INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Optimizing Educational Assessment: The Practicality of Computer Adaptive Testing (CAT) with an Item Response Theory (IRT) Approach

Asrul Huda [a,c,*], Firdaus [a,b,c], Dedy Irfan [a,c], Yeka Hendriyani [a,c], Almasri [a,c], Murni Sukmawati [a,b,c]

[a] *Department Teknik Elektronika, Universitas Negeri Padang, North Padang, Padang, 25131, Indonesia*
[b] *Pendidikan Teknologi dan Kejuruan, Universitas Negeri Padang, North Padang, Padang, 25131, Indonesia*
[c] *Digital Learning Research Center, Universitas Negeri Padang, , North Padang, Padang, 25131, Indonesia*
Corresponding author: [*]asrulhuda@gmail.com

***Abstract*— This research aims to develop a Computer Adaptive Test (CAT) system using the Items Response Theory (IRT) approach. This study is part of developing a web-based system using the Research and Development (R&D) method, employing the Four-D (4-D) model. At its core, this system is similar to a Computer-Based Test (CBT). Still, the critical difference lies in its ability to randomize and provide questions that align with the test-taker's skill levels using the Items Response Theory (IRT) algorithm. The system employs the 3-PL model from the Items Response Theory, considering the difficulty level of questions, the discriminative power of questions, and the likelihood of guessing or interference in the questions. The examination system randomly assigns questions to students based on their responses to previous questions, ensuring that each test-taker receives a unique question sequence. The exam concludes when a test-taker accurately estimates their ability, i.e., SE <= 0.01, or when all questions have been answered. The outcome of this research is a Computer Adaptive Test (CAT) system based on the Items Response Theory (IRT), which can be used to assess students' learning outcomes. This research was implemented in the Multimedia Department of SMK Negeri 1 Gunung Talang, with 90 students as the research sample. The evaluation of the practicality of this system received very high scores, indicating that the Computer Adaptive Test (CAT) system based on the Items Response Theory (IRT) is considered highly practical and effective in achieving the established measurement goals.**

***Keywords*— Computer adaptive test; items response theory; research and development; Four-D Model; educational assessment.**

Manuscript received 8 Oct. 2023; revised 31 Jan. 2024; accepted 26 Feb. 2024. Date of publication 31 Mar. 2024.

I. Introduction

Education is closely related to evaluating or assessing students' learning outcomes [1], [2]. One form of evaluation in education is testing, such as daily quizzes, mid-term exams, and final exams, which aim to measure students' achievement of learning objectives [3]. Meanwhile, evaluation in education is an assessment of students' growth and progress toward the goals and values set in the curriculum [4], [5].

Learning evaluation has three elements that cannot be separated: assessment, measurement, and the tools used [6]–[8]. In an assessment, measurement is required, and measurement requires a measuring instrument [9], [10]. Assessment aims to determine the results of the learning process that has taken place, while measurement is the score obtained from test results. Tests are measuring instruments used in evaluating learning outcomes.

In the era of information technology development, the education sector faces increasingly complex challenges in evaluating students' progress and learning effectiveness [11]–[13]. Constant advancements in technology access and abundant educational data have opened up new opportunities to improve how we measure students' understanding [14], [15]. However, along with this development, equally important testing and educational evaluation issues need to be effectively addressed.

One fundamental problem in educational evaluation is the inability of conventional evaluation methods to consider variations in students' levels of understanding [16]–[19]. In a traditional exam, all students are presented with questions of the same difficulty level. This can be a severe problem because each student has a different level of understanding

[20], [21]. Some students may understand certain concepts well, while others may need further assistance. As a result, exam results often do not accurately reflect the actual abilities of each student.

Furthermore, using time and resources in testing becomes less efficient because students may face questions irrelevant to their abilities [22]. This can also disrupt the overall learning process because students may feel frustrated when faced with questions that are too difficult or boring when faced with too easy questions. This problem not only disadvantages students but also negatively impacts educational decision-making, curriculum design, and teacher performance evaluation [23]. Therefore, there is an urgent need to address this issue and improve educational evaluation methods.

In this context, using a Computer Adaptive Testing (CAT) system integrated with the Item Response Theory (IRT) approach offers an exciting solution. Computer Adaptive Testing (CAT) allows exams to be more adaptive, with the difficulty level of questions dynamically adjusted based on students' performance during the exam [24]–[26]. Students who answer correctly will face more difficult questions, while those who answer incorrectly will face more straightforward questions. This allows for a more accurate measurement of individual student's abilities. On the other hand, the Item Response Theory (IRT) approach allows for a deeper analysis of each exam item [27], [28]. This helps identify the strengths and weaknesses in exam questions, creating better and more relevant tests based on the material taught [9].

In this article, the researcher will explain in more detail how integrating Computer Adaptive Testing (CAT) with the Item Response Theory (IRT) approach can enhance the effectiveness of educational evaluation. The researcher will explore the practical applications of this method in various educational contexts, including classrooms, higher education institutions, and standardized testing. The researcher will describe its benefits in addressing existing evaluation problems, including improving the accuracy of measuring students' abilities, increasing the efficiency of time usage, and enhancing the quality of exam results. We discuss the practicality of implementing Computer Adaptive Testing (CAT) with the Item Response Theory (IRT) approach, as evaluated by teachers and students who use the system to assess learning processes.

