The Implementation of Computer Adaptive Test (CAT) Using Expert System to Test Student Competencies in Higher Education

Irwan[1], Yenny Desnelita[1(✉)], Wilda Susanti[1], Gustientiedina[1], Fery Wongso[2], and Wira Jaya Hartono[2]

[1] Institut Bisnis dan Teknologi Pelita Indonesia, Pekanbaru, Indonesia
yenny.desnelita@lecturer.pelitaindonesia.ac.id
[2] STMIK Dharmapala Riau, Pekanbaru, Indonesia

Abstract. This research is based on a need analysis which finds the phenomenon that conventional learning models are considered not effective enough in developing students' learning potential to the fullest and various kinds of problems related to problems made by lecturers, most of which only come to measuring low-level thinking and problems cognitive mastery skills with the tools and test instruments used. This study aims to create a Computer Adaptive Test (CAT) test model using an expert system as a test measuring tool for students' ability to think, namely Higher Order Thinking/HOT which consists of 3 levels, namely low, middle and high. This research is a Research and Development research with ADDIE method. The data that is processed after the application of this model came from lecturers and students as respondents. The results of the practicality test of this study were 87.5 which were categorized as very practical. It can be concluded that the CAT test model is more effective than the conventional model in testing student competencies in higher education. The result of this study is the application of the CAT test model based on the Expert System as a tool for measuring the ability of Higher Order Thinking/HOT in supporting a practical partial assessment of learning outcomes (Assessment of Learning/AoL).

Keywords: CAT · Expert System · HOT · ADDIE · Computer science

1 Introduction

The test model is used to gather data and information about test takers' knowledge, attitudes, and skill levels with regard to information and communication technologies (ICT). In the meantime, [1] developed CAT for evaluating using various characteristics of items calibrated with the Samejima Graded Response Model, where the item questions have a primarily one-dimensional fit with the model and the CAT demonstrates good accuracy for evaluating scores.

Information networks have been upgraded by advancements in computer and internet technologies [2], enabling students to connect with teachers online [3]. What must be done to guarantee that advancements in ICT support and promote educational innovations

Ambiyar et al. (Eds.): ICTVET 2022, ASSEHR 747, pp. 257–267, 2023.
https://doi.org/10.2991/978-2-38476-050-3_29

such as computer-assisted student learning outcomes evaluation systems. Where exam results can be known instantly when using a computer system, and the items displayed can adjust the material and subject matter for the student's ability of thinking capacity [4–8].

Competency test activities can utilize information and communication technology (ICT), where to assess student learning outcomes using ICT with the computer adaptive test (CAT) paradigm as an alternative media used [9]. Multiple-choice assessment activities, such as e-quizzes, e-assessments, and e-evaluations, are examples of ICT-assisted CATs. When multiple choice questions are used as a robust foundation for evaluating computer-based tests, they become an objective, valid, and trustworthy test tool [10]. There have been several studies on the diagnosis of thinking skills using ICT, such as the Computer Adaptive Test (CAT), such as [11–13].

A person's mastering skill in managing and finishing a number of specific domains of work is determined by the CAT, which is better known as a set of tools/instruments in providing questions to an assessment procedure. The use of computer adaptive testing (CAT), a type of computerized assessment, is growing in many different industries, including education. CAT has a solid foundation in licensing and certification testing. In contrast to a direct linear test, the CAT chooses an item for each test-taker based on how well he did on the preceding item. In order to suit the test taker's anticipated ability, assessments are changed online and present the examinee with the item that best evaluates that ability [14].

The diagnosis of thinking ability according to Bloom's version of cognitive taxonomy can be a strong basis as support before deciding whether or not students pass or fail in certain courses or subjects. From the results of the research Irwan et al. [14] with CAT using an expert system that models the learning evaluation process where items are analyzed based on the level of difficulty, discriminatory, and distractor function using a rule base. The rule-based method can be applied to the calibration process of all test items, starting from the selection or selection of items, level of difficulty, discrimination, guessing factors, scoring and estimation of the ability of test takers [15].

