Journal of Educational Research on Children, Parents & Teachers, Volume 5, Number 2, 2024, 310-324
ISSN: 2664-3812, https://ercptjournal.org/

APPLICATION OF ITEM FACILITY INDICES IN THE DEVELOPMENT OF CAT TEST THROUGH A SUPERVISED MACHINE LEARNING TECHNIQUE IN EDUCATIONAL RESEARCH

Christian Ngozi[1]; Moses Offor[2]; Emmanuel Sunday Odo[1]. & Christian S. Ugwuanyi[3]
[1]Department of Science Education, University of Nigeria, Nsukka
[2]Faculty of Education, Kogi State University, Kogi State
[3]Faculty of Education, University of the Free State, Bloemfontein, South Africa
Corresponding email: Ngozi.christian.pg00252@unn.edu.ng

Abstract
A lot of academics in educational research have been unable to link up with usage of artificial intelligence towards the actualization of quality education (SDG 4) goal due to a gap on the technical knowledge required for AI development and deployment in educational assessment. Therefore, researchers carried out a study titled "application of Item facility indices in development of CAT test adaptive logic following a supervised machine learning technique in educational research" This study employed a Research and Development (R&D) design to create a Computer Adaptive Test (CAT) utilizing achievement tests and a concise decision tree based on item facility indices. The sample consists of 50 participants, with a reliability index of 0.74. The X-calibre using IPLM was used to ascertain and categorize test items based on the b-parameter, representing item difficulty. Items were classified as low difficulty ($b < -1$), moderate difficulty ($-1 \leq b \leq 1$), and high difficulty ($b > 1$). Easy items included Items 1, 4, 5, 6, 7, 9, 21, 27, 30, 39, 41, 44, and 45, while moderate items comprised Items 3, 8, 10, 11, 13, 14, 15, 16, 22, 23, 24, 26, 32, 38, 40, 42, 43, 46, 47, 48, and 50. Hard items included Items 2, 12, 17, 18, 19, 20, 25, 29, 34, 35, 36, 37, and 49. Facility indices were used to construct a decision tree for adaptive logic, categorizing b-parameters as easy ($-4 \leq x \leq -1$), moderate ($-0 \leq x \leq 0$), and hard ($+1 \leq x \leq +4$). The system dynamically adjusts item difficulty based on examinee responses, tailoring the test to match individual abilities. This approach exemplifies how facility indices can be leveraged to develop an AI-driven adaptive assessment tool, improving testing efficiency and accuracy by personalizing the experience for each examinee. The results highlight the potential of AI technologies in revolutionizing educational assessment systems.

Keywords: Computer Adaptive test, Artificial intelligence, Facility indices & CAT design

INTRODUCTION

The United Nations Sustainable Development Goals (SDGs) place a significant emphasis on the integration of technology, particularly artificial intelligence (AI), in achieving quality education (SDG 4). AI has been increasingly recognized as a transformative tool that can improve educational outcomes by personalizing learning, enhancing teacher support, and providing accessible education to marginalized communities (United Nations Educational, Scientific Cultural Organization, 2020). AI-driven adaptive learning systems offer personalized learning paths for students, addressing their individual strengths and weaknesses (Organization for Economic Co-operation Development, 2019). Moreover, AI enables teachers to focus on higher-order tasks by automating administrative duties and facilitating differentiated instruction (Schiff, 2020). In developing regions, AI applications help bridge the educational gap by providing remote access to learning materials and instruction (United Nations Children Fund, 2021). Ethical AI practices in education, as advocated by the United Nations, focus on ensuring inclusive and equitable learning environments (UNESCO, 2021). Additionally, AI contributes to vocational education, equipping learners with the digital skills needed in the evolving job market (World Economic Forum, 2018). Despite these advances, the UN emphasizes that AI in education must be aligned with principles of fairness, transparency, and data privacy (UNESCO, 2019).

Artificial intelligence (AI) has become an integral part of educational assessment, offering innovative approaches to evaluating student performance. One of the key benefits of AI in assessment is its ability to provide real-time feedback through automated grading systems. AI-driven tools like natural language processing (NLP) algorithms can assess written responses more consistently and objectively than human graders (Zawacki-Richter et al., 2019). These systems also enable scalability, allowing institutions to assess a large number of students efficiently (Luckin, Holmes, Griffiths, & Forcier, 2016). Furthermore, AI can analyze patterns in student responses to identify learning gaps, making the assessment process not only summative but also formative (Heffernan & Heffernan, 2014). This data-driven approach to assessment can significantly improve the personalization of learning by recommending specific interventions based on the student's performance (Shute & Ventura, 2013).

Additionally, AI in educational assessment facilitates adaptive testing, which adjusts the difficulty level of questions in real-time based on the learner's ability. This method enhances the precision of assessments by focusing on the most appropriate questions for each learner (Chen, 2019). AI also plays a critical role in reducing biases in traditional assessments by ensuring that algorithms are based on objective data rather than human intuition, thus promoting fairness in evaluation (Holmes, Bialik, & Fadel, 2019). Furthermore, AI can help monitor the psychometric properties of assessment tools, ensuring reliability and validity over time (Baker & Smith, 2019). Despite these advantages, the integration of AI in assessment also raises concerns about data privacy, ethics, and the potential for algorithmic bias, which underscores the importance of developing fair, transparent, and ethical AI assessment systems (Williamson, 2018).

