

The Development of Evaluation Instrument Based on Computer Based Test (CBT)

Mesra Damayanti, Muhammad Mifta Fausan, Indah Panca Pujiastuti

(Biology Education Department, Universitas Sulawesi Barat, Indonesia)

Corresponding Author: Muhammad Mifta Fausan

Abstract: *The utilization of computers as one of the media used for learning evaluation expected to be able to answer the shortcomings of Paper-Based Test (PBT). Assessment of learning biology is one of the required course in biology education department, Universitas Sulawesi Barat. But in this course, there has not been any evaluation instrument based on Computer Based Test (CBT). Therefore, the authors feel interested in developing an evaluation instrument based on CBT. The type of this study is R&D, adapting from the Plomp's model, with stages: the preliminary investigation, design, construction, evaluation, and implementation. To view the validity of the evaluation instrument, we used the validation sheets instrument. The evaluation instrument in this study was validated by the content expert and instructional design expert. Validation results showed the average percentage of assessment by the content expert is 87% (very valid), and the instructional design expert is 82% (valid). Furthermore, the results of testing readability of small groups consisting of 21 students who have different academic ability (high, middle, and low) showed the average percentage of assessment is 93% (very valid).*

Keywords - *Paper-based test, computer based test, evaluation instrument, assessment of learning biology*

Date of submissions: 21-08-2018

Date of acceptance: 04-09-2018

I. INTRODUCTION

Today's science and technology is growing so rapidly [1] [2]. This development certainly has an increasingly open impact and the spread of information and knowledge from and to the whole world through the boundaries of distance, place, space and time. Its influence extends to various lives, including education. Education is actually not an antipathy to the development of science and technology, but instead becomes the subject of its development. Progress in the field of knowledge and technology has a major influence on the field of education [2]. Renewal in the field of education carries the influence of attitudes, values behavior on people and society. To make progress in education, the right strategy is needed to observe supporting components such as material, methods, facilities, and evaluation. Therefore, the development of information technology and the advancement of education can be said to be closely related.

One of the media currently being developed in education is the creation of an evaluation instrument based on CBT [3]. The utilization of computers as one of the media used for evaluation of learning is expected to be able to answer the shortcomings of PBT. So the use of information, communication, and technology can serve as an evaluation medium, both locally and nationally by institutions. The using of CBT evaluation instruments in assessment of learning biology course intended for the following matters, including; a) improve the efficiency, quality, reliability, credibility and integrity of the exam, the important point is transparency in assessment and breaking the habit of cheating, because the questions are displayed automatic random by the computer; b) provide a good culture to student teacher candidates in using of computers to make measurable evaluation instruments; c) train or familiarize students in using computers, because the future all examinations are computer-based. One of the software can be used to develop evaluation instruments for assessment of learning biology courses based on CBT is Wondershare Quiz Creator 4 software. The use of this software making evaluation questions is very good, because the support features are complete enough, and the operation is easy.

II. LITERATURE REVIEW

In the development research, there are several models proposed by experts. In general, learning design models can be classified into class oriented models, system-oriented models, product-oriented models, and procedural models [4]. The development of an evaluation instrument based on CBT can basically refer to various types of development models. The development model chosen by the authors is the Plomp's development model. Plomp provides a model in designing education, which is divided into five stages, consisting of the preliminary investigation stage, the design stage, the construction stage, the evaluation and

implementation stages [5]. In this model, a development product based on the design has been produced. The products are textbooks, teaching materials, learning media and evaluation instruments [6].

Evaluation instruments in learning is a tool used to measure an object of measurement or collect data about a variable. In the field of education the instrument used to measure student learning achievement, factors have a relationship or influence their learning outcomes, the success of the teaching and learning process, and the success of achieving a program [7]. Forms of evaluation instruments are tests and non-tests [4]. Before the instrument used, it is analyzed first [8]. Two important characteristics of analyzing instruments are their validity and reliability. The instrument is valid if the instrument used to measure what should be measured. While reliability is something that can be trusted or something that shows the extent to which a measuring instrument can be trusted. Reliable tests if the scores obtained by the test participants are relatively the same despite repeated measurements [9]. A valid measuring instrument is not only able to produce the right data but also must give a right view of the data. The validity of a measuring instrument depends on the ability of the measuring instrument to achieve the desired measurement objectives appropriately [10].