The use of CAT in test implementation operations can improve the delivery of questions and assessments by permitting the inclusion of graphics, audio, and video as well as a more interactive virtual environment [16, 17]. Additionally, computer technology can manage greater knowledge, increase the efficiency of how many items are displayed next, and offer helpful feedback [18]. Therefore, it is presumed that the fusion of Bloom's taxonomy revision principles and expert systems is a highly relevant innovation to be carried out for the development of the scientific area, particularly in supporting a quality evaluation system.

Product trial after conducting the validation and revision processes. The empirical trial process is aimed at getting an illustration of whether the model being developed is effective and practical compared to previous models, and whether the developed model is feasible or not. The empirical trial process is aimed at getting an illustration of whether the model being developed is effective and practical compared to the previous one. The trial test will be conducted twice, namely limited trial test and extended trial test. Meanwhile, the subjects of the trial test are experts and students. It is possible to identify computer-assisted thinking abilities by displaying multiple choice items at

random, according to an analysis of the research findings [19]. This study has been successful in more precisely mapping test-takers' cognitive capacities based on their answers to the supplied questions. As opposed to essay-style test questions, multiple-choice assessments, according to Scully [20], have a solid ability to measure higher-order thinking skills in the area of knowledge.

The administration of computer-assisted multiple-choice test items, according to various research results, obtained the same and better validity and reliability than the traditional administration of test questions using paper with a pen [5, 21, 22]. Diagnostics on thinking skills, where educators can group and map students so they can find out the weaknesses and the strengths in answering tests [23]. The combination of Bloom's cognitive taxonomy revision concepts and expert systems and Computer Adaptive Tests (CAT) can be used to support a quality assessment system.

2 Method

The development procedure approach used in this study used the ADDIE method with the following stages:

Analysis. At this stage, an analysis of the CAT products to be developed includes an analysis of the material for the grid of thinking ability diagnostic test questions, an analysis of the rule base used in setting assumptions about the ability of the test takers, setting rules for the first item displayed is an easy level (low level). Namely remembering and understanding, setting middle and high-level question items namely application, analysis, evaluation, and creating so that later information or suggestions will come out from CAT products. In stopping multiple item questions by setting a rule base then making a table of advice or suggestions given that are in accordance with the rule base that has been determined.

Design. At the design stage, the interface is created, and the database is created by creating a CAT blueprint based on the analysis that has been done. The lecturer users, who can input passwords, add questions, modify, remove, and obtain test results reports that have been completed by test participants, primarily students, are included in the user interface design. Student users can start an online test by entering a name or nickname and responding to questions by picking responses. The test results come in the form of a diagnosis of low-level cognitive ability. At the conclusion of the test utilizing the CAT, pupils were given three levels of thinking ability: middle, high, and advice or suggestions.

Development. Conducting small- and large-scale product trials with consumers, particularly students, to gain feedback and suggestions for improving CAT products.

Implementation. Tools for CAT application testing should be made available. The user's feedback can be used to fix incorrect test results.

Evaluation. A testing instrument with 10 question items is used to evaluate CAT products for application users [24], as described in Table 1.

Table 1. CAT Product Practicality Instrument Items

No	Question Indicators
1	The application is easy to use
2	The application is not complicated
3	Simple menu interface
4	The menu interface has an identifier at the top of the view (header)
5	Memorable menu and features
6	Uncomplicated data input interface
7	The text contained in the application is clearly visible (easy to read)
8	The color composition is appropriate and attractive
9	The buttons are easy to use
10	Image looks clear

Data analysis techniques to prove the practicality of software products, namely CAT with a usability score (System Usability Score/SUS). Whereas in the proof of practicality, each respondent is counted for the total score with each statement with an odd number of 1, 3, 5, 7, 9, the scale of the answer from the respondent is reduced by 1 and for each statement with an even number 2, 4, 6, 8, 10, the scale of the respondent's answer is reduced by 5. The overall contribution score is multiplied by 2.5, divided by the total number of responders, and the overall value of the usability system is then determined. At least 8 responses were required to evaluate or demonstrate the viability of a software application using the algorithm [24–26]. According to research Hamzah [3], each component of the SUS instrument can have an item contribution score between 0 and 4, which can be used to compute the outcomes of testing software applications.

3 Result and Discussion

Figure 1 depicts the architecture of the created CAT model utilizing an expert system. The CAT application is built on an expert system, which illustrates how users are divided into lecturers and students, with lecturers developing a knowledge database of multiple-choice questions and students responding to questions on the CAT Expert System application.