II. Materials and Method

This system was implemented at SMK Negeri 1 Gunung Talang, West Sumatra, Indonesia, with a research sample of 90 students. The method used in this research is Research and Development (R&D). Research and Development (R&D) comprises several types of models. The 4-D (Four D) development model is used in product development. The 4-D development model was developed by S. Thiagarajan, Dorothy S. Semmel, and Melvyn I. Semmel and consists of four main stages: Define, Design, Develop, and Disseminate [32]. This method and model were chosen to produce a product as a Computer Adaptive Testing (CAT) System with an Item Response Theory (IRT) approach.

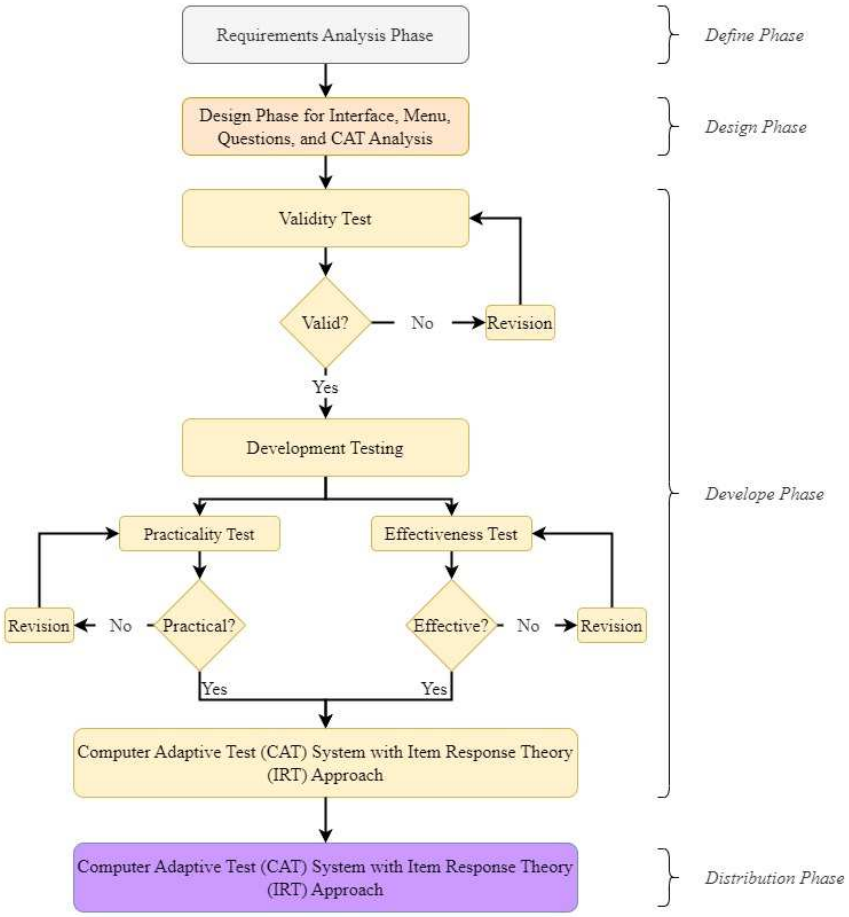

Fig. 1 The 4-D Model Research Design

Based on the research design in Figure 1, the procedure for designing the Evaluation Instrument for the Computer Adaptive Test (CAT) System with the Item Response Theory (IRT) Method consists of several stages, namely:

A. Define Phase

This stage begins with an initial analysis to identify the core issues that need to be addressed in the development process of the CAT evaluation instrument with an IRT approach. Subsequently, a needs analysis is conducted to establish the requirements that must be met by the product to be developed. Analyzing the students assists in understanding their characteristics, and a concept analysis is performed to identify the concepts required for instrument development. Finally, the objectives are analyzed to determine the achievement indicators in the system's development.

B. Design Phase

During this stage, the appropriate medium is selected based on the initial definition: a web-based system with an IRT approach. Subsequently, an initial design is carried out, including the design of system requirements. The use cases for the system can be observed in Figure 2. This process involves selecting technology and infrastructure that supports implementing a web-based system with an IRT approach. Furthermore, this initial design also encompasses the steps required to fulfill the previously identified needs. Figure 2 visually illustrates the system's primary use cases to be developed, providing a clear understanding of how the system will be used and interact with users. This is crucial in ensuring the system's design aligns with the initial objectives established in the preceding phase.

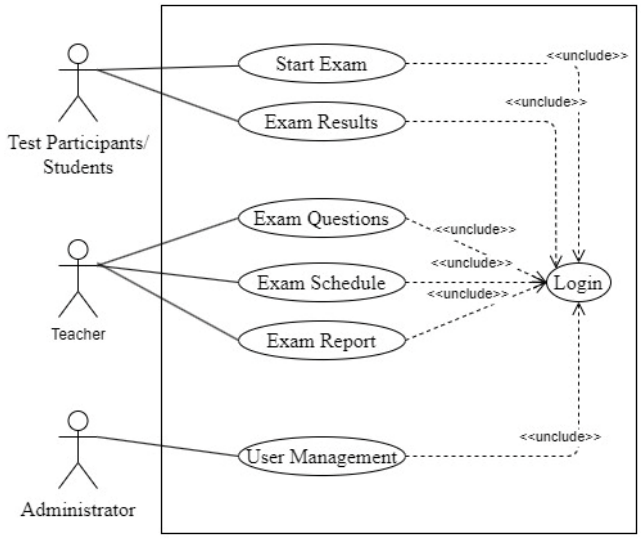

Fig. 2 Use Case Diagram

The use case diagram in the above image illustrates three user levels: test participants/students, teachers, and administrators. Test participants/students can take exams and view their exam results. Teachers can access exam questions, exam schedules, and exam result reports.

