

Optimizing Computer-Based Hots Instruments: An Analysis of Test Items, Stimulus, and Quiz Setting Based on Physics Teachers' Perceptions

Agus Suyatna*, Viyanti, Undang Rosidin

Department of Physics Education, Faculty of Teacher Training and Education, University of Lampung, Indonesia

Received December 9, 2019; Revised February 22, 2020; Accepted February 25, 2020

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Computer-based test (CBT) is still viewed to be incapable of measuring students' higher order thinking skills (HOTS). This research aims to describe the teachers' perceptions on the CBT model in stimulating HOTS among high school students in physics exams. This is a descriptive qualitative research involving lecturers and professional teachers in Lampung as respondents. The data were collected through Likert-scale questionnaire. The fifth scale shows strongly agree and first scale shows strongly disagree. Data analysis was conducted by counting the frequency of each scale for each questionnaire item to be then converted into percentage. The results show that teachers perceive that HOTS stimulus structure in the form of images/graphics, video/audio physical phenomena, animated/experimental simulations can enrich the diversity of items and the level of cognitive thinking. The items and options on a CBT are suggested to be raised randomly. In CBT, it is recommended to display the remaining time available for working on the test. Scoring is distinguished by cognitive level, formation of questions, and level of difficulty of the questions. Feedback from the system, in the form of follow-up for the students, is given both to the students who can work on the problem as a form of enrichment and the students who cannot answer the question as a guide to understand the material being tested.

Keywords CBT, Test Item Type, Teachers' Perception, HOTS

1. Introduction

A written test can no longer attract the interest of education practitioners to be administered in the era of

enormous development of technology. The development of technology, information, and communication has opened the way for a change on tests administration from paper-based (PBT) to computer-based (CBT). CBT is believed to be able to solve the problems related to the instability at preparation stage, multiplications during distribution of test sheets, treachery during the administration of test, and the demand for extra time and effort for checking the test results [1].

The administration of paper-based test with non-essay items are prone to the threat of treachery and relies heavily on the role of the invigilators. When it is not strictly supervised, the candidates can easily copy the works of other, often better, candidates. Some codes are often used to overcome the difficulties caused by seating arrangements so they can share the answers for test items with the same numbers. The order of the options, which is often made to be exactly the same one another, also provides with the ease of this misconduct. Problems related to treachery can actually be solved by administering test items demanding answers in essay form yet the checking process would require a great deal of time and energy, especially when there is a large number of candidates involved. Another problem is the demand for test checkers with thorough understanding on subject matter being tested. The last problem is often related to scoring stage at which the test checkers might be inconsistent in giving the judgement, despite the given rubrics, after being distracted by an activity as simple as pausing.

All these problems can be solved by altering the test model from PBT to CBT as the CBT offers a number of advantages. The first is that both options and test items in CBT can be randomized by the computer that disables candidates from copying the answers from other candidates. The checking can also be done by

computerized process. As the candidate clicks the submit button, the test items are checked and the results will be out in no time at all. Candidates may also receive feedback designed by the test crafter in order to notice their areas of weaknesses. There are various types of test items provided by CBT software: multiple choice, multiple respond, true-false, matching, sequence, fill in the blank. CBT also offers some aids in form of videos, audio, animation, simulation, and graphs. However, it is still considered to be unable to effectively measure candidates' creativity, problem solving skill, or critical thinking [2]. This notion has to be further investigated as the writer believes that CBT can be used for measuring, or even stimulating HOTS like critical thinking skills by optimizing the role of illustration, choice of test types, quiz setting, question setting, and feedback.

One of the learning objectives of physics education in high school is to develop the ability in comprehending the inductive and deductive thinking analysis by using the concepts and principles of physics education to elucidate various natural phenomena and to solve the problem both qualitatively and quantitatively. The ability of HOTS is very crucial to grow through the instructional process at school. HOTS focus on developing students' ability to effectively analyze, synthesis, evaluate existing information and create something new. HOTS is a major component of creative and critical thinking and creative thinking pedagogy can help students develop more innovative ideas, ideal perspectives and imaginative insights [3]. Related to this, it is necessary to train students to be able to create and incorporate these skills in the learning process, then these students already have HOTS. It is revealed that HOTS is easy to practice, and students have the right to learn and apply this thinking to solve problems [4]. The result of TIMMS in 2011 engaged Indonesia in the 38th place out of 42 countries [5] and PISA 2012 in the field of science placed Indonesia in the rank of 63th out of 64 countries [6]. Generally, the capability of the Indonesian students is so low in: (1) integrating the information, (2) generalizing cases into one solution, (3) formulating real life problem into a concept of school subject, and (4) doing an investigation [7]. The main reason of the low achievement of the Indonesian students is the lack of their ability in solving the items which urge them to have high level of thinking skills. HOTS development helps students avoid mistakes in thinking [8]. Thus, it is important to develop physics CBT assessment model to measure and stimulate the HOTS ability which can be easily accessed by the students continuously.