3.1 Rule Base System

The following is to create a rule base as a criterion for moving multiple choice items (moving rule). Table 1 is the rule base that has been set as a rule in moving the level of multiple choice items on CAT using the Expert System.

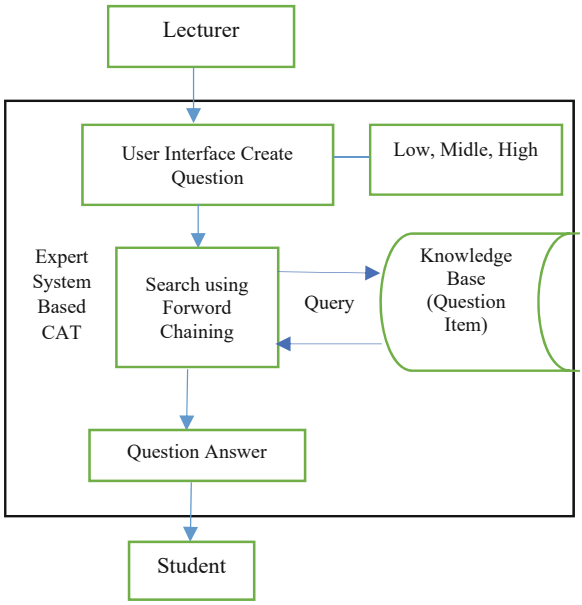

Fig. 1. Framework Computer Adaptive Test Expert System (CATExS).

After establishing the rule foundation, construct a Table 2 of advise or suggestions based on the established rule base. The advise table is a recommendation made by an expert based on the test taker's ability, and it is the result of the CAT employing the Expert System.

The development of the Computer Adaptive Test (CAT) in testing student competencies in higher education using a rule-based Expert System with the concept of Bloom's cognitive taxonomy does not decide whether or not the test takers pass based on the results of their answers, but rather to support assessment, obtain information, and map mastery on certain materials or subjects. And highlight kids' critical thinking abilities.

The rules are designed so that if 5 questions at level 1 are answered correctly, level 2 questions will be presented, and so on up to level 6. If a test taker answers less than 5 questions correctly, the following level of questions will not be presented, and guidance or suggestions will be provided to test takers in the form of a diagnostic result display.

The main menu when CAT uses the expert system is opened, then a page appears as shown in Fig. 2. This menu is the opening of the introduction of the CAT application using the expert system.

After the user enters through Fig. 2, the user can carry out activities on the application, namely education, starting from making test questions and students answering test items according to the level of low, middle and high questions.

The session score will be displayed after each session's questions have been answered. Additionally, pressing the next button will move students to questions based on the difficulty of the results of their responses as the session continues. Figure 3 is a page display when the test is in progress with a countdown timer according to the length of the test given.

Table 2. Rules Base on CAT using Expert System

Rule Name	Questions Level	Rule Base
Rule 1	The first 5 questions that appear are questions at the Middle level.	If the value of the answers from 5 questions >= 80, then the next 5 questions that appear are High questions.
		If the answer value of 5 questions >=60, then the next 5 questions that appear are Medium questions.
		If the answer value of 5 questions is <60, then the next 5 questions that appear are low questions.
Rule 2	The next 5 questions that appear are questions at a high level	If the value of the answers from 5 questions >=80, then the next 5 questions that appear are High questions.
		If the answer value of 5 questions >=60, then the next 5 questions that appear are Medium questions.
		If the answer value of 5 questions is <60, then the next 5 questions that appear are Low.
Rule 3	If the 5 questions that appear after the next rule 1 are questions at the Middle level	If the value of the answers from 5 questions >=80, then the next 5 questions that appear are High.
		If the answer value of 5 questions >=60, then the next 5 questions that appear are Medium.
		If the answer value of 5 questions is <60, then the next 5 questions that appear are Low.
Rule 4	If the 5 questions that appear after the next rule 1 are questions at low level	If the value of the answers from 5 questions >=80, then the next 5 questions that appear are High.
		If the answer value of 5 questions >=60, then the next 5 questions that appear are Medium.
		If the answer value of 5 questions is <60, then the next 5 questions that appear are Low.