C. Development Phase

The development phase involves creating the system product based on the initial design. This product will be validated by expert specialists in their respective field of study, and the product design may be improved based on the feedback from the validators. The method used in the Practicality Test of the Computer Adaptive Test (CAT) System with the Item Response Theory (IRT) Approach employs quantitative techniques, and data collection in this research utilizes a quantitative questionnaire. The aspects assessed by the experts can be found in Table 1.

TABLE 1
PRACTICALITY QUESTIONNAIRE INSTRUMENT FRAMEWORK

No.	Aspect	Number of Statements	Instrument Format
1.	Content and Objectives	1-4 (4 Items)	Likert Scale Questionnaire
2.	Appearance and Process Rules	5-14 (10 Items)	Likert Scale Questionnaire
3.	Usefulness	15-17 (3 Items)	Likert Scale Questionnaire

Source: [9] with modifications

Table 1 outlines the assessment instrument or questionnaire used in the research or evaluation. Three aspects are assessed: Content and Purpose, Presentation and Process Rules, and Utility. Each element is evaluated with a different number of statements or items. Content and Purpose are assessed using four items, Presentation and Process Rules are assessed using ten items, and Utility is evaluated using three items. The instrument used to collect data in this study is the Likert scale questionnaire.

The Likert scale questionnaire is a commonly used tool in research to measure respondents' levels of agreement or disagreement with specific statements using a predefined scale, such as a scale of 1 to 5. This instrument allows researchers to delve into respondents' perceptions and views of the assessed aspects and gain deeper insights into how respondents evaluate the evaluated system's content, presentation, process, and utility.

D. Dissemination Phase

In this phase, the developed product will be publicly uploaded for use by a broader user base. Users are encouraged to provide valuable feedback and suggestions to enhance the product's quality further. Furthermore, the product will be implemented as a pilot project at SMK Negeri 1 Gunung Talang within the scope of this research. The entire sequence of these phases forms a comprehensive procedure for designing and developing the Computer Adaptive Test (CAT) Evaluation Instrument using the Item Response Theory (IRT) Approach.

III. RESULT AND DISCUSSION

A. System Computer Adaptive Test (CAT) based on Item Response Theory (IRT)

The Computer Adaptive Test (CAT) system based on Item Response Theory (IRT) involves three distinct user types with varying roles: Students/Exam Participants, Teachers, and Administrators. To utilize this examination system, every user must possess a registered account and undergo a login procedure to access the available features. This login process entails users entering their account information, consisting of a username and password, on a specially designed login page.

Detailed information regarding the login page's appearance can be found in Figure 3, which has been presented as a visual illustration.

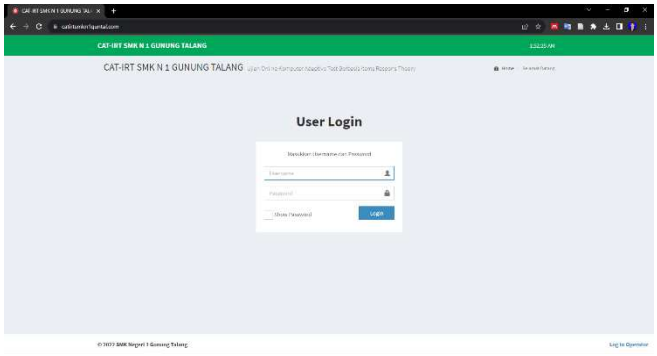

Fig. 3 Login Page Interface

After completing the login process, each user will be directed to an interface tailored to their role, tasks, functions, and permissions within the system. For instance, users with the role of an administrator will have full access to manage other users, as illustrated in Figure 4, which demonstrates the user management functions available to administrators. This allows administrators to perform actions such as creating, deleting, or modifying other users' account information as needed. Furthermore, users with different roles and responsibilities will encounter distinct menus and options within the interface, aligning with their specific requirements and duties in using the system. With this personalized setup, users can easily carry out their tasks without being distracted by features or options irrelevant to their responsibilities.

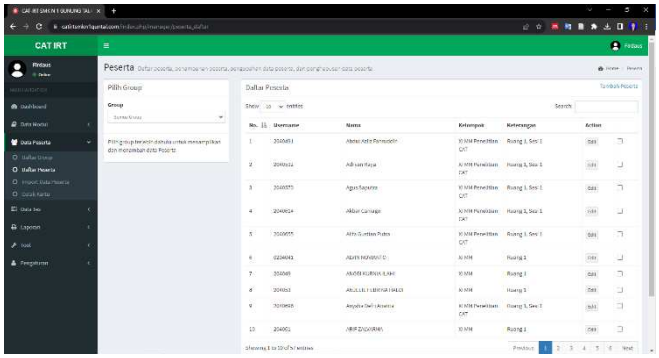

Fig. 4 User Management

Based on Figure 4, users with the role of a teacher have access rights that involve several critical tasks in exam management. Among these access rights, Teachers can manage the collection of exam questions, set exam time parameters, print student exam cards, and view exam results. This indicates that Teachers are central to the testing and evaluation process.

Furthermore, the pages illustrating the user interface for Teachers can be found in Figures 5 to 8. These images provide a visual overview of how Teachers can access and utilize features related to exam management, including adding or editing exam questions, scheduling exams, printing student exam cards, and viewing exam results in more detail. With this specially designed interface, teachers can efficiently carry out their tasks during the testing process according to their needs.

