TRANSFORMATION OF THE COMPUTERIZED MATHEMATICS FINAL EXAM: HOW WAS THE RESULTS OF COMPUTERIZED ADAPTIVE TEST-BASED ASSESSMENT?

Abdul Rahim¹, Samsul Hadi², Marlina³, Dyah Susilowati³, Muti‘ah³, Irhas³
¹Universitas Musamus
²Universitas Negeri Yogyakarta
³Universitas Bumigora

Email: abdulrahim.060691@gmail.com

DOI: http://dx.doi.org/10.26418/jpmipa.v15i2.69485

Abstract

Computerized Adaptive Test (CAT) allows the use of targeted tests, that is, each test taker obtains items that match his or her ability level. The purpose of this research is to find out the final semester exam with the CAT model. The research methodology is descriptive quantitative. Purposive sampling technique was used to take a sample of 73 students. The results of this study indicate that the CAT used in this semester final exam based on Moodle can select test items given to test takers adaptively according to the user's ability level. The very high ability category was 42 students, the high ability category was 11 students, the medium ability category was 9 students, the low ability category was 4 students, and the very low ability category was 7 students. Overall, the ability of learners is in the high category.

Keywords: Computerized Adaptive Test, Final Exam, Mathematics, Rasch Model.

INTRODUCTION

The rapid technological advances achieved have integrated into the teaching, learning, testing and assessment process thus transforming the teaching and learning process and proactively supporting user activities based on their context (Fadzil, 2018; Istiyono et al., 2019). The era of the industrial revolution 5.0 is one way to change the assessment strategy by utilizing technology (Febliza & Okatariani, 2020) including assessment in mathematics learning (Radović et al., 2019). Mathematics is considered a difficult subject by some students (Ernest 2018; Kusuma et al., 2021; Mazana et al., 2018; Noperta & Sari 2023; Rahim et al., 2020; Sufanti & Santosa 2021), There are even learners who feel bored, afraid, and uninterested in learning mathematics (Cetin et al., 2019; Furner 2017; Gresham 2018; Martínez-Sierra & García-González, 2017). Mathematics is a scary subject for students (Ramirez...
Transformation Of The Computerized Mathematics Final Exam: How Was The Results Of Computerized Adaptive Test-Based Assessment et al., 2018) especially during tests or school exams (Diego, 2017). Mathematics is still an important concentration for schools (Wang et al., 2017). Educators do their best to help students in the process of learning mathematics (Kenedi et al., 2019). Therefore, educators need to know how the actual path or process of mathematics can be understood or mastered by students (As’ari et al., 2017; Hendriana et al., 2019).

Learning implementation includes evaluation. Evaluation of student learning outcomes is carried out by educators on an ongoing basis and data to determine the extent to which, in what ways, and how educational goals have been achieved (Stufflebeam & Shinkfield, 2012). Evaluation can be in the form of assignments after teaching and learning activities, Midterm Examinations, and Final Semester Examinations.

The education evaluation model adopted in Indonesia uses the Paper Based Test (PBT) and Computer Based Test (CBT) designs (Jamiludin et al., 2017; Pramono & Retnawati, 2020). PBT is a classic paper-based assessment process, while CBT is a computer-based assessment process. The design of CBT is actually the same as PBT, but in CBT the work is done using a computer and there is no need to duplicate test questions like PBT (Istiyono et al., 2019). CBT uses a classical approach with the assumption that learners have the same abilities (Gibbons & deGruy, 2019; Toroujeni, 2021). Whereas in reality there are significant variations in ability (Dillon & Smith, 2017; Mahanal et al., 2019).

The Computerized Adaptive Test (CAT) testing model can overcome the weaknesses of the existing educational evaluation model with PBT and CBT (Plajner, 2016; Frey et al., 2020). This model allows the use of targeted tests, i.e. each test taker gets items that match his or her ability level (Aybek & Demirtasli, 2017; Delgado-Gómez et al., 2019). In other words, if the test taker can answer the previous question correctly, he will get a more difficult question and vice versa, if the test taker cannot answer the previous question correctly, he will get an easier question. The use of this adaptive test can accurately determine learners' abilities (Sahin et al., 2018; Wang et al., 2019). In addition, bright students will feel more challenged, while less intelligent students will not feel frustrated (Narciss et al., 2014). The suitability of the items on the adaptive test is based on the response to the answers given by the test takers (Kezer, 2021; Plummer et al., 2019).

The previous CAT research and development has been conducted by (Abidin et al., 2019) with the title computerized adaptive test to measure physics critical thinking skills. Another research was conducted by (Istiyono et al., 2019) with the title developing of computerized adaptive testing to measure physics higher order thinking skills of senior high school students and its feasibility of use. Based on the research and problems that have been described previously, in this study find out the mathematics final exam with the Computerized Adaptive Test (CAT) model.

