasibility of PhysTePSoS-CBT before used in the trial is given in table 5. The validators assess the quality of PhysTePSoS-CBT based on aspects of quality of the media.The result of media validation of PhysTePSoS-CBT can be seen in table 5.

Table 5. Theresult of media validation of PhysTePSoS-CBT

Aspect	Percentage (%)	Category
Correctness	100	Excellent
Reliability	100	Excellent
Integrity	88	Excellent
Usability	100	Excellent
Interface	100	Excellent
Navigation	100	Excellent

Aspect of correctness has 100% and excellent category. It means that PhysTePSoS-CBT can give the correct resultsof the softwere. The reliability aspect has 100% and excellent category, which shows that PhysTePSoS-CBT is excellent in accuracy and tolerancy in failure. The third aspect is integrity having 88% and excellent catgeory.This means that PhysTePSoS-CBT has excellent category in instrumentation and safeness. The usability aspect has 100% and excellent category, which means that the PhysTePSoS-CBT has excellent in ease in using for assessing the students' PSS. The interface aspect has 100% and excellent category, which means thatthe menu and button is easy to use, layout and visiblility in excellent category. The last aspect is navigation, which has 100% and excellent category. It means that the functioning mechanism of the PhysTePSoS-CBT has excellent catgeory. The average of media validation aspect is 98% with excellent category. Hence, PhysTePSoS-CBT is valid and has good quality. PhysTePSoS-CBT fulfills all aspect of the feasible media tated Pressman [17]. The media experts agree that PhysTePSoS-CBT has advantages and efficient in assessment and giving feedback in short time.The above results are in accordance with Redecker & Johannessen [1] as CBT saves time and can give feedback directly.

4. Conclusion

Based on this developmental research,a CBT-based on IRT to measure students' problem-solving skills in physics (PhysTePSoS-CBT) is developed. The PhysTePSoS-CBT is assesed by experts to know the feasibility of PhysTePSoS-CBT. The feasibility of PhysTePSoS-CBT is assessed in the aspects of correctness, reliability, integrity, usability, interface and navigation. Based on the assessment from the experts, all aspects of feasibility isin excellent category with an average percentage of 98%. So, it can be concluded that PhysTePSoS-CBT is valid and has good quality.

References

[1] Redecker C and Johannessen Ø 2013 *European Journal of Education* **48** 79–96. https://doi.org/10.1111/ejed.12018
[2] McNeill M, Gospe M and Xu J 2012 *Research in Learning Technology* **20** 283–96. https://doi.org/10.3402/rlt.v20i0.17595
[3] Winarti, Cari, Suparmi, Sunarno W and Istiyono E 2015 *Proc. Int. Conf. Science and Applied Science (Solo)* vol 795 (Bristol: IOP Publishing) p 1-5 https://doi.org/10.1088/1742-6596/795/1/012052
[4] Nadapdap A T Y, Lede Y A and Istiyono E 2016 *Proc. International Seminar on Science Education (Yogyakarta)* (Bristol: IOP Publishing) p 37-42
[5] Permatasari A K, Istiyono E and Kuswanto H 2019 *International Journal of Educational Research Review* **4** 358–66 https://doi.org/10.24331/ijere.573872
[6] Nadapdap A T Y and Istiyono E 2017 *Research and Evaluation in Education* **3** 114 https://doi.org/10.21831/reid.v3i2.14982
[7] Istiyono E, Mustakim SS, Widhiastuti, Suratno and Mukti 2019 *Jurnal Pendidikan IPA Indonesia* **8** 170-6 https://doi.org/10.15294/jpii.v8i2.17766
[8] Azizah R, Yuliati L and Latifah E 2015 *Jurnal Penelitian Fisika dan Aplikasinya (JPFA)* **5** 44 https://doi.org/10.26740/jpfa.v5n2.p44-50
[9] Polya G 1957 *How to Solve it: A new Aspect of Mathematical Method* (Doubloday Garden City)
[10] Selçuk G S, ÇalS and Erol M 2008 *Lat. Am. Journal Physics Education* **2** 151–66. https://doi.org/10.1.1.669.3132
[11] Istiyono E, Mardapi D and Suparno 2014 *Jurnal Penelitian dan Evaluasi Pendidikan* **18** 1–12 https://doi.org/10.21831/PEP.V18I1.2120
[12] Akdemi O and Oguz A 2008 *Computers & Education* **51** 1-5 http://dx.doi.org/10.1016/j.compedu.2007.11.007
[13] Istiyono E, Dwandaru W S B and Faizah R 2018 *Research and Evaluation in Education* **4** 144–54. https://doi.org/10.21831/reid.v4i2.22218
[14] Hambleton R K and Swaminathan H 1985 *Item Response Theory Principles and Applications* (New York: Springer Science) pp 23-31
[15] Meyer J and Zhu S 2013 *Research & Practice in Assessment* **8** 26–39
[16] Thiagarajan S S and Semmel M I 1974 *Instructional Development for Training Teachers of Exceptional Children*
[17] Pressman 2010 *Software Engineering: A Practitioner's Approach* (New York: McGraw Hill)