Assessment is the main factor in learning. Thus, the implementation of an assessment should be facilitated optimally in the term of technique, method, or even the quality of the items. Accordingly [9], LMS-based online test is an advantageous tool to administer knowledge assessment. Automatic marking is possible for multiple

choice items and short essay of which the keywords loaded on the answer cell. So far, the feedback of essay assessment is still difficult to implement, especially the diagram figures. The CBT is still believed to be ineffective in evaluating the creativity, problem solving, and critical thinking [2]. This argument needs to be further investigated. In developing CBT, it is important to choose an application or a software of which the feature is providing with various types of items, inserting the illustration as the HOTS stimuli, and proving with various quiz or question setting so that the author is flexible in arranging and making the instruments, especially for measuring and exercising HOTS. For instance, the program called Wondershare Quiz Creator (WCS) is really user friendly in the items making, so it is easy to use and operate which does not require any complex programming language. The results of the items, quizzes, and tests made by using this software can be stored in a flash format which can stand alone in a website. The publication of the interactive evaluation program using WCS is also varied in form of SWF, HTML, or .EXE. file. By using WCS, the user is able to create and arrange any kind of test form such as True/False, multiple choice, multiple respond, fill in the blank, sequence, matching, closed test, word bank, click map, or short essay. Even, pictures and movies can be inserted in WCS to support the learners' comprehension in doing the test.

Some other facilities available in WCS are (1) the feedback based on the answers or the response of the test takers, (2) the preview of test score result and the steps followed by the test takers based on their answers, (3) the feature of changing the language of the button and label of the application, (4) the feature of inserting voice notes and colors on each item based on the desire of the test items maker, (5) the hyperlink which can facilitate the users to send the result or the score through e-mail or LMS, (7) the privacy setting with user account/password, (8) the display setting which can be modified, etc. [10]. The response to the students can be given by the computer. Those powers make it possible to develop the assessment item instruments that can be used to measure and practice the HOTS ability.

The capability of higher order thinking skills includes metacognitive and critical, logical, reflective, and creative thinking skills [11]. The higher order thinking skill trains the students to think rather than to memorize [12]. It means applying new information or previous knowledge manipulate the information to reach possibility of any response in a new situation [13]. It is also a thinking process which involves mental activities in exploring complex, reflective, and creative experiences consciously done to obtain the objectives. Higher order thinking skills involves analytical, critical, syntactical, and evaluative thinking along with problem solving [14]. Learning to foster higher-order thinking skills requires clear communication to reduce ambiguity and confusion and

improve students' attitudes regarding the task. Learning that exercises higher-order thinking skills facilitates students to analyze, describe, interpret basic principles of contextual [15]. The lesson plan must include the model of thinking skills, the example of applying thinking, and the adaptation to the students' different needs. Scaffolding (providing support at the beginning of the lesson and gradually leading to independence) will help students develop high-level learning skills. Small group activities such as discussions, peer tutorials, and cooperative learning will be effective in developing thinking skills. Activities such as challenging assignments, encouragement to work on assignments, and providing feedback on group progress must be carried out. Computers mediate communication and learning need to be held to facilitate access to remote data sources and allow collaboration with students in other locations. Three task items / formats useful in measuring higher-order thinking skills are: (a) choosing, including plural choices, matching, and sorting; (b) short answers, essays, and performance or assignments; and (c) explaining, including giving reasons for the answers chosen [11]. High-level thinking assessment includes three principles: (a) present a stimulus for students to think about, usually in the form of introductory texts, visuals, scenarios, discourse, or problems (cases), (b) use new problems for students, which have not been discussed in class, and not just questions for the process of remembering, and (c) distinguish between the difficulty level of questions (easy, medium, or difficult) and cognitive levels (low-level thinking and high-level thinking) [7].