CBT is a test held using the computer. The characteristic of CBT is the delivery technique for items are no longer using paper or paperless, both questions and answer sheets. Scoring or correction system by computer. Usually, participants can be working on and see items from the first number to the last [11]. According to [12] [13] [14] several advantages of computer-based testing are including reducing paper used; Checking answers made by the computer, so it saves time for checking; Test results can be directly seen by students; Questions are automatically randomized by computer, so the questions are different; Familiarize students in dealing with CBT.

III. RESEARCH METHOD

The development research model was adapted from Plomp (1997) development model with five stages of development, consisting of the preliminary investigation, design, construction, evaluation, and implementation. The Plomp's development model is schematically illustrated in Figure 1.

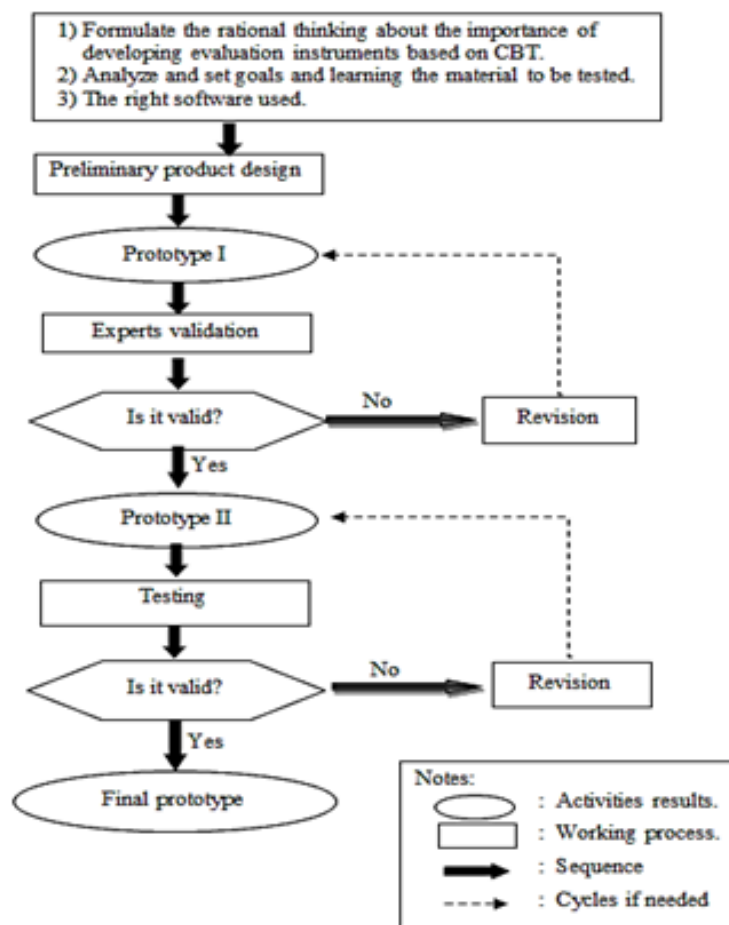


Figure 1. Design of development of evaluation instrument based on CBT with small group testing.

Preliminary Investigation stage, this activity is to formulate the rational thinking of the importance of developing an evaluation instrument based on CBT. Analyzed and set goals and learning the material to be tested. As well as the right software used for the development of an evaluation instrument based on CBT. **Design stage**, activities in this phase are more focused on the results obtained in the preliminary investigation phase, then the solution is designed. The result is a design document. Design includes a systematic process in which the complete problem of the previous phase divided into parts of the problem and the parts of the solution are applied. Then it connected to a complete structure. Evaluation instruments based on CBT are designed using Wondershare Quiz Creator 4 software. **Construction stage**, this phase is one of the production phases in addition to the design phase. In this phase, the technique for making an evaluation instrument based on CBT is made. In this phase development products are also produced based on designs have been previously designed. So that the final product of this study is an evaluation instrument based on CBT, assessment of learning biology course using Wondershare Quiz Creator 4 software. **Evaluation and Revision stages**, the evaluation phase in this study aims to determine whether the product design specifications have been good or not. Then it was revised, then returned to design activities and so on. This cycle is a feedback cycle and stops after obtaining the desired solution. In this phase also produces revised products based on input from the experts. This stage includes validation by content expert and instructional design expert, also followed by a revision process. Validator suggestions or input are taken into consideration in product improvement. So that at this stage, products can be produced in the form of valid evaluation instrument based on CBT. **Implementation stage**, the implementation in this study is small group testing which aims to see the readability of the products have been developed.