Ignorance about the methods of implementing artificial intelligence (AI) in educational assessment can hinder the effective integration of AI-driven tools and technologies in the educational system. A lack of understanding of AI's capabilities and limitations often leads to skepticism among educators and administrators, preventing its widespread adoption (Holmes, Bialik, & Fadel, 2019). Many educators may be unfamiliar with how AI-based systems, such as automated grading and adaptive testing, function or how to align them with pedagogical goals (Zawacki-Richter et al., 2019). This gap in knowledge results in underutilization or improper use of AI tools, reducing their potential benefits. Moreover, teachers may fear that AI will replace their roles rather than enhance their capacity to provide personalized instruction and feedback (Luckin, Holmes, Griffiths, & Forcier, 2016). Consequently, professional development programs that focus on AI literacy and its application in educational contexts are essential for overcoming these challenges (Williamson, 2018).

Furthermore, ignorance regarding the ethical and technical considerations of AI in assessment can lead to the misuse of these systems. Without a deep understanding of how AI algorithms work, educators may overlook important factors like algorithmic bias, data privacy, and the validity of AI-driven assessments (Baker & Smith, 2019). Inaccurate or biased algorithms can disadvantage certain groups of students, exacerbating educational inequalities (Holmes, Bialik, & Fadel, 2019). Additionally, the lack of awareness of AI's role in adaptive learning or formative assessment means that educators might only focus on summative uses, missing opportunities for continuous feedback and personalized learning (Heffernan & Heffernan, 2014). Bridging this knowledge gap requires collaboration between AI experts and educational practitioners to create frameworks that are transparent, fair, and aligned with pedagogical best practices (Chen, 2019).

Perceived difficulty in implementing artificial intelligence (AI) in educational assessment arises from a variety of technical, pedagogical, and infrastructural challenges. One of the key obstacles is the complexity of AI technologies themselves. Many educators and administrators lack the technical expertise to understand or effectively implement AI tools, such as machine learning algorithms, adaptive testing systems, and automated grading software (Holmes, Bialik, & Fadel, 2019). The integration of these systems often requires significant technical support, specialized training, and a rethinking of traditional assessment models, making the implementation process appear daunting (Zawacki-Richter et al., 2019). Furthermore, the customization of AI-driven tools to fit specific educational contexts, curricula, and learning goals is a challenge that many institutions find difficult, further contributing to the perception that AI is too complex for practical use in assessments (Chen, 2019).

In addition to technical hurdles, the perceived difficulty of implementing AI in educational assessment is compounded by concerns related to infrastructure and cost. Many schools, particularly in developing regions, lack the necessary digital infrastructure to support AI-based assessments, such as high-speed internet, adequate hardware, and reliable software systems (Luckin, Holmes, Griffiths, & Forcier, 2016). Moreover, the cost of AI solutions, both in terms of initial investment and ongoing maintenance, can be prohibitive for many educational institutions, leading to a perception that AI is not a feasible option (Baker & Smith, 2019). These barriers also intersect with concerns about data privacy, ethical considerations, and the need for ongoing professional development, which can make AI adoption seem overwhelming (Williamson, 2018). As a result, even when educators recognize the potential benefits of AI in assessments, the perceived difficulty of implementation often leads to resistance or hesitation in adopting these technologies (Shute & Ventura, 2013).

The lack of training programs available for implementing artificial intelligence (AI) in educational assessment in Nigeria presents a significant barrier to adopting these advanced technologies in the education sector. One of the main challenges is the limited availability of AI literacy programs designed for educators and school administrators, which leaves many ill-equipped to understand or integrate AI tools into their assessment practices (Okoye et al., 2021). Without adequate training, teachers are unable to harness the potential of AI-driven technologies like automated grading systems, adaptive learning platforms, and data analytics for formative assessment (Ade-Ojo, 2019). This lack of training is compounded by the minimal inclusion of AI-related courses in teacher education and professional development programs in Nigeria, further widening the gap between the current teaching methods and AI-driven innovations (Azeez et al., 2022). Consequently, many educators remain unaware of how AI can enhance their teaching and assessment strategies, perpetuating a cycle of reliance on traditional assessment models.

Moreover, the absence of institutional support for AI training programs exacerbates the issue. While some private institutions may offer limited AI-related workshops, there is a general lack of government-backed initiatives to promote the development of AI skills in the educational sector (Ogunode, 2021). Without a robust framework for AI education, Nigerian educators face challenges in staying updated on global trends in educational technology (Ajayi & Fashola, 2020). Moreover, the few existing AI training programs are often inaccessible to teachers in rural areas, where internet connectivity and access to technology are limited (Babalola, 2020). To bridge this gap, concerted efforts are needed from both the government and private sector to develop and implement training programs focused on AI in education, ensuring that all educators are empowered to use these technologies effectively (Uwadia & Adebayo, 2018). Thus, this study intends to provide the application of Item facility indices in development of Computer Adaptive Test adaptive logic following a supervised artificial intelligence machine learning technique. Little or no literature has been provided on this study, this informs the gap that the current study intends to fill.