Table 3. Advice/Suggestions

Ability Code	Advice
Cognitive Code 1	You have not met the minimum competency standards for this course. It is appropriate that you get additional lectures (remedial) or repeat.
Cognitive Code 2	According to System Diagnostics, your thinking ability in the subjects tested is still at a low level (Lower Order Thinking/LOT), namely **Remembering**. You should read a lot of books, journals, focus more on learning and use methods as well as learning strategies such as asking lecturers, friends and seniors.
Cognitive Code 3	According to System Diagnostics, your thinking skills in the subjects tested are still at a low level (Lower Order Thinking/LOT), namely **Understanding**. You should read a lot of books, journals, focus more on learning and use methods as well as learning strategies, for example asking lecturers, friends and seniors to have higher competence.
Cognitive Code 4	According to System Diagnostics, your thinking skills in the subjects tested are at the intermediate level (Middle Order Thinking/MOT) namely **Application**. You should focus on studying and study more diligently in order to have higher competence.
Cognitive Code 5	According to System Diagnostics, your thinking skills in the subjects tested are at a high level (Higher Order Thinking/HOT) namely **Analyzing**. You should focus on studying and study more diligently in order to have higher competence.
Cognitive Code 6	According to System Diagnostics, your thinking ability in the subject being tested is at a high level (Higher Order Thinking/HOT) namely **Deciding**. You should focus on studying and study more diligently in order to have higher competence.

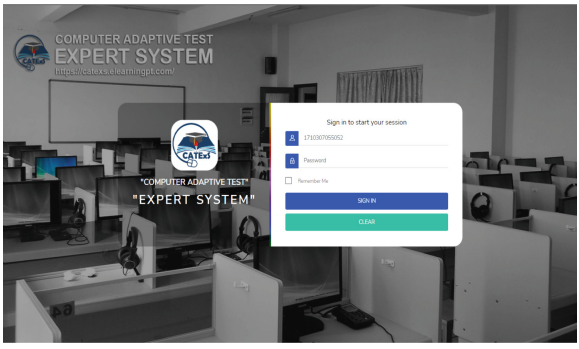

Fig. 2. Display Login CATExS

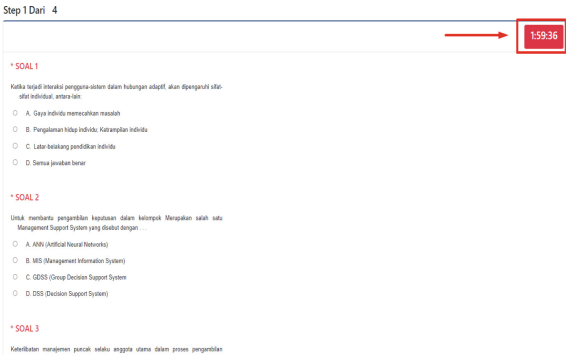

Fig. 3. Form CATExS is in progress

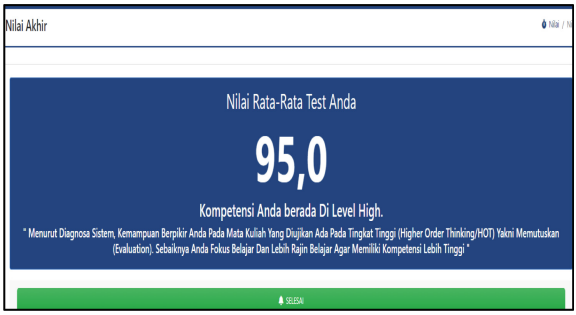

Fig. 4. Form Final Score and Diagnosis

Starting with five multiple-choice questions at level C1 being displayed at random, the administration of the multiple-choice questions for diagnosing thinking skills using CATExS has been organized linearly well. According to the stated guidelines, if the testee is successful in answering 10 questions accurately, the next five questions of level C2 will be shown. The progression continues until level C5. The test model will automatically stop, and the test taker will immediately receive help or ideas from an expert, in this case, the lecturer, if less than five questions cannot be answered properly. This makes it possible for the multiple-choice question-stopping rule system to swiftly satisfy the requirements for measuring the test subject's capacity for abstract thought.

The display of the page to see the results of the final scores and student competency level categories after the student has passed all the steps of the test on CATExS, that is illustrated in Fig. 4.