The subjects consisting of expert and students. The expert consisting of one content expert and one learning design expert. The small group testing subjects in this study were students who had taken the assessment of learning biology course. Students were chosen 21 people based on their academic abilities. We used the validation sheets which has been developed as the data collection instrument. Scoring criteria used in the validation of evaluation instrument and small group testing, consisting of score 1 (unsuitable), score 2 (less suitable), score 3 (suitable), score 4 (very suitable) [4] [15]. Data validity of evaluation instrument is data describes the validity of evaluation instrument has been developed. Validity data of evaluation instrument were analyzed using the following formula [16]. Determination of evaluation instrument has been developed using validity criteria [17], are shown in Table 1.

$$P = \frac{\sum(\text{overall score of questionnaire answers})}{n \times \text{highest indicator} \times \text{number of respondents}} \times 100\%$$

Notes: P is the percentage of appraisal; n is the number of questionnaire items.

Table 1. The criteria of validity of evaluation instrument

Achievement level (%)	Qualification	Testing decision
86-100	Very valid	Not revised
71-85	Valid	Not revised
56-70	Valid enough	Revised
41-55	Less valid	Revised
<40	Invalid	Revised

Analysis of cognitive test questions consisting of analysis of validity, and reliability. The formula used to analyze the validity of the item is Moment Product [4]. Criteria: if $R_{(t)} < R_{(c)}$ it is valid, if $R_{(t)} >$ from $R_{(c)}$ it is invalid. The formula used to analyze the reliability of cognitive test questions is Cronbach Alpha [4]. The criteria of the reliability level of the test, consisting of 0.80 until 1.00 (very high criteria); 0.60 until 0.79 (high criteria); 0.40 until 0.59 (high enough criteria); 0.20 until 0.39 (low criteria); 0,00 until 0,19 (very low criteria).

IV. RESULTS AND DISCUSSION

A. Preliminary Investigation Stage.

The results of observations have been made in the assessment of learning biology course, indicate there is not yet an evaluation instrument based on CBT. In this course, the evaluation instrument still uses paper-based tests which makes it difficult, because used many papers in each evaluation. Universitas Sulawesi Barat already has a sophisticated computer laboratory that can be used as the place to conduct evaluations, but this is not used properly because of the unavailability of evaluation instruments based on CBT. Therefore the existence of evaluation instruments based on CBT is very urgent and needed to help lecturers and students.

The characteristics of students in the biology education department generally come from various families. Cognitive development of students has shown the level of formal operational cognitive development. At this level, students have the ability to coordinate both simultaneously and sequentially by two types of cognitive abilities. Observations also show that students are more interested in conducting the evaluation based on CBT

because they are not troublesome and do not need paper for questions and answer sheets. They want a lecturer to initiate the evaluation based on CBT. One of the software can be used to develop an evaluation instrument based on CBT for assessment of learning biology courses is Wondershare Quiz Creator 4 software. The use of this software in making evaluation questions is very good, because the support of the features is complete, and the operation is easy.

Assessment of learning biology courses have a general purpose that is for students to understand the nature of evaluation and its role in the teaching and learning process of biology, evaluation of student learning processes and outcomes and able to use it in compiling biology learning evaluations or educational research instruments. Assessment of learning biology courses provide a basis for understanding educational evaluations which consisting of mastery of objectives and assessment functions; the role of evaluation in the teaching and learning process; procedures and forms of evaluation; assessed ability; planning, compilation, analysis of test principal, validity and reliability of tests/test materials (concept and process skills); processing of assessment results; performance assessments, including portfolios, practical assessments; and class-based assessment.

B. Design and Construction Stages

The stage begins by analyzing cognitive test questions consisting of analysis validity and reliability. This important to ensure the questions included in Wondershare Quiz Creator 4 software have validity relate to accuracy and reliability relate to provisions [4]. The results of the analysis of cognitive test items are shown in Table 2.