The author has also developed an interactive quiz program to practice the ability to explore physical phenomena. Quizzes are made in four types of questions, which are multiple respond, multiple choice, true-false, and fill in the blank. Each item is equipped with illustrations that illustrate physical phenomena. Illustrations can be in the form of videos or interactive simulations related to the items. From the trial results, it is known that the interactive quiz program is declared to be interesting and easy to operate. Installation of the program can be done easily. Programs and navigation buttons can be run without difficulty. Likewise, the illustrations in the form of animation can be run easily. Through the practice using this quiz program, there is an increase in the ability to explore physical phenomena [16-18]. Exploration of new physical phenomena is the first step towards HOT training. Therefore, it needs to be followed up with research that develops quiz programs to measure and train HOT. These findings will complement the role of ICT in learning.

The aim of this research is to describe (1) the physics

teachers' perception of the CBT model to stimulate the HOTS of high school students, (2) the type of questions suitable for measuring HOTS high school students in learning physics, and (3) the quiz settings and question settings appropriate for CBT. Based on this perception an effective CBT model was designed to measure and stimulate HOTS.

2. Research Method

This research aims to provide the instrument model of computer-based assessment to measure and practice the critical thinking of the high school students in comprehending physics education. Hopefully, the model is easy to use and also positively responded by the students and the teachers. In order to find it out, the teachers' perception of CBT model to stimulate the students' HOTS in learning physics education should be described initially. The method of this research is descriptive qualitative. Moore over, the subjects taken were 62 professional physics teachers and lecturers in Lampung Province. The data were obtained through Likert-scale questionnaire using Google Form. The scale of 5 (five) represents strongly agree and the scale of 1 (one) represents strongly disagree. The data analysis was done by calculating the frequency of each scale of the questionnaire items, and it is converted into percentage form. The results are used to design a certain CBT model which is effective to measure and stimulate the HOTS.

3. Result and Discussion

The questionnaire was analyzed by calculating the frequency of each answer scale on each statement. The calculation results are converted to a percentage. The results of data analysis are presented in Table 1 and Table 2.

Based on Table 1, it is known that the respondents' perceptions about the form of illustrations that can be made using the CBT application in this case refer to the WQC application. All respondents (98%) already know the form of illustrations inserted into the item can be in the form of video, audio, images, animation, simulation experiments. Movies or videos that can be inserted are in flv, swf, mp4, mov, wmv, and avi format. Thus, the test developer can insert a video from filming using a mobile phone or downloaded from YouTube. Image or images that can be inserted are in jpg, jpeg, bmp, png, gif, emf, and wmf format. Audio or sound can be inserted by recording directly or through an mp3 format audio file. The percentage perception of CBT and its random questions can be described on the tables below.

Table 1. CBT model perception

No	Statement	SDA		DA		LA		A		SA	
1	Computer Based Test (CBT) Model allows the stimulus items in a form of video, audio, animation, and interactive simulation					1	2%	19	30%	42	68%
2	The stimulus items in the form of images, graphics, video, audio, animation, and interactive simulations can enrich the diversity of HOTS items.							22	36%	40	64%
3	Stimulus items in the form of images, graphics, video, audio, animation, and interactive simulations enrich the level of HOTS cognitive thinking that will be measured.			1	2%	3	5%	20	32%	38	61%
4	Stimulus items in the form of video, audio, animation, and simulation can enrich the indicators of competency achievement (GPA) to be measured.			1	2%	2	3%	25	40%	34	55%
5	Stimulus items in the form of video, audio, animation, and interactive simulations strongly support assessment as learning.					2	3%	28	45%	32	52%
6	The form of stimulus items in the form of video, animation, simulation can reduce verbalism (too long items/questions)					2	3%	18	29%	42	68%

Note:

SDA : Strongly Disagree

DA: Disagree

LA: Low Agree

A: Agree

SA: Strongly Agree

Table 2. Randomization of questions

No	Statement	SDA		DA		LA		A		SA	
1	Items appear randomly	1	2%	4	6%	11	18%	14	22%	32	52%
2	Option (answer choices) appear randomly	2	3%	9	14%	6	10%	19	31%	26	42%