From the implementation of the CATExS product that was tested on lecturers (L) and students (S) with 10 questions from Table 1, the practicality test results were obtained as shown in Table 4.

Based on Table 3 it can be seen that the mean value for the user practicality test for CATExS products is 87.7 in the practical category. The results of proving the practicality of CATExS products that have been developed to carry out the feasibility of using the product so that it can be used widely by users.

Table 4. CATExS Product Practicality Test Results

Respondents	Score	Category
S1	90	Very Practical
S2	92.5	Very Practical
S3	67.5	Practical Enough
S4	80	Practical
S5	82.5	Practical
S6	95	Very Practical
S7	90	Very Practical
S8	82.5	Practical
S9	90	Very Practical
L1	87.5	Practical
L2	100	Very Practical
L3	92.5	Very Practical
Average SUS Score	87.5	Very Practical

4 Conclusion

The CATExS application product has supplied an alternative for diagnosing students' thinking skills, including aspects of low-level thinking skills, medium-low-level thinking, and high-order thinking skills that might cause a change in the educator's evaluation pattern of students. By applying the created products, the CATExS application with its three aspects of thinking skills can be diagnosed successfully.

Acknowledgments. The researchers express their gratitude for the funding provided by the Director of Technology Research and Community Service (DRTPM) of the Ministry of Education, Culture, Research and Technology for applied research funding in 2022. And don't forget to thank the authors for the contributions from students and lecturers of STMIK Dharmapala Riau and the Institut Bisnis dan Teknologi Pelita Indonesia (IBTPI) which has participated in conducting trials of this research product.

References

1. Á. Postigo, M. Cuesta, I. Pedrosa, J. Muñiz, and E. García-cueto, "Development of a computerized adaptive test to assess entrepreneurial personality," 2020.
2. Y. Liu and X. Zhao, "Design Flow of English Learning System Based on Item Response Theory," *Int. J. Emerg. Technol. Learn.*, vol. 12, no. 12, pp. 91–102, 2017, doi: https://doi.org/10.3991/ijet.v12i12.7958.