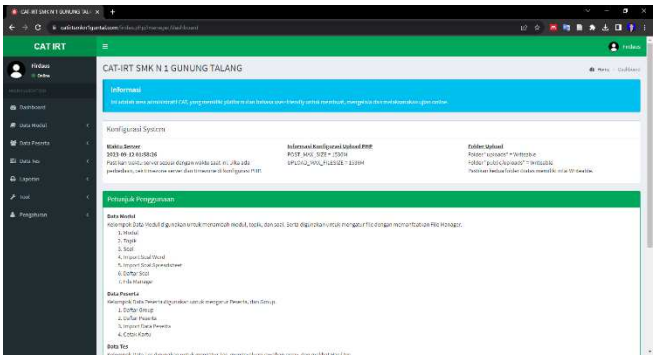

Fig. 5 Teacher Dashboard

The Guru Dashboard Interface serves as the initial screen for teachers when they successfully log into the Computer Adaptive Test (CAT) system based on Item Response Theory (IRT). The Guru Dashboard is purposefully designed to provide easy access and relevant information regarding the management of exams and student evaluations. Within this interface, teachers can swiftly access information about upcoming, ongoing, or completed exams, including details on timing, location, and the list of exam participants. Furthermore, teachers have complete control over exam question management, allowing them to edit existing questions or add new ones to the question bank. Teachers can also easily schedule exams in alignment with their instructional plans, print examination participant cards for students taking the exam, and analyze individual or group exam results. This specially designed interface empowers teachers to efficiently carry out their examination and student evaluation tasks, ultimately supporting enhanced teaching practices.

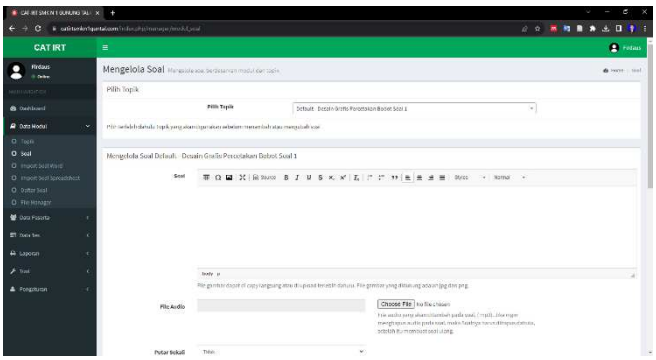

Fig. 6 Question Management

Teachers have full access rights in all aspects of exam question management. They can freely input new questions into the system, edit existing question content, and arrange the availability of questions for each exam session. This gives teachers robust control in ensuring the exams align with the curriculum and predefined learning objectives. With the ability to manage and create various questions, teachers can ensure exam variation and relevance, contributing to a more accurate measurement of students' understanding of the subject matter. In other words, teachers can design exams that suit learning needs and optimize the student evaluation process.

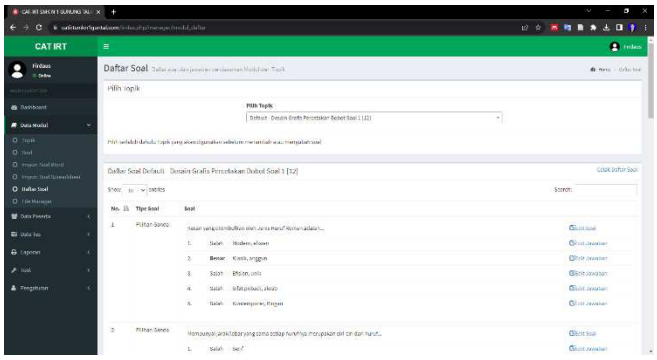

Fig. 7 The Exam Question Data Summary Interface

Figure 7 represents a display that shows the question data that the teacher has input. These questions are stored in the form of a question bank that can be accessed by the teacher when they are creating an exam. The question bank is a storage place for all the questions entered into the system. Teachers can easily access, manage, and select the questions used in exams, whether daily quizzes, mid-term exams, or end-of-term exams. By having access to a question bank rich in various questions, teachers can design exams that match the desired difficulty level and cover multiple aspects of the subject matter. This enables teachers to create diverse and relevant exams that align with the learning objectives and the student's level of understanding. Thus, Figure 7 illustrates how teachers can optimize the use of the question bank to create practical exams for evaluating student progress.

Furthermore, teachers have the privilege of accessing and reviewing detailed exam results for each student who has completed the examination, as depicted in Figure 8.

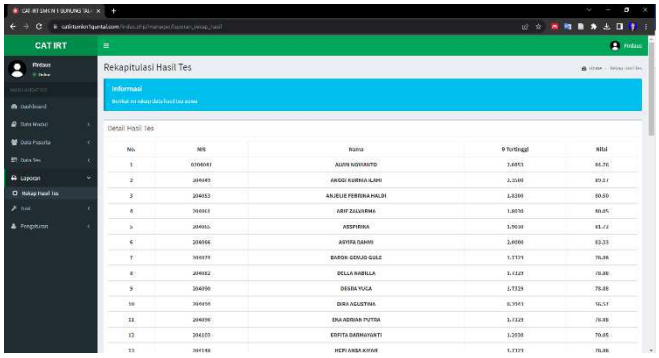

Fig. 8 Exam Result

Based on Figure 8, teachers can view students' exam results with more detailed information than conventional exams. In the Computer Adaptive Test (CAT) system based on Item Response Theory (IRT), the exam results data display the Theta values obtained by each student based on their responses to each question. This data also presents the students' exam scores based on the Theta values they obtained. Furthermore, the student page provides students with the right to take exams based on the exam sessions they are assigned to, as seen in Figure 9.

The Computer Adaptive Test (CAT) system based on Item Response Theory (IRT) is a computerized testing system designed to measure an individual's abilities or knowledge adaptively. This means that the system will adjust the difficulty level of the questions given to test takers based on their performance. This system utilizes an Item Response Theory (IRT) theory to manage this adaptive process. This theory enables the system to measure the test takers' abilities or knowledge levels accurately.