Table 2. Summary of data from analysis of cognitive test items

No	Questions validity		No	Questions validity		Questions reliability	
	R _{xy}	Decision		R _{xy}	Decision	Cronbach alpha coefficient	Decision
1	0.47	valid	24	0.63	valid	0.66	Reliable
2	0.38	valid	25	0.46	valid		
3	0.49	valid	26	0.52	valid		
4	0.49	valid	27	0.74	valid		
5	0.43	valid	28	0.58	valid		
6	0.50	valid	29	0.44	valid		
7	0.50	valid	30	0.61	valid		
8	0.37	valid	31	0.37	valid		
9	0.59	valid	32	0.47	valid		
10	0.41	valid	33	0.47	valid		
11	0.63	valid	34	0.68	valid		
12	0.41	valid	35	0.71	valid		
13	0.38	valid	36	0.79	valid		
14	0.80	valid	37	0.64	valid		
15	0.69	valid	38	0.83	valid		
16	0.74	valid	39	0.86	valid		
17	0.57	valid	40	0.55	valid		
18	0.45	valid	41	0.72	valid		
19	0.61	valid	42	0.71	valid		
20	0.42	valid	43	0.47	valid		
21	0.51	valid	44	0.61	valid		
22	0.43	valid	45	0.37	valid		
23	0.47	valid					

Table 2 shows the cognitive test items have varying levels of validity. The results have shown $R_{xy} > R_{tab}$, so the test items developed to fulfill the valid criteria (R_{tab} is 0.361). Table 2 also shows the coefficient Cronbach alpha on cognitive test questions is 0.66. Based on these results, it could be understood the developed cognitive test items fulfill reliability, with a high category.

The techniques for making evaluation instrument based on CBT are: 1) open the Wondershare Quiz Creator software, 2) select and click create a new quiz, 3) select and click quiz properties, then the display will appear as follows, on the quiz information menu, please fill in the name of the exam in the quiz title, 4) on the quiz settings menu, you can determine the minimum grade for students by filling the passing rate, you can also set how long it takes students to answer the questions, the way by checking to enable time limit, click Options, 5) on the quiz result menu, you can write the sentence "congratulations on you pass" or "you have not passed" by marking according to the result, 6) on the others menu, you can provide a password to the questions that you created by marking access with password only, and finally click OK, 7) next select and click question, after clicking it will appear several types of questions, please choose the type of question that will be made, 8) after all the questions are made, the next step is to design the appearance of the questions by clicking the player template menu, 9) review the questions have been made, the way is by clicking the preview menu, 10) the last step is publish the question by clicking publish on the menu then click CD / EXE.



Figure 2. The appearance introduction page of the evaluation instrument for assessment of learning biology courses.

After the cognitive test questions have valid and reliable criteria, and then they are included in Wondershare Quiz Creator 4 software. The appearance of evaluation instruments for assessment of learning biology courses is shown in Figure 2.

C. Evaluation and Implementation Stages

After the construction phase is complete, then it validated by experts. Validation aims to obtain data of suggestions on the accuracy and suitability of the evaluation instruments. The implementation in this study is a small group testing which aims to see the readability of the product. Summary of the results of data analysis by content expert, instructional design expert, and small group testing are shown in Table 3.

The indicators assessed by content expert and instructional design expert on the lesson plan consisting of systematically display the identity of course; Systematically test instructions for students; Information on the number of questions, passing rate, and passing scores are available on the introduction page; Navigation buttons are available; Systematically display questions and time to work on questions; Systematically display the result page; Easy to operate the software; Suitability of Background color and writing; Questions are presented interactively and are not rigid; Question items use the enhanced spelling; Font type and size are suitable and easy to read; Suitability of the questions with the indicators of learning outcomes; Suitability of the time provided with the number of questions; Question levels are compiled based on Bloom's revised taxonomy [18]; and Questions suitable with the level of cognitive development of students. Table 3 shows the content experts and instructional design expert giving appraisal for the final product with an average of 87% and 82%. Generally, components are appropriate and valid.