It is well-known that WQC or other question maker application is equipped with facilities to make various types of questions such as True / False, Multiple Choice, Multiple Respond, Fill in the Blank, Sequence, Matching, Short Answer, Word Bank, Click Map, and Short Essay. In the structure of HOTS questions, it generally uses stimulus. Stimulus is the basis for understanding information. In the context of HOTS, the stimulus presented must be contextual and attractive. HOTS that are equipped with a stimulus affect the ability of speed and effectiveness of learning. Therefore, students are more thinkable to increase understanding in solving Science problems. Finally, students will get used to thinking competitive, developing intellectually and helping to avoid mistakes in thinking. All respondents (100%) agree and strongly agree that stimulus HOTS questions in the form of images, graphics, video, audio, animation, interactive simulations can enrich the diversity of questions. Stimulus in the form of video, audio, animation, interactive simulation is only possible in CBT. The stimulus in the form of video, audio, animation,

interactive simulation is not widely available on CBT that is used today. This is a challenge going forward. The belief of physics teachers that the diversity of stimulus forms can enrich the diversity of forms of questions that can be arranged, is a very valuable asset for the HOTS physics assessment developer. For example, procedural knowledge whose stimulus in the form of video will be easily measured using the form of sequence problems. Likewise, factual knowledge about objects whose characteristics can be ordered from the results of reading the graph will be more suitable to be measured using the form of sequence problems. During this time, we are too rigid with the choice of the form of CBT questions, according [16] it's time to consider steps to adopt the right number of multiple-choice items and the variables are not limited to multiple choice items with three, four, or five choices. Others have recommended that multiple choice items should not be used because they are susceptible for guessing [20]. Statistical analysis considering the abilities and gender of students shows that elimination testing with adjusted assessments is a valuable alternative to negative

assessments when looking for assessment methods that do not support guesswork. This study shows that elimination testing with adaptation scores reduces blank answers and finds strong indications for reduced guesses compared to negative ratings. Students prefer elimination testing with adaptation scores rather than negative assessments, and it reports lower stress levels in elimination testing with adaptation scores compared to negative assessments [21].

Stimulus items in the form of images, graphics, video, audio, animation, interactive simulations can enrich the level of cognitive thinking HOTS that will be measured to be believed by 93% of respondents. This belief is very rational because test developers will be more flexible in choosing the form of stimulus. The stimulus used must be interesting, meaning that it must be able to encourage students to read, contain new facts or current issues, be contextual or be found in students' daily lives. Some things that need to be considered to compile a stimulus about HOTS are: (1) selecting some information can be in the form of pictures, graphics, tables, or discourse who have a connection in a case, (2) demanding the ability to interpret, look for relationships, analyze, conclude, or create, (3) choosing cases / contextual and interesting (recent) problems that motivate students to read, and (4) directly related to questions (subject matter), and functioning [7]. In line with the requirements of a stimulus about HOTS, it is clear that the stimulus will enrich measurable cognitive levels.

Stimulus items in the form of video, audio, animation, and simulation can enrich the indicators of competency achievement (GPA) to be measured. This statement was approved by 90% of respondents. Information that can be obtained from a stimulus in the form of a video, although of short duration, will be far more than written down in a discourse, or in the form of a case. Likewise, the stimulus in the form of an experimental simulation, which allows changing the observed variables, will be able to enrich the GPA to be measured.

Stimulus items in the form of video, audio, animation, and interactive simulations strongly support assessment as learning, approved by 97% of respondents. This is possible because the stimulus in the form of video, audio,

animation, and interactive simulations contains real and contextual visualization of physical phenomena so that it can be used as a source of learning. Through stimulus on the problem, students can explore, practice interpreting, look for relationships, analyze, conclude, or create [16-18].

The stimulus items in the form of video, animation, simulation can reduce verbalism or too long discourse, believed by 97% of respondents. Explaining a case through video displaying, animation, or simulation can reduce the number of sentences and can clarify the physical phenomena that are the subject of the problem.

The majority of the respondents (74%) agreed and strongly agreed that items on a test appear randomly. This can overcome the problem of students' tendency to emulate their friends work or cheating. If students already know that the questions appear randomly, it is hoped that the students will focus on working on the problems, not trying to ask friends in one room.

The options (answer choices) appeared randomly were agreed by 73% of respondents. This can only be applied to multiple choice and multiple respond type questions. The advantage of randomizing the answer option is that if the appearance of the question is easily recognized remotely, it can prevent the tendency of the students from asking the people closest to them when the test is in progress.

Most respondents (85%) agreed and strongly agreed to display the time spent on the work on the items. However, they agreed (98%) that only if the computer screen displayed the remaining time available for the test. By seeing the remaining time available to take a test, students can organize their time and give more peace to think.