**METHOD**

This research is quantitative descriptive research. The quantitative descriptive research method is a way to objectively describe a situation using
numbers, from data collection to interpretation and presentation of results. Sampling using purposive sampling technique. The subjects of this study were 73 public junior high school students in Yogyakarta City. The data collection technique is a test packaged in a web-based Computerized Adaptive Test (CAT). The data analysis technique used is descriptive quantitative to process data from student test results using CAT.

The principle of presenting questions in CAT is that initially participants are presented with questions with a medium level of difficulty (b=0). If the question can be answered correctly, the participant is presented with a new question with a higher difficulty, while if the participant answers incorrectly, a new question with an easier difficulty is presented. The flowchart of the adaptive test can be seen in Figure 1.

In computerized adaptive tests (CAT), item difficulty is estimated using the Rasch model or item response theory. The Rasch model was developed by a Danish mathematician named George Rasch. The Rasch model is mathematically written as follows (Sumintono 2014; Sumintono & Widhiarso 2015):

\[ P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \] with \( i = 1, 2, \ldots, n \)

\( P_i(\theta) \) = the probability of correctly answering a participant with ability \( \theta \) on the \( i \)-th item
\( b_i \) = \( i \)-th item difficulty level
\( e \) = number with value 2,718
\( \theta \) = ability scale
\( n \) = number of test items

RESULTS AND DISCUSSION

The Computerized Adaptive Test (CAT) used in this end of semester exam is based on a learning management system (Moodle) by
utilizing the Adaptive Quiz Plugin that has been installed. LMS has certain characteristics, such as in the management of the learning content, the management of learning process, evaluation, by implementing online examination system, as well as the administration process that can be completed using online chatting and discussion forum. One of outstanding LMS is Moodle (Modular Object-Oriented Dynamic Learning Environment) (Campo et al., 2021; Raja et al., 2021). The results of all students who have taken the test and the details of the answers of each learner can be seen in Figure 2 and Figure 3.

Figure 2. Summary of test takers' answers

Figure 3. Test takers' answers

Based on Figure 3, it shows that the selection of items is shown by a blue line. The ability of students is shown by a red line. While the standard error is described by a pink area pattern. Based on these results (blue line) it can be seen that the ups and downs of the line depend on the correct
Transformation of the Computerized Mathematics Final Exam: How Was the Results of Computerized Adaptive Test-Based Assessment

and incorrect responses of the test takers. The user's ability can be directly seen from the red line. Similarly, other test takers get different images immediately after the test, depending on whether the user responded correctly or incorrectly. This shows that the Moodle adaptive test is able to adaptively select test items given to test takers according to the user's ability level.

The test participants in this study were public junior high school students in Yogyakarta City involving 73 students. Data from the test results obtained the ability category of students. The results of the end-of-semester exam test with CAT media obtained ability estimation (θ) by utilizing item response theory. Assessment with a learning management system serves as the right platform to assess their self-learning skills, so that they become more useful and relevant in today's world of work. Moodle-based CAT media provides the lowest learner ability score is 1 and the highest learner ability score is 10. The ability of students obtained from this test is then categorized into five ability levels. The learner ability categories can be seen in Figure 4.

![Figure 4. Categories of learners’ ability](image)

Figure 4. Categories of learners’ ability

Based on Figure 4, the ability of test participants obtained with CAT media shows that the very high ability category is 42 students, the high ability category is 11 students, the medium ability category is 9 students, the low ability category is 4 students, and the very low ability category is 7 students. Overall, the ability of students is in the high category.

Computerized Adaptive Test (CAT) based assessment is based on item response theory (IRT) (Magis & Barrada 2017; Makransky et al., 2017; Clark & Watson 2019; Hadi et al., 2022). The scoring technique uses the Rasch Model. The Rasch model was developed to analyze test items that require a single step (Rahim & Haryanto, 2021; Rashid et al., 2021; Rizbudiani et al., 2021; Atikah et al., 2022; Futri et al., 2022), where items follow the pattern of the Rasch model so that individuals with higher skills are expected to have higher scores than individuals with lower abilities (Tigchelaar et al., 2017; Planinic et al., 2019; Chan et al., 2021). CAT is based on the principle that items should be selected that best measure a person at his or her ability level (Magis et al., 2017; Thompson, 2017). In general, an item is selected that provides most of the information on the estimated ability of the participant (Martin & Lazendic 2018). Then, based on the pattern of responses (correct/incorrect), including the item, the ability level is re-estimated and a new item is selected at the newly estimated ability level (Ling et al., 2017; Choi & McClennen, 2020). This process is continued until a certain level of precision of the test taker's ability level is reached.