Content expert and instructional design expert have been provided suggestions are including (1) instructions for working on the questions should be made points, (2) note the writing errors in items number 4, 7, 10, 15, 20, 21, and 38, (3) the size of the questions must be proportional, not too small and large (4) the choice of answers in multiple choice questions should be randomized, (5) the formulation of the question is sorted by the questions item have been made, (6) the file should also be copied to the CD. Feedback: all suggestions from content experts and instructional design expert have been received, and the evaluation instrument has been revised well

Table 3. Summary of data analysis results by experts and students

No	Indicators	Experts appraisal		Average assessment of small group testing
		Content expert	Instructional design expert	
A. Cognitive assessment instrument format based on CBT				
1	Systematically display the identity of course	4	3	4
2	Systematically test instructions for students	4	3	4
3	Information on the number of questions, passing rate, and passing scores are available on the introduction page.	4	4	4
4	Navigation buttons are available	4	3	4
5	Systematically display questions and time to work on questions.	4	3	4
6	Systematically display the result page.	3	3	4
B. The quality of cognitive assessment instruments based on CBT				
7	Easy to operate the software.	3	3	3.7
8	Suitability of Background color and writing	3	3	3.7
9	Questions are presented interactively and are not rigid.	3	3	3.7
10	Question items use the enhanced spelling	4	4	3.7
11	Font type and size are suitable and easy to read.	4	4	3.7
12	Suitability of the questions with the indicators of learning outcomes.	3	3	-
13	Suitability of the time provided with the number of questions.	3	3	-
14	Question levels are compiled based on Bloom's revised taxonomy.	3	4	-
15	Questions suitable with the level of cognitive development of students.	3	3	-
Total		52	49	41
Percentage		87%	82%	93%
Qualification		Very valid	Valid	Very valid

Readability indicators assessed by students in evaluation instrument consisting of systematically display the identity of course; Systematically test instructions for students; Information on the number of questions, passing rate, and passing scores are available on the introduction page; Navigation buttons are available; Systematically display questions and time to work on questions; Systematically display the result page; Easy to operate the software; Suitability of Background color and writing; Questions are presented interactively and are not rigid; Question items use the enhanced spelling; Font type and size are suitable.

Table 3 shows the small group testing results. Students with high, middle, and low academic abilities [19] [20] gave an average of 93% for evaluation instrument. Generally, the level of legibility of the components in the evaluation instrument is valid. Therefore, the evaluation instrument could be used as one of evaluation instrument based on CBT in the assessment of learning biology course. There are no suggestions and feedback from students after small group testing. Many researchers have already done studies investigating the relationship between computer usage ability and achievement tests. They are stressing that computer usage ability is an important predictor of respondent achievement; therefore, those poor students at computers may show low achievement in CBT. However, they added that with the increase in computer technologies and access opportunities, such problems might decrease [21] [22] [23] [24] [25].

V. CONCLUSION

Evaluation instruments in learning is a tool used to measure an object of measurement or collect data about a variable. Evaluation instruments if developed correctly and systematically, a valid evaluation instrument could be produced. As in this study, the evaluation instrument based on CBT could be said valid because in the development stages follow the systematic Plomp's development. Development of evaluation instruments based on CBT in the assessment of learning biology courses is the first development in biology education department. This could be a good example to apply to other courses. This evaluation instrument will always update according to demands of the applicable curriculum so that the contents remain dynamic.

Acknowledgements

The Authors would like to thank DRPM-RISTEKDIKTI for research grant has been given.