3. M. L. Hamzah, Ambiyar, F. Rizal, W. Simatupang, D. Irfan, and Refdinal, "Development of Augmented Reality Application for Learning Computer Network Device," *Int. J. Interact. Mob. Technol.*, vol. 15, no. 12, pp. 47–64, 2021, doi: https://doi.org/10.3991/ijim.v15i12.21993.
4. W. Malec, "Automating The Construction Of Selected-Response Items With A Text-To-Items ConverteR," *CBU Int. Conf. Innov. Sci. Educ.*, pp. 864–872, 2016.
5. H. Khoshsima, S. Morteza, and H. Toroujeni, "Comparability of Computer-Based Testing and Paper- Based Testing : Testing Comparability of Computer-Based Testing and Paper- Based Testing : Testing Mode Effect, Testing Mode Order, Computer Attitudes and Testing Mode preference," no. February, 2017.
6. I. Kalender and G. Berberoglu, "Can computerized adaptive testing work in students' admission to higher education programs in Turkey?," *Kuram ve Uygulamada Egit. Bilim.*, vol. 17, no. 2, pp. 573–596, 2017, doi: https://doi.org/10.12738/estp.2017.2.0280.
7. M. Ajinaja, "'The Design and Implementation of a Computer Based Testing System Using Component-Based Software Engineering,'" *J. Appl. Comput.*, vol. 8, no. 01, pp. 58–65, 2017, [Online]. Available: http://www.ijcst.com/vol8/2/13-michael-ajinaja.pdf
8. V. J. Shute and S. Rahimi, "Review of computer-based assessment for learning in elementary and secondary education," *J. Comput. Assist. Learn.*, vol. 33, no. 1, pp. 1–19, 2017, doi: https://doi.org/10.1111/jcal.12172.
9. C. Redecker and Ø. Johannessen, "Changing Assessment — Towards a New Assessment Paradigm Using ICT," vol. 48, no. 1, 2013.
10. M. Ghaderi, "Comparing Between Computer based Tests and Paper-and-Pencil based Tests," vol. 2, no. 4, pp. 2005–2007, 2014, doi: https://doi.org/10.7575/aiac.ijels.v.2n.4p.36.
11. B. C. Kuo, C. H. Chen, and J. de la Torre, "A Cognitive Diagnosis Model for Identifying Coexisting Skills and Misconceptions," *Appl. Psychol. Meas.*, vol. 42, no. 3, pp. 179–191, 2018, doi: https://doi.org/10.1177/0146621617722791.
12. A. Huebner, M. Finkelman, and A. Weissman, "Factors Affecting the Classification Accuracy and Average Length of a Variable-Length Cognitive Diagnostic Computerized Test," *J. Comput. Adapt. Test.*, vol. 6, no. 1, pp. 1–14, 2018, doi: https://doi.org/10.7333/1802-060101.
13. D. Tu, S. Wang, Y. Cai, J. Douglas, and H. H. Chang, "Cognitive Diagnostic Models With Attribute Hierarchies: Model Estimation With a Restricted Q-Matrix Design," *Appl. Psychol. Meas.*, vol. 43, no. 4, pp. 255–271, 2019, doi: https://doi.org/10.1177/0146621618765721.
14. R. N. I.Irwan, Y. Desnelita, W. Susanti, D. Oktarina and Putri, "The Application of Computer Adaptive Test Using Expert System in E-Learning," *Ann. Rom. Soc. Cell Biol.*, vol. 25, no. 6, pp. 5962–5974, 2021, [Online]. Available: https://annalsofrscb.ro/index.php/journal/article/view/6615
15. H. Geerlings and W. J. Van Der Linden, "Optimal Test Design With Rule-Based Item Generation," 2012, doi: https://doi.org/10.1177/0146621612468313.
16. M. Llamas-nistal, M. J. Fernández-iglesias, J. González-tato, and F. A. Mikic-fonte, "Computers & Education Blended e-assessment : Migrating classical exams to the digital world," *Comput. Educ.*, vol. 62, pp. 72–87, 2013, doi: https://doi.org/10.1016/j.compedu.2012.10.021.
17. S. J. Snodgrass, S. E. Ashby, D. A. Rivett, and T. Russell, "ascilite Implementation of an electronic Objective Structured Clinical Exam for assessing practical skills in pre-professional physiotherapy and occupational therapy programs : Examiner and course coordinator perspectives," vol. 30, no. 2, pp. 152–166, 2014.
18. J. Vie and B. Marteau, "A Heuristic Method for Large-Scale Cognitive-Diagnostic Computerized Adaptive Testing," pp. 323–326, 2017.

19. M. D. Finkelman, W. Kim, A. Weissman, and R. J. Cook, "Cognitive Diagnostic Models and Computerized Adaptive Testing : Two New Item-Selection Methods That Incorporate Response Times Cognitive Diagnostic Models and Computerized Adaptive Testing : Two New Item-Selection Methods That Incorporate Response Times," vol. 2, no. 4, 2014, doi: https://doi.org/10.7333/1412-0204059.
20. D. Scully, "Constructing Multiple-Choice Items to Measure Higher-Order Constructing Multiple-Choice Items to Measure Higher-Order Thinking," vol. 22, 2017.
21. A. J. Boevé, R. R. Meijer, C. J. Albers, Y. Beetsma, and R. J. Bosker, "Introducing Computer-Based Testing in High- Stakes Exams in Higher Education : Results of a Field Experiment," pp. 1–13, 2015, doi: https://doi.org/10.1371/journal.pone.0143616.
22. W. Malec, "Computer-Based Testing : A Necessary Evil Or A Sensible Choice ?," vol. 7635, pp. 100–114, 2020.
23. Y. Kaya and W. L. Leite, "Assessing Change in Latent Skills Across Time With Longitudinal Cognitive Diagnosis Modeling: An Evaluation of Model Performance," *Educ. Psychol. Meas.*, vol. 77, no. 3, pp. 369–388, 2017, doi: https://doi.org/10.1177/0013164416659314.
24. J. Brooke, "SUS: A Retrospective," *J. Usability Stud.*, vol. 8, no. 2, pp. 29–40, 2013.
25. I. S. G. Baxter, "Socio-technical systems: From design methods to systems engineering," *Interact. Comput.*, vol. 23, no. 1, pp. 4–17, 2011.
26. J. R. Lewis and J. Sauro, "Item Benchmarks for the System Usability Scale," *J. Usability Stud.*, vol. 13, no. January, pp. 158–167, 2018.