Adaptive classroom assessment tools based on the learning management system with the CAT system provide the opportunity for teachers to assess comprehensive
Transformation Of The Computerized Mathematics Final Exam: How Was The Results Of Computerized Adaptive Test-Based Assessment

Abdul Rahim, Samsul Hadi, Marlina, Dyah Susilowati, Muti’ah, & Irhas

Transformation Of The Computerized Mathematics Final Exam: How Was The Results Of Computerized Adaptive Test-Based Assessment

adjustment of questions in Computer Adaptive Testing (CAT) based on the ability of the participant on the difficulty level of the question with the participant's previous answers (Huda et al., 2024; Litvak et al., 2024; Xu 2024). This makes for a more accurate evaluation of participants' abilities. There is less trust in the administration of questions as questions are quickly customized to the ability of the participant and can reduce the time required to complete the test. Questions are unique and relevant to the ability level of the test taker and provide a more individualized test. CAT dynamically adjusts questions to provide a more accurate and reliable assessment of a participant's ability across different ability levels (Kurisu et al., 2022; Puskulluoglu et al., 2022).

CAT avoids giving questions that are too easy or too difficult for participants, so that each question is more relevant to the participant's ability. Participants work on different questions, making it difficult to ensure uniformity of questions that are customized with complex and more difficult to implement adaptive algorithms (Kaat et al., 2023). Analysis in CAT is more complex as the questions are different for each participant and validity and reliability depend on the quality of the adaptive algorithm and question bank (Chang et al., 2021; Liu et al., 2023).

This is inversely proportional to the evaluation method that uses CBT questions given with the same level of difficulty for all participants (Ameri et al., 2024; Domínguez-Rodrigo et al., 2024) without adjustments based on individual performance. CBT require more questions for all participants as there is no difficulty adjustment, so they can take longer. Participants work on the same questions, so the test experience is less personalized (Barkhatova et al., 2024; Choudhary & Garcia, 2024).

Assessment accuracy may be lower because there is no item customization, especially for participants with very high or very low ability. Participants will face questions that are too easy or too difficult, which can be less relevant to their ability and receive the same questions, which are considered fairer in the context of uniform evaluation. CBT is easier to implement because all participants take the same test, without the need for adaptive algorithms with easier and more consistent analysis of results because all participants take the same questions. Validity and reliability are easier to maintain as all participants face the same questions (Blavt et al., 2023; Tran & Nguyen 2023; Chai et al., 2024). Therefore, the use of evaluation methods with CAT depends on the test objectives, assessment needs, and available resources and has the advantage of providing more accurate and efficient assessments. The findings of this study support previous research that each test taker gets a
different number of items because it is adaptive to the test taker's ability to answer the items obtained. Thus, the CAT-based final exam assessment shows that the assessment is more accurate by adjusting the test takers' abilities, and saving time and costs in conducting the final exam assessment.

CONCLUSION
The Computerized Adaptive Test (CAT) used in this final semester exam is based on a learning management system (Moodle) by utilizing the Adaptive Quiz Plugin that has been installed. CAT-based tests with Moodle can select test items given to test takers adaptively according to the user's ability level. The ability level of test takers with CAT showed that the very high ability category was 42 students, the high ability category was 11 students, the medium ability category was 9 students, the low ability category was 4 students, and the very low ability category was 7 students. Overall, the ability of students is in the high category.

REFERENCES


Blavt, O., Bodnar, A., Mykhalskyi, A.,


Abdul Rahim, Samsul Hadi, Marlina, Dyah Susilowati, Muti’ah, & Irhas
Transformation Of The Computerized Mathematics Final Exam: How Was The Results Of Computerized Adaptive Test-Based Assessment
Abdul Rahim, Samsul Hadi, Marlina, Dyah Susilowati, Muti’ah, & Irhas

Transformation Of The Computerized Mathematics Final Exam: How Was The Results Of Computerized Adaptive Test-Based Assessment
Transformation Of The Computerized Mathematics Final Exam: How Was The Results Of Computerized Adaptive Test-Based Assessment

Abdul Rahim, Samsul Hadi, Marlina, Dyah Susilowati, Muti`ah, & Irhas

Transformation Of The Computerized Mathematics Final Exam: How Was The Results Of Computerized Adaptive Test-Based Assessment
Shoulder and Elbow Surgery, 28(7), 1–8. doi: https://doi.org/10.1016/J.Jse.2018.11.068


doi: https://doi.org/10.1007/S11336-019-09662-9