REFERENCES

- [1]. Henriques, I. C., Sobreiro, V. A., and Kimura, H. Science and Technology Park: Future Challenges. *Technology in Society*, 53, 2018, 144-160.
- [2]. Kiyici, F. B. The Investigation of The Effect of Visiting Science Center on Scientific Process Skills. *Procedia-Social and Behavioral Sciences*, 64, 2012, 291–296.
- [3]. Bennett, R. E. Inexorable and Inevitable: the Continuing Story of Technology and Assessment. *The Journal of Technology, Learning, and Assessment*, 1(1), 2002, 1-23.
- [4]. Johnson, R. B., and Christensen, L., *Educational research: Quantitative, qualitative, and mixed approaches*. (California: SAGE Publications, 2014).
- [5]. Plomp, T. *Educational & Training System Design: Design of Education and Training*. (Netherland: University of Twente, 1997).
- [6]. Hobri. *Developmental Research*. (Jember: Jember University, 2009).
- [7]. Oliver, B., Tucker, B., Gupta, R., and Yeo, S. Evaluate: an Evaluation Instrument for Measuring Students' Perceptions of Their Engagement and Learning Outcomes. *International Journal of Assessment & Evaluation in Higher Education*, 33(6), 2008, 619-630.
- [8]. Smith, C. Building Effectiveness in Teaching through Targeted Evaluation and Response: Connecting Evaluation to Teaching Improvement in Higher Education. *International Journal of Assessment & Evaluation in Higher Education*, 33(5), 2008, 517-533.
- [9]. Gibbons, B. A. Supporting Elementary Science Education for English Learners: A Constructivist Evaluation Instrument. *The Journal of Educational Research*, 96(6), 2010, 371-379.
- [10]. Alkharusi, H. Effects of Classroom Assessment Practices on Students' Achievement Goals. *Journal Educational Assessment*, 13(4), 2008, 243-266.
- [11]. Dorans, N. J. *Computerized Adaptive Testing: A Primer*. (USA: Lawrence Erlbaum Associates, 2014).
- [12]. Noyes, J. M., and Garland, K. J. Computer- vs. Paper-based Tasks: Are they Equivalent?. *Ergonomics*, 51(9), 2008, 1352–1375.
- [13]. Kim, D. H., and Huynh, H. Comparability of Computer and Paper-and-pencil Versions of Algebra and Biology assessments. *Journal of Technology, Learning, and Assessment*, 6(4), 2007, 1-35.
- [14]. Kingston, N. M. Comparability of computer- and paper-administered Multiple-Choice Tests for K-12 Populations: A Synthesis. *Applied Measurement in Education*, 22(1), 2009, 22-37.
- [15]. Sugiyono. *Metode Penelitian Pendidikan, Pendekatan Kuantitatif, Kualitatif, dan R&D*. (Bandung: Alfabeta, 2013).
- [16]. Fausan, M. M., and Pujiastuti, I. P. Pengaruh pendekatan CTL berbasis NHT terhadap motivasi hasil belajar IPA dan retensi siswa. *Jurnal Pendidikan Biologi Indonesia*, 3(2), 2017, 133–140.
- [17]. Sudarma, I. K. Pengembangan Paket Pembelajaran dengan Model Dick & Carey Mata Kuliah Pengembangan Media Pendidikan II Program S1 Teknologi Pendidikan IKIP Negeri Singaraja. Magister thesis, Universitas Negeri Malang, Indonesia, 2006.
- [18]. Anderson, L. W., and Krathwohl, D. R. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. (New York: Addison Wesley Longman, 2001).
- [19]. Palennari, M. Exploring the Correlation between Metacognition and Cognitive Retention of Students using some Biology Teaching Strategies. *Journal of Baltic Science Education*, 15(5), 2016, 617-629.
- [20]. Prayitno, B. A., Corebima, D., Susilo, H., Zubaidah, S., and Ramli, M. Closing the Science Process Skills Gap between Students with High and Low Level Academic Achievement. *Journal of Baltic Science Education*, 16(2), 2017, 266-277.
- [21]. Yurdabakan, I. Primary School Students' Attitudes Towards Computer Based Testing and Assessment in Turkey. *Turkish Online Journal of Distance Education*- 13 (12), 2012, 177-188.
- [22]. Pomplun, M., and Custer, M. The score comparability of computerized and paper-and-pencil formats for K-3 reading tests. *Journal of Educational Computing Research*, 32(2), 2005, 153-166.
- [23]. Pomplun M., Ritchie, T., and Custer, M. Factors in Paper-and-Pencil and Computer Reading Score Differences at the Primary Grades. *Educational Assessment*, 11(2), 2006, 127-143.
- [24]. Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., and Yan, F. Does it matter if I take My Mathematics Test on Computer? A Second Empirical Study of Mode Effects in NAEP. *Journal of Technology, Learning, and Assessment*, 6(9) , 2008, 1-39.
- [25]. Hosseini, M., Abidin, M. J. Z. , Baghdarnia, M. Comparability of Test Results of Computer Based Tests (CBT) and Paper and Pencil Tests (PPT) among English Language Learners in Iran. *Procedia - Social and Behavioral Sciences*, 98, 2014, 659–667.

Mesra Damayanti, "The Development of Evaluation Instrument Based on Computer Based Test (CBT) "IOSR Journal of Research & Method in Education (IOSR-JRME) , vol. 8, no. 4, 2018, pp. 50-56.