Fig. 9 Student Exam Question Interface

In IRT, each question has specific characteristics such as difficulty level and discrimination power. The system selects questions based on the test takers' responses to previous questions. If a test taker answers correctly, the next question will be more challenging. Conversely, the next question will be easier if a test taker answers incorrectly. This allows the system to measure the test takers' abilities with greater precision and efficiency than conventional tests that provide questions with fixed difficulty levels.

Using the CAT system based on IRT makes the measurement of test takers' abilities or knowledge more accurate and fairer. Each test taker receives a set of questions tailored to their abilities, and no two test takers receive the same sequence of questions. Students will take the exam according to the concept of the Computer Adaptive Test (CAT) system based on Item Response Theory (IRT). This means that students will start the exam with questions of moderate difficulty, and subsequent questions will adjust based on each student's ability to answer them correctly. Further details about the question presentation process can be seen in the flow chart in Figure 10.

From Figure 10, which is the flowchart of the question presentation above, it can be seen that the presentation of questions in this system is carried out through several steps as follows:

1. Test participants begin working on the questions.
2. The system will check if the participant has previously answered questions; if not, the system will display questions with a medium weight of 2.
3. If they have, the system will read the participant's highest and lowest theta values to determine if they have reached the maximum (3) or minimum (-3) limit. If they have, the system decides the exam is finished.
4. If not, the system will read the participant's last answer, whether correct (1) or incorrect (0). If it is accurate, the system will increase the question weight by +1; if it is erroneous, it will decrease the question weight by -1.
5. After obtaining the weight, the system will check if any questions are left that match the participant's weight and haven't been answered. If no more matching questions exist, the system will decide that the exam is finished.

The system will display the following question according to the weight if there are matching questions.

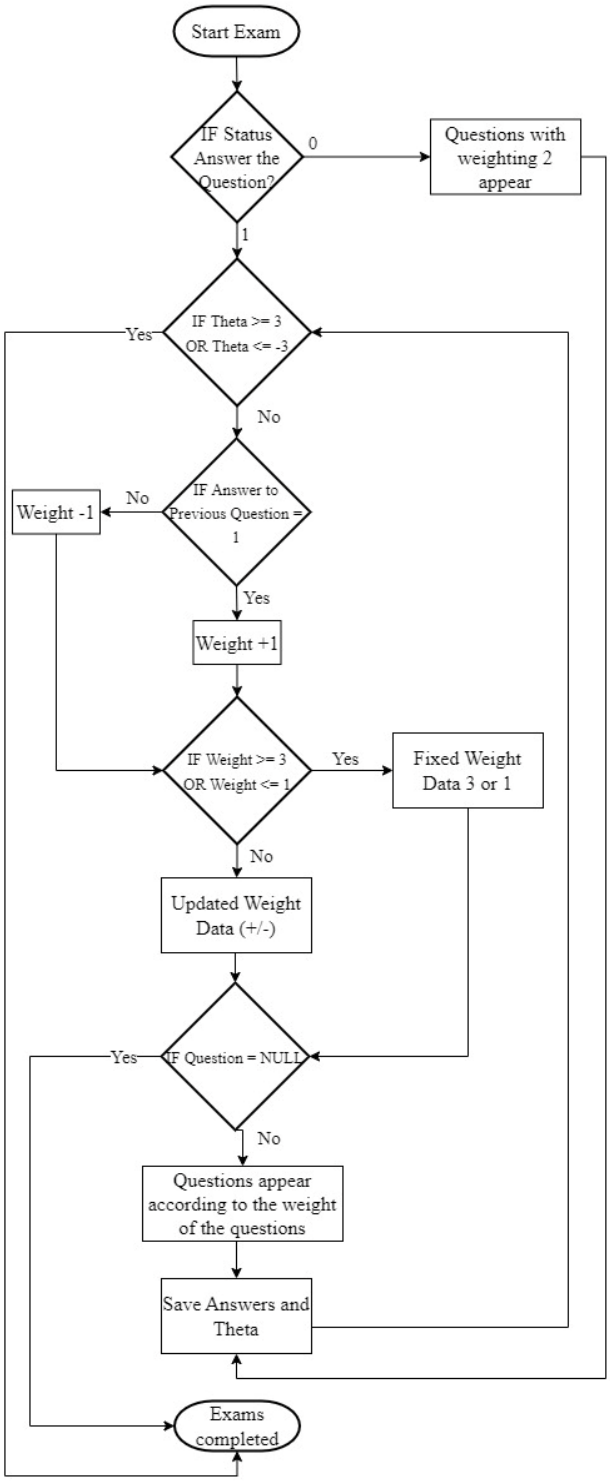

Fig. 10 Flowchart of Question Presentation

The Computer Adaptive Test (CAT) based on Item Response Theory (IRT) differs from the conventional exam score calculation process. In this system, the score calculation begins with the calculation of the Theta value, following these rules.

1. The system checks the participant's answers.
2. Is the score of the participant's answer = correct = 1?
3. If yes, then θ after answering is calculated using Bayesian Estimation method, with the formula:

$$\theta = b_i \frac{1}{Da_i} ln\left(0{,}5\left(1+\sqrt{(1+8c_i)}\right)\right) \quad (1)$$

4. If not, then θ after answering = θ initial.

Next, the process of calculating student scores is carried out based on the answers provided by the students and the Theta values generated by the students themselves, with the following provisions.