Purpose of the study

The main purpose of the study was to examine the application of Item facility indices in development of Computer Adaptive Test (adaptive logic) following supervised artificial intelligence technique in educational research. Specifically, the study sought to find out the:

1. facility indices of the test items used for CAT development in Educational Research?
2. way to organize the facility indices for CAT development in Educational Research?

Research Questions

1. What are the facility indices of the test items used for Computer adaptive test (CAT) development in Educational Research?

2. How are the facility indices organized for CAT assessment logic in educational research

Method

Design of the study: The study employed Research and Development (R&D) because it was concerned with the development, deployment and testing of the Computer Adaptive test software. Research and Development (R&D) is a systematic approach to innovation, improvement, and problem-solving, encompassing both research (theoretical and empirical investigation) and development (application and implementation of solutions) (Baxter, 2017). As a research design, R&D involves a cyclical process of identifying needs, conceptualizing solutions, testing and refining prototypes, and implementing and evaluating outcomes (Crossan & Apaydin, 2010). By integrating research and development, R&D facilitates the translation of theoretical knowledge into practical applications, driving innovation and competitiveness (Freeman & Soete, 2009). In the context of computer adaptive testing, R&D enables the development of more efficient, accurate, and user-friendly assessment systems, ultimately enhancing the testing experience and outcomes.

CAT Design: this involves the process of *Algorithm Development.* The description of the development of the adaptive testing algorithm, including the mathematical models and programming languages the following steps provide a complete picture:

Steps on the design of the AI driven CAT

1. The AI approach used for CAT: The researchers utilized the pencil approach to develop the Concise decision tree model that would make use of smaller item bank. This is because popular decision trees can spend about 1,500 to 750 items for a bank just to administer a 50-item achievement test.
2. The machine learning approach utilized by the researchers: While engaging in the use of educational assessment the researchers adopt the use of supervised machine learning technique, because in this technique the solution or the answer of the item is already revealed to the computer. Hence, the researchers made use of CTT and IRT to train the data.
3. The type of data used for the AI driven CAT: The researchers made use of item difficulty parameter to develop the test. However, other parameters such as the discrimination, distractor, guessing, parameter was not active in the work. This is because the adaptive item selection method was based on “a stratification without blocking method” where items are navigated based on their difficulty level. According to Hans (2016) The reliance on the `b` parameter (item difficulty) in a stratification without blocking method in Computerized Adaptive Testing (CAT) is primarily because this method aims to balance item selection across a wide range of difficulties, ensuring a fair and representative test. The `b` parameter allows the test to cover various difficulty levels without focusing on the precision of ability estimation that other parameters like `a` (discrimination) and `c` (guessing) provide. This approach simplifies the item selection process, making it computationally efficient and effective in maintaining a balanced test, which is crucial for content fairness and representativeness across diverse test-taker populations. The item selection method actually makes decision tree become operational.

4. How the researchers visualized the logic: The researchers made use of word insert tools to develop a visual representation of the adaptive logic.
5. How the researchers transformed the logic to an AI program: The researchers first transformed the visual representation of the adaptive logic to a pseudocode, after which they convert them to the proper program code. The researchers transformed the adaptive logic to a python code using Visual studio code and Python IDE (3.11; 64 Bit). This is the stage of the software development, see figure 4.0

6. Converting python software to executable: The software was converted from a python file (.py) to a executable file (exe.) This was done with the assistance of a software expert using their resources.

However, in the process of CAT design, the researchers engaged on Item response theory to train data and get the psychometric properties of the item. The researchers plugged the items to their respective categories such as easy, moderate and hard using the scales -4 to -1 for easy, -0 to +0 for moderate and +1 to +4 for hard. Afterwards, the researchers developed a graphical chart of the concise decision tree (CDT) for the adaptive logic. The researchers converted the decision tree to a pseudocode. Then, the researchers choose the python language for the program code. The researchers downloaded and utilized the Visual Studio Code for the CAT business logic (Adaptive logic) development and deployment.

The Adaptive Logic

The following chart formed the business logic that guides the sequence upon which the instrument is administered.

The subscript numbers 1,2,3 refers to the sorting of items based on their categories. This shows how items are retrieved by the system from the item bank. Some items would be presented to the examinee before other items. All items from the same category cannot be presented to the examinee at the exact same time. The arrows show the movement of the examinee across items. If the examinee passes the item, they move right (more difficult items). If they fail, they move left (less difficult items). The levels show the process upon which the examinee follows, he begins with a moderate item, in level two he proceeds to either an easy or hard item, in level three he is exposed to either easy, moderate or difficult item, the same occurs in level four. But in level five a stopping rule is provided, showing items the respondents would have to answer before the test comes to an end irrespective of the magnitude of items available.

The model predicts that a test-taker is more likely to answer correctly if their ability exceeds the item's difficulty level. This predictive power allows the CAT system to adaptively select items that are neither too easy nor too difficult, providing a tailored assessment experience. By focusing on the “b” parameter, the system simplifies the prediction of item responses, making real-time item selection more computationally efficient. Using only the “b” parameter, the model estimates a test-taker's ability by comparing their responses to items of varying difficulty levels. As test-takers answer items correctly or incorrectly, the model updates the ability estimate, seeking a level

Journal of Educational Research on Children, Parents & Teachers, Volume 5, Number 2, 2024, 310-324
ISSN: 2664-3812, https://ercptjournal.org/

where the probability of correct responses aligns with expectations based on item difficulty.