Most respondents (92%) agree and strongly agree. Each item should be weighted according to the level of thinking (C1, C2, C3, C4, C5, C6) which is needed to do the problem. This can help to gather accurate information related to the constraints (strengths and weaknesses) of learning, as well as the important role that children need in the learning process. The recapitulation data taken from the aspects of time setting, score weighting, and providing feedback are presented on the tables as follow.

Table 3. Time setting

No	Statement	SDA		DA		LA		A		SA	
1	On CBT, the time that has been used for working on the test items should be displayed on the computer	6	10%	0	0%	3	5%	19	30%	34	55%
2	On CBT, the remaining time for working on the test items should be displayed on the computer	0	0%	0	0%	1	2%	16	25%	45	73%

Table 4. Score Weighting

No	Statement	SDA		DA		LA		A		SA	
1	Each item should be weighted according to the level of thinking (C1, C2, C3, C4, C5, C6) needed to do the problem.	0	0%	0	0%	5	8%	26	42%	31	50%
2	Each item should be weighted according to the level of difficulty of the questions	0	0%	0	0%	5	8%	24	39%	33	53%
3	Each item should be weighted according to the type of items	0	0%	0	0%	6	10%	25	40%	31	50%

Table 5. Providing feedback

No	Statement	SDA		DA		LA		A		SA	
1	In the CBT form test, it should be designed to give feedback from the system, in the form of follow-up for students who can answer the questions correctly	0	0%	0	0%	4	6%	29	47%	29	47%
2	In the CBT form test, it should be designed to provide feedback from the system in the form of follow-up for students who cannot answer the questions correctly	0	0%	0	0%	5	8%	19	31%	38	61%
3	In CBT feedback or system feedback, it should be given for each item that has been done	0	0%	2	3%	9	15%	26	42%	25	40%
4	Feedback should be given after all items have been worked on	1	2%	0	0%	6	10%	22	35%	33	53%

Each item should be weighted according to the level of difficulty agreed by 92% of respondents. This is a form of appreciation for students and is a principle of justice. Difficult questions and solutions, which require a longer thinking time, should be given a higher weight. And conversely, problems that can be done easily, are given a lower weight. Each item should be weighted according to the type of problem approved by 90% of respondents. Complex types of questions with a greater chance of answering wrong should be given a higher weight. Each type of question has different stages of thinking for completion. Problem type T-F only has a 50% chance of being wrong. Problem multiple choice with 5 options, has a 20% chance of answering correctly, while the type of multiple response questions, has the opportunity to answer the correct varies, depending on the number of correct answers provided. Thus, it is only natural that the items are weighted according to the type of problem.

The majority of the respondents (94%) agreed and strongly agreed on the CBT form test designed to be given feedback from the system, in the form of follow-up for the students who could answer the questions correctly. This can facilitate the development of self-assessment (reflection) in learning and provide high-quality information to students about learning. In addition, feedback provides an opportunity to close the gap between current and desired performance. [22] revealed the feedback function: (1) helping clarify student performance, (2) providing high quality information to students about their learning, and (3) encouraging dialogue among

teachers related to learning. Online feedback helps enrich student participation and discussion [23].

Furthermore, 92% of respondents agreed and strongly agreed on the CBT form designed to be given feedback from the system in the form of follow-up for the students who could not answer the questions correctly. This increases the student's role and teacher's responsibility for the weaknesses of student answers. [24] suggest that quality feedback focuses not only on strengths and weaknesses, but is timely and contains high-quality information. Related to this, we need educators who have knowledge of pedagogical content to provide feedback [25]. In line with this, giving feedback to the students who cannot answer the questions correctly enables the students to have the mindset, skills, and motivation to prioritize the material that has not been mastered in a sustainable manner. It was further revealed that the feedback given with frequency often improved learning performance and assignments [26]. In line with the opinion [27] it states that feedback can enrich students' cognitive resource abilities.

In CBT feedback or system feedback, it should be given for each item that has been worked on, which was approved by 82% of respondents. Feedback is provided for this type in order to facilitate the time constraints students have regarding the knowledge they have just gained. Besides this activity has an impact on: (1) the ability of students to identify which material has been explained they can understand and vice versa, (2) the identification of students' concept understanding, (3) the

ability of teachers to understand quickly to material that is not yet understood by students, and (4) the ability of teachers to find anything that has not been conveyed clearly. [28] revealed that quality feedback if: (1) has an impact outside the feedback task; and (2) enhance the role of students interpreting and engaging in feedback activities. Everything that was apparently not clearly understood by the students. It should be noted and repeated again at the next opportunity. Another way that is better and will give more certainty is to hold a short exam.