1. The system checks all participant answers to see if all student answers are incorrect.
2. If yes, then the final θ = the lowest θ achieved by any participant.
3. If not, then the final θ = the highest θ achieved by any participant.
4. Then the final score is calculated using the formula score $= 50 + \left(\left(\frac{50}{3}\right) x\ final\ \theta\right)$
5. After the final θ and score are calculated, the final θ and score are displayed.

After students have completed their exams, information related to their exam process will be displayed. This information includes the difficulty level of the questions that the students answered, the discrimination index of each question, whether the answers to each question were correct or incorrect, the Theta value of the student after answering each question, and the final score of the student. All of this information is crucial for analyzing students' performance and understanding the extent to which they have grasped the exam material. For example, Figure 11 illustrates a detailed display of this information. With this information, teachers or evaluators can provide more accurate feedback to students and design more effective learning strategies to help students improve their understanding of specific subjects.

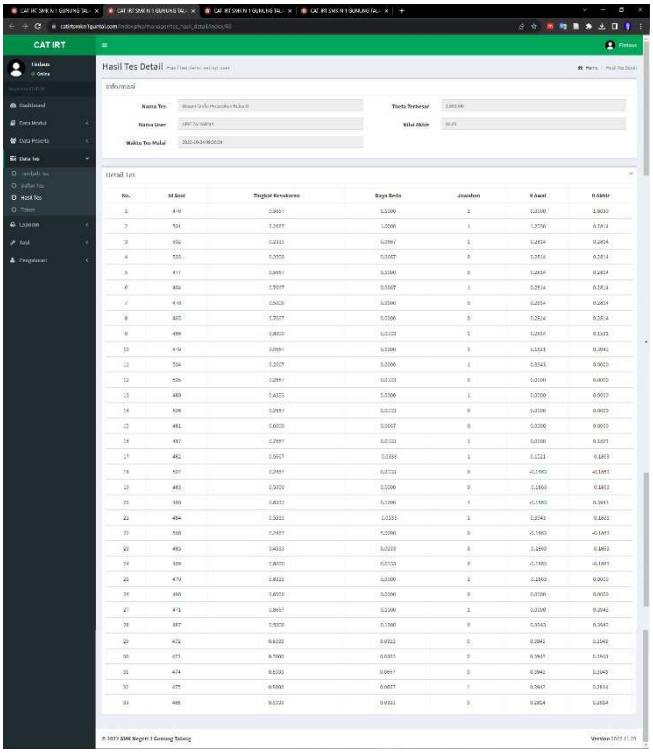

Fig. 11 Detail of one student's examination process

Based on Figure 11, it can be observed that students are not required to answer all the questions to complete the exam.

From Figure 12, it is evident that students have answered only 33 out of the total 50 questions available in the exam. Students can complete the exam following the rules and procedures specified which can be seen in the flowchart of question presentation in Figure 10. This demonstrates the flexibility of the system, where students can finish the exam according to their abilities and the time available to them. In other words, they are not obligated to answer all the questions if they have already achieved a sufficient score or if the exam time has expired in accordance with the applicable rules.

B. Practicality System Computer Adaptive Test (CAT) based on Item Response Theory (IRT)

The aspects assessed in the practicality test consist of 17 questionnaire items divided into three different aspects. There are 4 questionnaire items used to measure the Content and Purpose aspect, 10 questionnaire items to measure the Appearance and Rule Process aspect, and 3 questionnaire items to measure the Usefulness aspect. In the context of this research, the data obtained from 90 students who have used this system are then processed using the SPSS statistical software. The results of the practicality assessment of these 90 students can be found in the distribution of practicality questionnaire scores documented in Table 2.

TABLE II
FREQUENCY DISTRIBUTION OF PRACTICALITY QUESTIONNAIRE SCORES

Interval Class	Frequency	Percentage (%)
71 - 72	1	1%
72 - 73	2	2%
73 - 75	5	6%
75 - 76	4	4%
76 - 78	19	21%
78 - 79	12	13%
79 - 81	47	52%
Total	**90**	**100%**

Table 2, presented above, is a frequency distribution table that depicts data distribution into interval classes along with their frequencies. This table consists of seven interval classes used to group the data, with values ranging from 71 to 81 and varying interval widths. Each interval class has a specific frequency, reflecting the number of data occurrences within that class. For instance, in the interval class 71-72, one data point falls into this category.

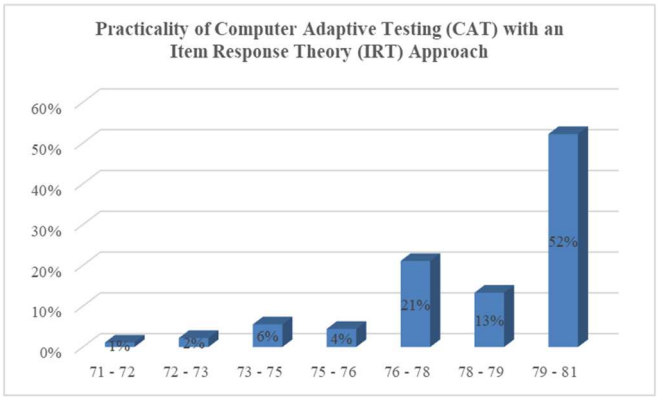

Fig. 12 Histogram Practicality of Computer Adaptive Testing (CAT) with an Item Response Theory (IRT) Approach

Furthermore, the table also provides percentage information, indicating the contribution of each interval class to the total data. As an illustration, the interval class 79-81 has the highest frequency, which is 47, contributing to 52% of the entire dataset. This table offers a clear visual representation of the data distribution and the relative contributions of each interval class to the dataset. This information is precious for statistical analysis and aids in gaining a deeper understanding of the data distribution under evaluation. Thus, this table serves as a crucial tool in comprehending the characteristics of the practical data being assessed. We can refer to the histogram in Figure 12 for a clearer understanding.