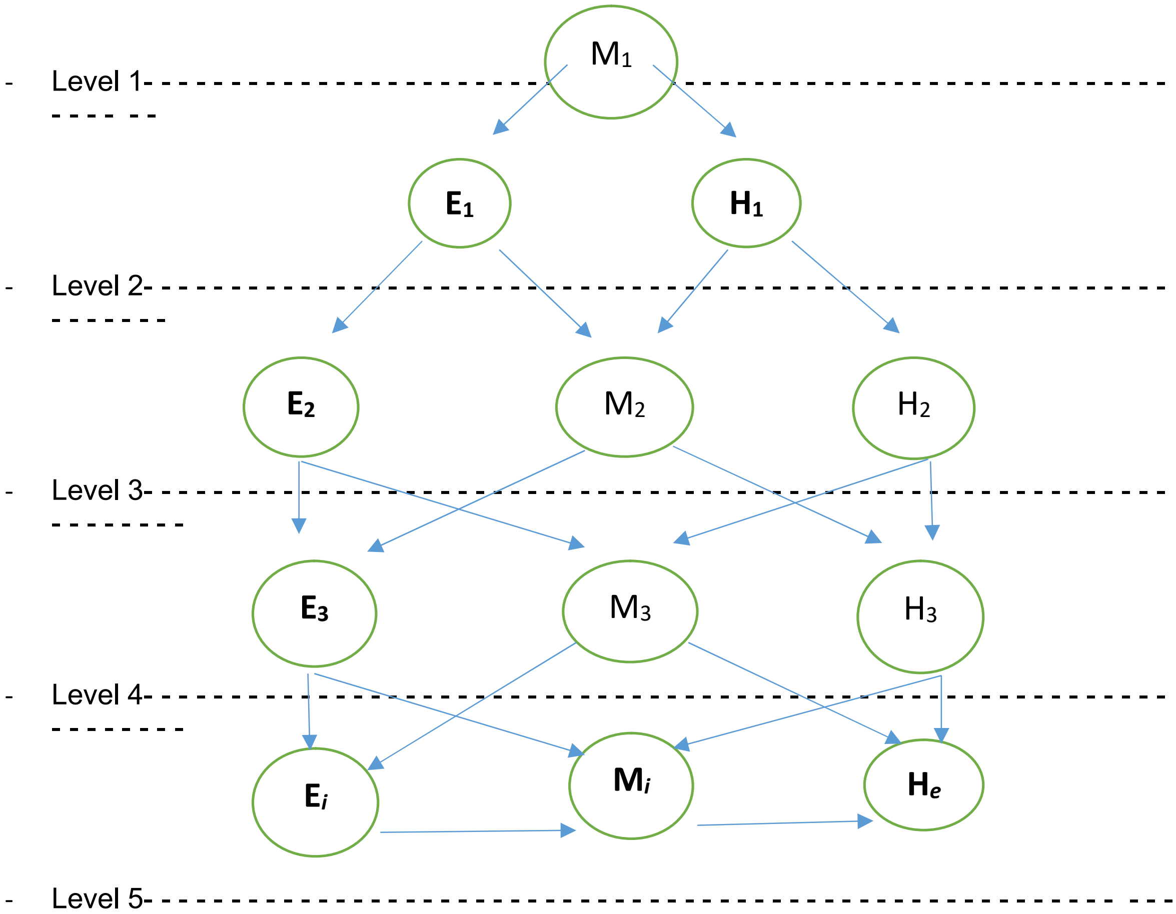


Figure 1 above showing a graphical representation of the adaptive logic.

E = Easy item, M = Moderate Item, H = Hard item.

Reliability: The instrument was trial tested using 50 education students from Enugu State University of Science And Technology (ESUTH) from the faculty of education. The internal consistency of the instrument was determined using the Microsoft Excel statistical tool and an index of .74 was obtained using KR-20; this indicates that the instrument was reliable for the study as noted by George (2020) whose benchmark is 0.70 and above.

Method of Data Analysis: The study made use of X-calibre to determine the psychometric properties of the Items are the a,b & c parameters.

Results

Research Question One: What are item facility indices of the test items used for Computer adaptive test (CAT) development in Educational Research?

Table 1: Difficulty Parameters of the Items Estimated from 2PLM

Easy Items		Moderate Items		Hard Items	
Item ID	B	Item ID	B	Item ID	B
001	-4.000	003	-0.985	002	1.430
004	-1.454	008	0.015	012	2.424
005	-1.519	010	-0.212	017	2.553
006	-1.357	011	0.636	018	3.272
007	-1.348	013	-0.577	019	1.996
009	-1.141	014	-0.328	020	1.378
021	-1.151	015	0.423	025	1.217
027	-3.457	016	0.270	029	1.776
030	-1.391	022	-0.772	034	1.355
039	-1.154	023	0.357	035	3.036
041	-1.172	024	-0.590	036	3.397
044	-1.009	026	-0.791	037	3.625
045	-3.300	032	-0.454	049	2.740
		038	-0.544		
		040	0.002		
		042	-0.085		
		043	0.062		
		046	0.235		
		047	-0.212		
		048	0.969		
		050	0.969		

The scale for categorizing parameters is: Low difficulty: $b < -1$, Moderate difficulty: $-1 \le b \le 1$. High difficulty: $b > 1$. The Easy Items include Items 1, 4, 5, 6, 7, 9, 21, 27, 30, 39, 41, 44 & 45. On the other hand, Moderate Items include 3, 8, 10, 11, 13, 14, 15, 16, 22, 23, 24, 26, 32, 38, 40, 42, 43, 46, 47, 48, and 50. In addition, Hard Items includes 2, 12, 17, 18, 19, 20, 25, 29, 34, 35, 36, 37 and 49.