Feedback should be given after all the items being worked on have been approved by 88% of respondents. This provides opportunities for students to: (1) increase student awareness about learning, (2) facilitate students to be stimulated to develop, monitor and evaluate their own learning, and (3) increase the capacity of students for life-long continuous learning. [29] suggests that ongoing feedback supports students in the learning process.

The majority of the respondents (76%) agreed and strongly agreed that on the CBT test, the students should be given the opportunity to repeat the test. This is related to the opportunity for students to pay attention to the items that have not been completed as one of the stimuli that lead to the achievement of learning objectives. Meanwhile, this form of test can process the learning outcomes effectively and actively and also physically, intellectually, and emotionally.

In the CBT test, it is better for the answer submission (click submit) to be done after each item has been completed, with the risk that students cannot redo the previous questions, which are approved by 45% of respondents. That is, more than half of the respondents disagreed with the CBT test in the form of answer submission after each item was completed because it is risky. The students cannot redo the questions. This is believed to have an impact on the students' motivational resources that can be used to stimulate thinking activities.

The practice of students' ability to link between what is taught and what they already know and the way the students think raises the independence of thinking which makes them ready to convey the information in their head and process it into the right answer. Being consistent with the previous opinion, the respondents agreed by 88% that click submission is done after all items have been completed so that students can redo the questions. This will provide an opportunity for the students to think again.

Based on the perception of the physics teachers who were the respondents in this study, a CBT design model was developed to measure and stimulate physics HOTS as in Figure 1. In this model, the stimulus items can be presented in the form of video, audio, and animation of contextual physical phenomena that can arouse the curiosity. The stimulus can be in the form of an experimental simulation involving several variables. There are some fixed variables and also some variables which are changed in value. The simplest stimulus is in the form of cases, graphs, or data tables of research results. Stimulus as above will enrich: the indicators of competence achievement to be measured, cognitive level, even the complexity of thinking. Types of questions made can be very varied such as True-False, Multiple Choice, Multiple Respond, Sequence, Fill in The Blank, and Matching. Scoring is distinguished by cognitive level, form of questions, and level of difficulty of the questions. For the purposes of assessment as learning, the question instrument can be supplemented with constructive feedback. To prevent cheating, randomization can be arranged which includes setting questions and question options. This model is of course just an initial idea that still needs to be tested. The percentage achievement of repetition and submission in CBT and the design of HOTS in CBT for physics learning are presented on table 6 and figure 1 below.

Table 6. Repeating and submitting

No	Statement	SDA		DA		LA		A		SA	
1	In the CBT form test, it is better to provide the chance of re-doing the test	2	3%	4	6%	9	15%	21	34%	26	42%
2	In the CBT form test, it is better to click the submission after each of the items has been done, with the risk that the students are not able to redo the previous test anymore	16	26%	11	18%	7	11%	16	26%	12	19%
3	In the CBT form test, it is better to click the submission after each of the items has been done, with the risk that the students are able to redo the previous test	1	2%	1	2%	5	8%	20	32%	35	56%