In conclusion, the Practicality scores can be summarized in the following Table 3.

TABLE III
THE PRACTICALITY SCORES FOR THE THREE ASPECTS OF THE COMPUTER ADAPTIVE TESTING (CAT) SYSTEM WITH AN ITEM RESPONSE THEORY (IRT)

Aspect	Percentage (%)	Category
Content and Objectives	91	Very Practical
Appearance and Process Rules	90	Very Practical
Usefulness	90	Very Practical

Based on Table 3, it can be concluded that the instrument used in the assessment has a very high level of practicality. The three aspects measured, namely content and objectives, appearance and rule processes, and the utility of the instrument, all received high percentages, namely 91% for content and objectives, and 90% for appearance and utility. The "Very Practical" designation indicates that the Computer Adaptive Test (CAT) System based on Item Response Theory (IRT) is considered very practical and beneficial in achieving the established measurement objectives. This indicates that the instrument suits user needs and can be used optimally in relevant contexts. This conclusion indicates the practicality of the Computer Adaptive Test (CAT) System based on Item Response Theory (IRT) that is used.

IV. CONCLUSION

The Computer Adaptive Test (CAT) system based on Item Response Theory (IRT) has three primary users: students/test takers, teachers, and administrators. Each user must have an account and log in to the system. After a successful login, users have a different interface tailored to their roles and responsibilities. The IRT-based CAT system allows for adaptive testing, where the difficulty level of questions is adjusted according to the test takers' abilities. Each test taker receives a different sequence of questions, and the test concludes when the test taker accurately estimates their ability or when all available questions have been answered. The test results are presented with Theta values that describe the students' abilities.

The results of the practicality assessment indicate that this system is efficient and beneficial. The assessed aspects, including content and objectives, user interface design and rule processes, and the system's usefulness, all received very high scores. Therefore, the Computer Adaptive Test (CAT) system based on Item Response Theory (IRT) is efficient and effective in achieving the established measurement objectives. This conclusion suggests that the system aligns with user needs and can be optimally used in relevant contexts.

AUTHOR CONTRIBUTIONS

Asrul Huda: Design, analysis, Reviewing, Supervision. Firdaus: Conceptualization, Design, Analysis, Writing, Final Approval. Dedy Irfan: Analysis, Editing. Yeka Hendriyani: Analysis, Editing. Almasri: Analysis, Editing. Murni Sukmawati: Analysis, Editing. All authors have approved the final version.

FUNDING

Universitas Negeri Padang funded this research through the Master's Thesis research program with contract number 1391/UN35.15/LT/2023.

ACKNOWLEDGMENT

We thank everyone who helped create this work, especially the Faculty of Engineering Universitas Negeri Padang, LPPM UNP, PTM, and the entire team.

REFERENCES

[1] H. Goss, "Student learning outcomes assessment in higher education and academic libraries: A review of the literature," *J. Acad. Librariansh.*, vol. 48, no. 2, p. 102485, 2022.

[2] H. Y. Suhendi, M. A. Ramdhani, and F. S. Irwansyah, "Verification concept of assesment for physics education student learning outcome," *Int. J. Eng. Technol. UEA*, vol. 7, no. 3.21, pp. 321–325, 2018.

[3] U. Levent and Ş. Ertok, "Student Opinions on task-based approach as formative evaluation versus exam-based approach as summative evaluation in education," *Sak. Univ. J. Educ.*, vol. 10, no. 2, pp. 226–250, 2020.

[4] R. Sun, H. Zhang, J. Li, J. Zhao, and P. Dong, "Assessment-for-learning teaching mode based on interactive teaching approach in college English," *Int. J. Emerg. Technol. Learn. Online*, vol. 15, no. 21, p. 24, 2020.

[5] C. K. Chan and L. Y. Luk, "Academics' beliefs towards holistic competency development and assessment: A case study in engineering education," *Stud. Educ. Eval.*, vol. 72, p. 101102, 2022.

[6] M. Nurtanto, N. Kholifah, A. Masek, P. Sudira, and A. Samsudin, "Crucial Problems in Arranged the Lesson Plan of Vocational Teacher.," *Int. J. Eval. Res. Educ.*, vol. 10, no. 1, pp. 345–354, 2021.

[7] I. W. Mertha and M. Mahfud, "History Learning Based on Wordwall Applications to Improve Student Learning Results Class X IPS in Ma As'adiyah Ketapang," *Int. J. Educ. Rev. Law Soc. Sci. IJERLAS*, vol. 2, no. 5, pp. 507–612, 2022.

[8] K. A. Gamage, D. I. Wijesuriya, S. Y. Ekanayake, A. E. Rennie, C. G. Lambert, and N. Gunawardhana, "Online delivery of teaching and laboratory practices: Continuity of university programmes during COVID-19 pandemic," *Educ. Sci.*, vol. 10, no. 10, p. 291, 2020.

[9] L. A. Clark and D. Watson, "Constructing validity: New developments in creating objective measuring instruments.," *Psychol. Assess.*, vol. 31, no. 12, p. 1412, 2019.

[10] S. J. Eloranta, R. Kaltiala, N. Lindberg, M. Kaivosoja, and K. Peltonen, "Validating measurement tools for mentalization, emotion regulation difficulties and identity diffusion among Finnish adolescents," *Nord. Psychol.*, vol. 74, no. 1, pp. 30–52, 2022.