Journal of Educational Research on Children, Parents & Teachers, Volume 5, Number 2, 2024, 310-324
ISSN: 2664-3812, https://ercptjournal.org/

Research Question Two: How are the facility indices organized for CAT assessment logic in educational research?

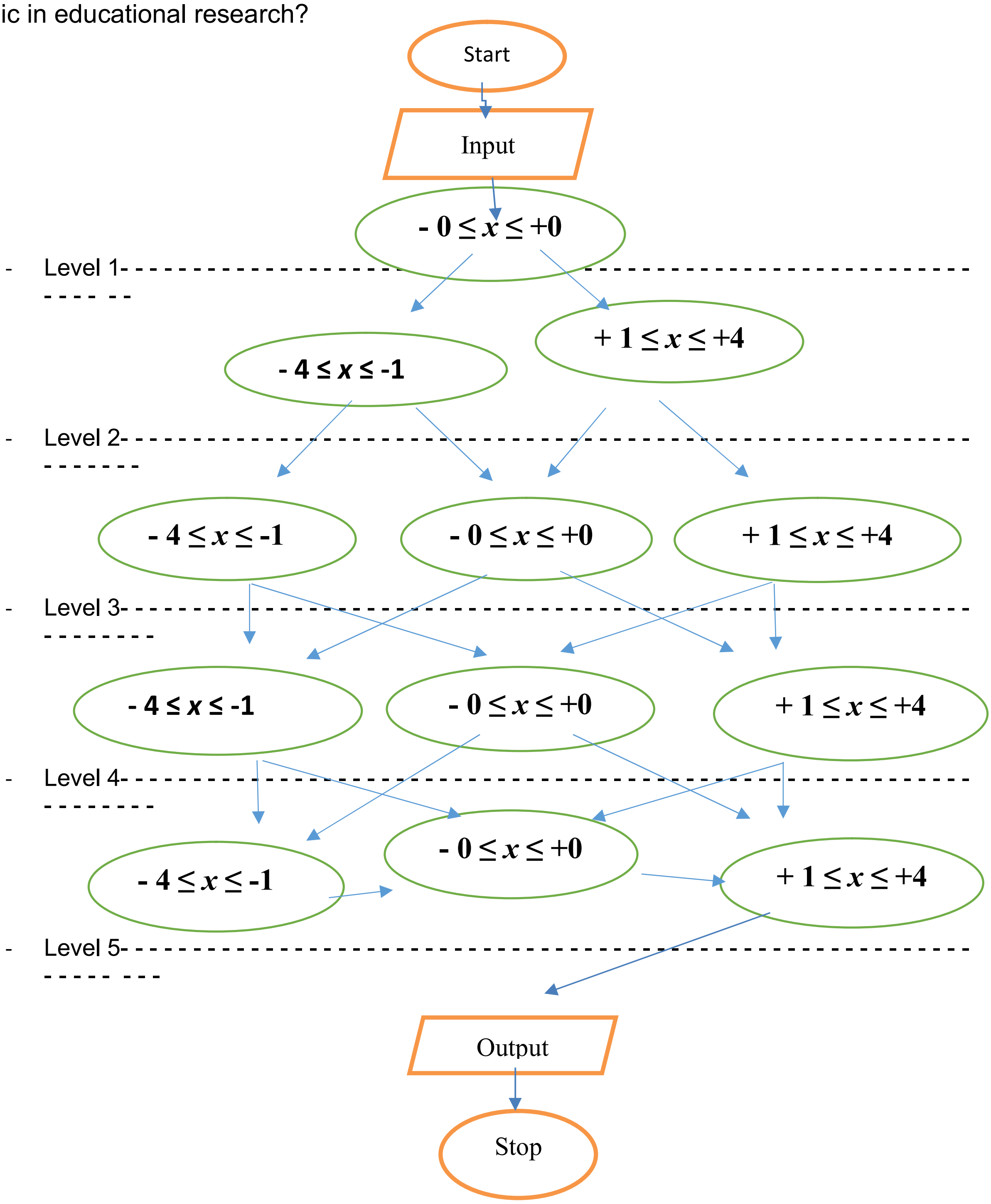


Figure 1. above showing a graphical representation of how the facility indices from table 1. is organized to form the CAT adaptive logic

Findings from above shows how the facility indices are organized to create the decision tree for the adaptive logic. Based on the nature of the facility indices collected from the b distributions, The b parameters are categorized as $-4 \le x \le -1$ as easy items, $-0 \le x \le +0$ as moderate Items and $+1 \le x \le +4$ as hard items, The arrow shows how the computer would intelligently navigate examinees across items and their difficulty levels from the start till the finish of the test process to tailor the assessment to fit their ability. This exemplifies how the facility indices would be used to develop a computer adaptive test which functions as a form of Artificial intelligence technology.