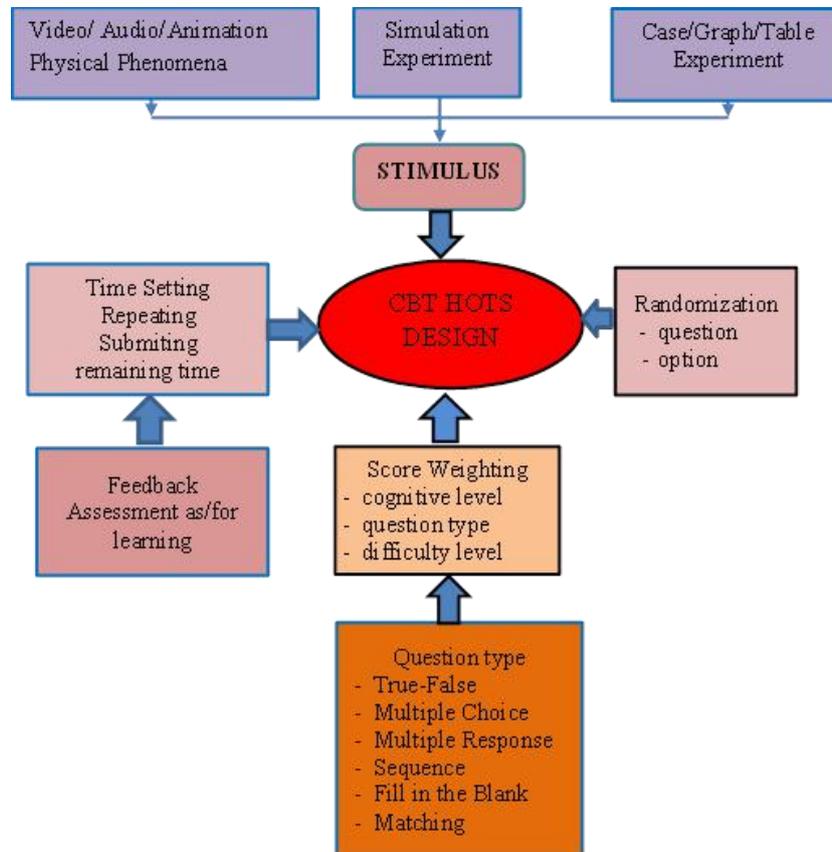


Figure 1. CBT design to measure and stimulate HOTS in learning physics

4. Conclusions

According to the perception of the physics teacher, the HOTS items need to contain stimulus to encourage the students to think. HOTS stimulus structure in the form of images, graphics, video/audio physical phenomena, animated/experimental simulations can enrich the diversity of items, and the level of cognitive thinking, and stimulus items in the form of video, audio, animation, interactive simulations strongly support assessment as learning and reduce verbalism about test questions. The items and options on a CBT are suggested to be raised randomly. In CBT, it is recommended to display the remaining time available for working on the test. Each item should be weighted according to the level of thinking, the level of difficulty of the questions, and the type of questions. Feedback from the system, in the form of follow-up for the students, is given both to the students who can work on the problem as a form of enrichment and the students who cannot answer the question as a guide to understand the material being tested.

Acknowledgments

I would love to thank the Research and Community Service Institutions (LPPM) of Lampung University which has funded this research through the professor's

grant scheme.

REFERENCES

- [1] Temitayo, M. F, Adebisi, A.A & Alice, O.O. (2013). Computer Based Test System for University Academic Enterprise Examination. *International Journal of Scientific & Technology Research*, 2(8), 336-342.
- [2] Fluck, A., Pullen, D., & Harper, C. (2009). Case study of a computer based examination system. *Australasian Journal of Educational Technology*, 25(4), 509-523.
- [3] Chinedu, C.C. & Kamin, Y. (2015). Strategies for Improving Higher Order Thinking Skills in Teaching and Learning of Design and Technology Education. *Journal of Technical Education and Training (JTET)*. 7(2)
- [4] Heong, Y. M., Yunos, J. M., Hassan, R. B., Othman, W. B., & Kiong, T. T. (2011). The perception of the level of higher order thinking skills among technical education students. In *International Conference on Social Science and Humanity journal*. Faculty of Technical Education, Universiti Tun Hussein Onn Malaysia (Vol. 5, No. 2, pp. 281-285).
- [5] HSRC & IEA TIMSS 2011. (2012). Highlights from TIMSS 2011, The South African Perspective. Diakses 7 Desember 2012 dari <http://www.hsrc.ac.za>.