[11] A. Alam, "Employing adaptive learning and intelligent tutoring robots for virtual classrooms and smart campuses: reforming education in the age of artificial intelligence," in *Advanced Computing and Intelligent Technologies: Proceedings of ICACIT 2022*, Springer, 2022, pp. 395–406.

[12] A. S. Rahmatullah, E. Mulyasa, S. Syahrani, F. Pongpalilu, and R. E. Putri, "Digital era 4.0: The contribution to education and student psychology," *Linguist. Cult. Rev.*, vol. 6, no. S3, pp. 89–107, 2022.

[13] S. Vincent-Lancrin and R. Van der Vlies, "Trustworthy artificial intelligence (AI) in education: Promises and challenges," 2020.

[14] R. Oktrilani, V. I. Delianti, B. R. Fajri, and A. D. Samala, "Rancang Bangun Media Pembelajaran Berbasis Augmented Reality pada Materi Sistem Pernapasan Mata Pelajaran Biologi Kelas XI MIPA Tingkat SMA," JAVIT : Jurnal Vokasi Informatika, pp. 79–86, Jun. 2023, doi:10.24036/javit.v3i2.156.

[15] R. Husna and S. Husna, "Pembelajaran Seluler dan Efek Motivasi dalam Hal Kinerja Guru," JAVIT : Jurnal Vokasi Informatika, pp. 95–101, Jun. 2023, doi: 10.24036/javit.v3i2.114.

[16] S. Mystakidis, A. Christopoulos, and N. Pellas, "A systematic mapping review of augmented reality applications to support STEM learning in higher education," *Educ. Inf. Technol.*, vol. 27, no. 2, pp. 1883–1927, 2022.

[17] R. Martínez-Jiménez and M. C. Ruiz-Jiménez, "Improving students' satisfaction and learning performance using flipped classroom," *Int. J. Manag. Educ.*, vol. 18, no. 3, p. 100422, 2020.

[18] J. Singh, K. Steele, and L. Singh, "Combining the best of online and face-to-face learning: Hybrid and blended learning approach for COVID-19, post vaccine, & post-pandemic world," *J. Educ. Technol. Syst.*, vol. 50, no. 2, pp. 140–171, 2021.

[19] L. Clouder, M. Karakus, A. Cinotti, M. V. Ferreyra, G. A. Fierros, and P. Rojo, "Neurodiversity in higher education: a narrative synthesis," *High. Educ.*, vol. 80, no. 4, pp. 757–778, 2020.

[20] H. Gardner, *Disciplined mind: What all students should understand*. Simon & Schuster, 2021.

[21] P. Putra, F. Y. Liriwati, T. Tahrim, S. Syafrudin, and A. Aslan, "The students learning from home experience during covid-19 school closures policy in indonesia," *J. Iqra*, vol. 5, no. 2, 2020.

[22] J. G. Nguyen, K. J. Keuseman, and J. J. Humston, "Minimize online cheating for online assessments during COVID-19 pandemic," *J. Chem. Educ.*, vol. 97, no. 9, pp. 3429–3435, 2020.

[23] H.-C. Lin, G.-J. Hwang, S.-C. Chang, and Y.-D. Hsu, "Facilitating critical thinking in decision making-based professional training: An online interactive peer-review approach in a flipped learning context," *Comput. Educ.*, vol. 173, p. 104266, 2021.

[24] S.-C. Cheng, Y.-P. Cheng, and Y.-M. Huang, "To implement computerized adaptive testing by automatically adjusting item difficulty index on adaptive English learning platform," *J. Internet Technol.*, vol. 22, no. 7, pp. 1599–1607, 2021.

[25] B. Yessingeldinov, Z. Rakhymbayeva, A. Zhapparova, and L. Tursynova, "Utilising a differentiated assessment method in mathematics class: computer adaptive testing for tracking student progress," *Glob. J. Eng. Educ.*, vol. 25, no. 1, 2023.

[26] M. Boussakuk, A. Bouchboua, M. El Ghazi, M. El Bekkali, and M. Fattah, "Design of Computerized Adaptive Testing Module into Our Dynamic Adaptive Hypermedia System," *Int. J. Emerg. Technol. Learn. IJET*, vol. 16, no. 18, pp. 113–128, 2021.

[27] K. Pliakos, S.-H. Joo, J. Y. Park, F. Cornillie, C. Vens, and W. Van den Noortgate, "Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems," *Comput. Educ.*, vol. 137, pp. 91–103, 2019.

[28] P.-C. Bürkner, "Bayesian item response modeling in R with brms and Stan," *ArXiv Prepr. ArXiv190509501*, 2019.

[29] I. Lee, S. Grover, F. Martin, S. Pillai, and J. Malyn-Smith, "Computational thinking from a disciplinary perspective: Integrating computational thinking in K-12 science, technology, engineering, and mathematics education," *J. Sci. Educ. Technol.*, vol. 29, pp. 1–8, 2020.

[30] M. O. Ogunjimi, M. A. Ayanwale, J. I. Oladele, D. S. Daramola, I. M. Jimoh, and H. O. Owolabi, "Simulated evidence of computer adaptive test length: Implications for high stakes assessment in Nigeria," *J. High. Educ. Theory Pract.*, vol. 21, no. 2, pp. 202–212, 2021.

[31] M. A. Sorrel, J. R. Barrada, J. de la Torre, and F. J. Abad, "Adapting cognitive diagnosis computerized adaptive testing item selection rules to traditional item response theory," *Plos One*, vol. 15, no. 1, p. e0227196, 2020.

[32] D. Sugiyono. *Metode penelitian pendidikan pendekatan kuantitatif, kualitatif dan R&D*. Bandung, Indonesia: Alfabeta, 2013.