Discussion of Findings

Item facility indices of the test items used for Computer adaptive test (CAT) development in Educational Research

Regarding difficulty, the study divides the items into easy, moderate, and hard categories based on the b-parameter. Easy items, including 1, 4, and 5, fall below a difficulty level of -1, suggesting that these items are suitable for examinees with lower ability levels (De Ayala, 2009). Moderate difficulty items such as 8, 10, and 11 have values between -1 and 1, while hard items like 12, 17, and 49 exceed 1, indicating that they are more challenging for respondents. Items with a high level of difficulty are crucial in assessments designed for high-ability individuals, but they may present challenges if used in tests targeting a more general population (Embretson & Reise, 2013).

The facility indices are filtered in such a manner that must reflect in the data storage layer of the CAT test using a layered approach. The categories of the items based on difficulty are made to enable real time estimation of test taker ability. This is because CAT systems require real-time data processing capabilities to adapt to the test-taker's performance dynamically (Ceri, 2016). MySQL, combined with in-memory processing and caching techniques, can handle real-time data ingestion and analytics, providing immediate feedback and updates to the system (Han, Pei, & Kamber, 2011). Stored procedures and triggers in MySQL can be used to automate and optimize complex data operations, such as recalculating ability estimates and selecting the next test item based on real-time performance data (Leis, Boncz, & Kemper, 2016). See below how item facility indices are used to categorize items to easy, moderate & hard items within a broader outlook of the CAT system.

Journal of Educational Research on Children, Parents & Teachers, Volume 5, Number 2, 2024, 310-324
ISSN: 2664-3812, https://ercptjournal.org/