- [6] OECD. (2016). Indonesian Student Performance (PISA 2012). Retrieve 7 October 2016 from <http://gpseducation.oecd.org>.
- [7] Mukhtar, M. & Haniin, K. (2019). Modul Penyusunan Soal Keterampilan Berpikir Tingkat Tinggi (Higher Order Thinking skills Fisika. Jakarta: Direktorat Pembinaan Sekolah Menengah Atas, Dirjen dikdasmen.
- [8] Heong, Y. M., Othman, W. B., Yunos, J. B. M., Kiong, T. T., Hassan, R. B., & Mohamad, M. M. B. (2011). The level of marzano higher order thinking skills among technical education students. *International Journal of Social Science and Humanity*, 1(2), 121.
- [9] Booth, R., Clayton, B., Hartcher, R., Hungar, S., Hyde, P. & Wilson, P. (2003). The development of quality online assessment in vocational education and training, Volume 1. National Centre for Vocational Education Research. Retrieve 7 October 2016 from <http://www.researchgate.net>.
- [10] Hernawati, Kuswari. (2009). Membuat Quiz/Evaluasi dengan WonderShare Quiz Creator. Modul PPM. Yogyakarta: UNY.
- [11] King, F.J., Goodson, L., & Rohani, F.(2015). Higher Order Thinking Skill, Definition, Teaching Strategies, Assessment. Center for Advancement of Learning and Assessment. Retrieve 7 October 2016 from <http://www.cala.fsu.edu>.
- [12] Leou, M., Abder, P., Riordan, M., & Zoller, U. (2006) Using HOCS-Centered Learning' as a Pathway to Promote Science Teachers' Metacognitive Development. *Journal Research in Science Education*, 36 (1), 69-84.
- [13] Heong, Y.M., Othman, W.D., Md Yunos, J., Kiong, T.T., Hassan, R., & Mohamad, M.M. (2011). The Level of Marzano Higher Order Thinking Skills Among Technical Education Students. *International Journal of Social and Humanity*, 1(2), 121-125.
- [14] Osborne, Jonathan. (2013). The 21st Century Challenge for Science Education: Assessing Scientific Reasoning. *J. Thinking Skill and Creativity* 10 (2013), 265-279.
- [15] Russel, M-R., (2011). Reflections on Learning: Students' Insights on Their Learning in a Legal Research Skill Course in The Core Curriculum. *The law Teach Journal* 45 (1) 45-62.
- [16] Ikhwan, M., Suyatna, A. & Sesunan, F. (2015). Pengembangan kuis interaktif tipe fill in the blank untuk melatih kemampuan eksplorasi fenomena fisika, *Jurnal Pembelajaran Fisika*, 3(1), 27-35.
- [17] Setiawati, A.A., Suyatna, A. & Suyanto, E. (2015). Pengembangan kuis interaktif tipe multiple choice untuk melatih kemampuan eksplorasi fenomena fisika, *Jurnal Pembelajaran Fisika*, 3(1), 49-55.
- [18] Sundari, A., Suyatna, A. & Sesunan, F. (2015). Pengembangan kuis interaktif tipe multiple response untuk melatih kemampuan eksplorasi fenomena fisika, *Jurnal Pembelajaran Fisika*, 3(1), 69-79.
- [19] Tweed, M. (2019). Adding to the debate on the numbers of options for MCQs: the case for not being limited to MCQs with three, four or five options. *BMC medical education*, 19(1), 1-4.
- [20] McKenna, P. (2019). Multiple choice questions: answering correctly and knowing the answer. *Interactive Technology and Smart Education*, 16(1), 59-73.
- [21] Vanderroost, J., Janssen, R., Eggermont, J., Callens, R., & De Laet, T. (2018). Elimination testing with adapted scoring reduces guessing and anxiety in multiple-choice assessments, but does not increase grade average in comparison with negative marking. *PloS one*, 13(10).
- [22] Arbaugh, J. B., & Hornik, S. (2006). Do Chickering and Gamson's seven principles also apply to online MBAs. *The Journal of Educators Online*, 3(2), 1-18.
- [23] Brower, H. H. (2003). On emulating classroom discussion in a distance-delivered OBHR course: Creating an on-line community. *Academy of Management Learning and Education*, 2(1), 22-36.
- [24] Nicol and Macfarlane (2006). Rethinking Technology-Supported Assessment Practices in Relation to The Seven Principles of Good Feedback Practice. In *Innovative assessment in higher education*, ed. C. Bryan and K. Clegg, 64–77. London: Routledge.
- [25] Parr, J. M., & Timperley, H. S. (2010). Feedback to writing, assessment for teaching and learning and student progress. *Assessing writing*, 15(2), 68-85.
- [26] Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: a review and critical reappraisal. *Psychological bulletin*, 95(3), 355.
- [27] Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of applied psychology*, 74(4), 657.
- [28] Hounsell, D. (2007). Towards More Sustainable Feedback to Students. In *Rethinking Assessment in Higher Education*, ed. D. Boud and N. Falchikov, 101–13. London: Routledge
- [29] Beaumont, C., M. O'Doherty, and L. Shannon. (2008). Staff and Student Perceptions of Feedback Quality in the Context of Widening Participation. Research report. York: Higher Education Academy. http://www.heacademy.ac.uk/assets/York/documents/ourwork/research/Beaumont_Final_Report.pdf (accessed August 21, 2009).