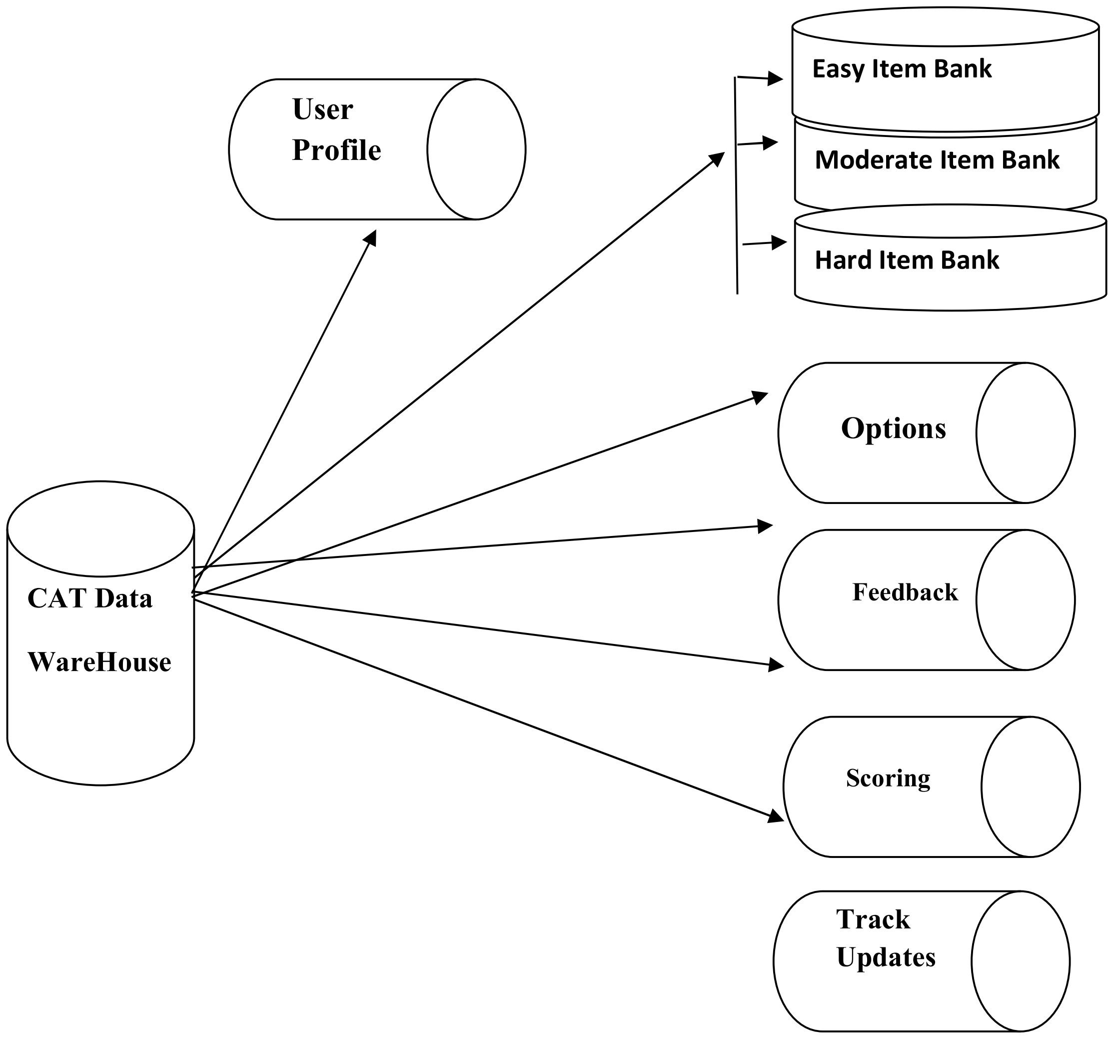


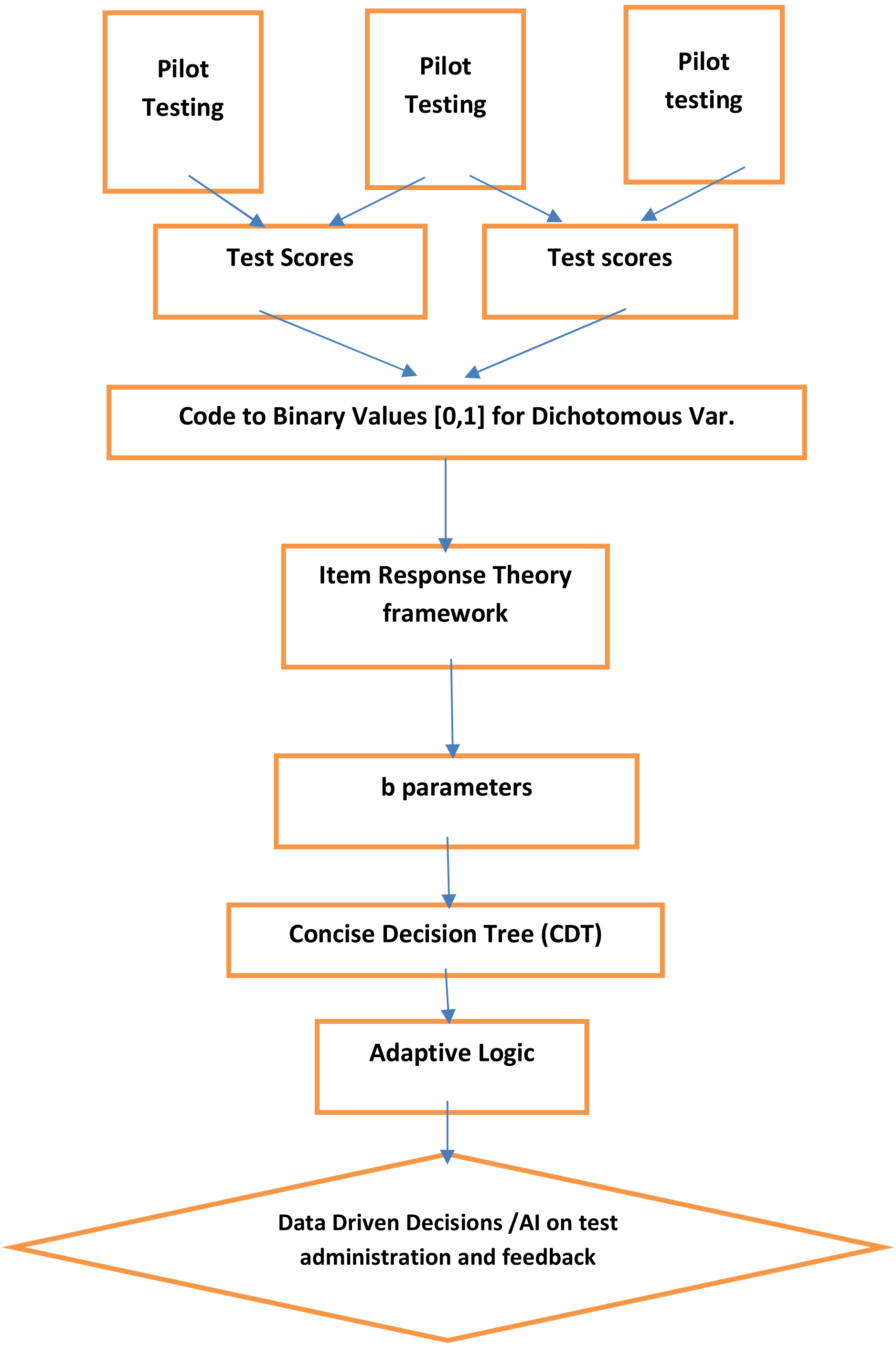
Figure 2. SQL for CAT Database Management

The organization of the facility indices for CAT assessment in educational research

The facility indices are crucial in developing the decision trees used to provide the assessment logic for the adaptive test, this is known as the adaptive logic. Without proper organization of the facility indices in a logical structure, the decision trees could not be formed. Decision trees play a crucial role in the development of computer adaptive testing (CAT) systems by facilitating intelligent decision-making, which directly impacts the software's autonomy (Breiman, 2017). In a CAT system, decision trees are used to determine the most appropriate test items based on a test-taker's responses and estimated ability levels. This ability to dynamically select and adapt test items based on real-time data allows CAT software to operate autonomously, tailoring the testing experience without requiring manual intervention (Van der Linden & Glas, 2010).

The decision-making process inherent in decision trees ensures that the CAT system can independently adjust its behavior to better assess a test-taker's skills. The intelligent decision-making capability of decision trees enhances the autonomy of CAT software by enabling it to make complex decisions based on statistical analysis of test-taker responses (Mitchell, 2017). Each node in a decision tree represents a decision point where data is evaluated to determine the next course of action, such as selecting a new test item (Quinlan, 2014). This hierarchical decision-making process allows CAT systems to autonomously adjust the difficulty of questions and provide a personalized testing experience, improving the efficiency and accuracy of the assessment (Baker & Inventado, 2014). As a result, CAT systems can operate with minimal human oversight, ensuring that the assessment process remains adaptive and responsive to individual test-taker needs. Below shows a diagram on the process of using the facility indices (b parameter) for creating the adaptive logic.

Adaptive Logic Development Framework



Conclusion

The study successfully demonstrates how the b-parameter distributions can be categorized and utilized to develop an intelligent, adaptive testing system. The scale of difficulty for the items, based on facility indices, is categorized into three levels: low ($b < -1$), moderate ($-1 \leq b \leq 1$), and high ($b > 1$). By classifying items according to their difficulty, the study has shown how an adaptive testing logic can be created. The adaptive test system intelligently navigates examinees across different levels of item difficulty, tailoring the assessment to match their ability levels. This adaptive approach ensures that examinees are presented with items that are neither too easy nor too difficult, enhancing the accuracy and efficiency of the assessment process. Furthermore, the findings illustrate how facility indices can form the basis of a Computer Adaptive Test (CAT), a sophisticated form of AI-driven educational assessment.

Recommendations

Based on the findings the following recommendations were made:

i. It is recommended that the model developed in this study be expanded to include additional parameters such as discrimination and guessing, to refine the adaptive logic and further improve the precision of the assessment.

ii. To ensure the effective implementation of AI-based assessments, professional development programs should be established to train educators on how to interpret b-parameters and how to integrate adaptive testing technologies into their assessment practices.

iii. Before large-scale deployment, it is advised that the adaptive test be piloted across different educational institutions to assess its effectiveness and ensure it works optimally across diverse student populations.

iv. For the adaptive testing system to be adopted widely, adequate technological infrastructure, including reliable internet access and AI-enabled testing software, should be developed, especially in under-resourced areas.

References

Ade-Ojo, G. (2019). Artificial intelligence in education: Implications for teaching and learning in Nigeria. *International Journal of Educational Technology*, *13*(1), 45-56.

Ajayi, I. A., & Fashola, O. S. (2020). Challenges of integrating artificial intelligence in educational assessment in Nigeria. *Nigerian Journal of Educational Technology*, *6*(2), 33-47.

Azeez, N. A., Alhassan, J. K., & Oyeyemi, G. M. (2022). AI in Nigeria: Prospects and challenges for education and skills development. *Journal of African Education*, *8*(1), 54-70.

Babalola, J. B. (2020). Technology and rural education in Nigeria: Overcoming barriers to AI adoption. *Educational Journal of Nigeria*, *9*(2), 29-40.

Baker, T., & Smith, L. (2019). Educ-AI-tion reconsidered: AI's impact on education assessment. Oxford University Press.

Chen, L. (2019). *Adaptive learning and AI-driven assessment.* Journal of Educational Technology, *15*(4), 75-89.

Heffernan, N., & Heffernan, C. (2014). The impact of AI in formative assessments. *Journal of Computer Assisted Learning, 30*(4), 298-312.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education.

OECD. (2019). AI and the future of skills and knowledge. https://www.oecd.org/

Ogunode, N. J. (2021). Barriers to the implementation of artificial intelligence in Nigerian schools: The need for governmental intervention. *International Journal of Education and Development*, *12*(4), 67-80.

Okoye, I., Akinwale, O., & Yusuf, A. (2021). Teacher preparedness for AI in education: A study of Nigerian educational institutions. *African Journal of Educational Research*, *19*(3), 87-101.

Schiff, D. (2020). *Artificial intelligence and the future of teaching and learning*. Brookings Institution.

Shute, V. J., & Ventura, M. (2013). *Measuring and supporting learning in games: Stealth assessment.* MIT Press.

UNESCO. (2019). *Guidelines for AI and education: Towards a fair and inclusive system*. United Nations Educational, Scientific and Cultural Organization.

UNESCO. (2020). *AI in education: Challenges and opportunities*. United Nations Educational, Scientific and Cultural Organization.

UNESCO. (2021). *Ethics of artificial intelligence in education*. United Nations Educational, Scientific and Cultural Organization.

UNICEF. (2021). *AI-powered solutions for remote education*. United Nations Children's Fund.

Uwadia, C., & Adebayo, M. (2018). Developing AI capacity in Nigerian education: The role of public and private sectors. *Nigerian Journal of Information Technology*, *14*(1), 55-64.

Williamson, B. (2018). Policy networks, performance metrics and platform markets: Charting the expanding data infrastructure of higher education. British *Journal of Educational Technology*, *49*(6), 1001-1018.

World Economic Forum. (2018). *AI and the future of work: Implications for education and skills.* https://www.weforum.org/

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education, 16*(1), 39.