Research Article / Review Article FUPRE JOURNAL 9(1):321-338(2025)

FUPRE Journal

of

Scientific and Industrial Research

ISSN: 2579-1184(Print) ISSN: 2578-1129 (Online)

http://fupre.edu.ng/journal

A Computer Adaptive Test Framework Leveraging Genetic Algorithm for Optimizing Testlet-Based Question Selection and Randomization

EYEHORUA, O. C. [1,*], ***OYEMADE, D. A.*** [2], ***AKO, R. E.*** [3]

[1,2,3]*Department of Computer Science, Federal University of Petroleum Resources, Effurun, Nigeria*

ARTICLE INFO

Received: 20/12/2024

Accepted: 07/04/2025

Keywords

Computer Adaptive Test, Genetic Algorithm, Genetic Algorithm Randomization, Heuristic Randomization, Randomization

ABSTRACT

Computer Adaptive Testing (CAT) has ushered in a new era in assessment by providing a more efficient, accurate, and tailored testing experience for measuring each test-takers' abilities. However, optimizing question selection and randomization remains a significant challenge. In the framework of CAT lies the question selection and randomization mechanism, which determines the next item to be administered based on the test-taker's responses. Traditional approaches to question selection often rely on Heuristic Randomization (HR), which can lead to suboptimal item selection and compromised test validity. To address these limitations, this study proposes a novel CAT framework that leverages Genetic Algorithms (GA) to optimize testlet-based question selection and randomization. The methodology adopted supports Object-Oriented Analysis and Design (OOAD) using Agile Methodology. The proposed Genetic Algorithm Randomization (GAR) employs a multi-parameter fitness function, incorporating question difficulty, discrimination, time, and learning objective coverage. Through iterative evolution, the algorithm identifies the optimal question set combination that maximizes test efficiency. By harnessing evolutionary principles, GAR optimizes the selection of CAT test questions. The results demonstrate that the possibility of optimizing CAT with GAR for a better question selection and randomization process. The findings of this research have significant implications for the development of more sophisticated CAT systems, ultimately leading to improved assessment outcomes and better-informed decision-making in education and professional certification.

1. INTRODUCTION

Computer-Adaptive Tests (CAT) have emerged as a pioneering technology in education and assessment. This assessment method has gained widespread acceptance among educational institutions and examination boards for various evaluations, including university and college entrance exams (Alsufayan and El-dakhs, 2023). The quality of questions in CAT has a profound impact on the validity, reliability, and accuracy of assessment results (Auwalu and Hayyo, 2023). Over the years, experienced item writers manually selected questions and implemented manual randomization before categorization. Manually, randomizing questions into categories like Category A, B, C, and D was acknowledged as a randomization method. Over the years, this approach introduced question diversity among test-takers and bolstered the security of CAT testlet questions (Rajiha and Rini, 2023; Senja et al., 2024).. The mechanism for

*Corresponding author, e-mail:ako.rita@fupre.edu.ng

DIO

selecting questions is a crucial component of CAT that requires enhancement, and optimizing it can significantly enhance the current CAT framework (Kara et al., 2024; Nurhikmah et al., 2021). Issues related to scalability, the selection of questions, and the usage of resources have contributed to less-than-optimal performance and possible system bottlenecks. While Heuristic Randomization (HR) effectively replaced the random selection process, it can introduce biases and provide less-than-ideal results with constrained accuracy. HR often compromises optimal performance for the sake of simplicity. As the use of CATs increases across different fields, there is a growing demand for a more effective question selection method (Auwalu and Hayyo, 2023; Supriyati et al., 2021). Routine assessments of the CAT process have become essential to pinpoint strengths, weaknesses, and potential areas for enhancing the CAT randomization process (Yilmaz, 2021; Abiola and Okegbemiro, 2021). GA, which is an optimization method modeled after the principles of natural selection and genetics, can be combined with HR to enhance the CAT question selection process (Shin et al., 2022). By integrating these approaches, the system can smartly choose questions and optimize resource distribution to enhance scalability, performance, and user satisfaction, taking advantage of the optimization capabilities of GA alongside the randomization features of HR (Sinha, 2024).

2. RELATED WORKS

CAT is gradually becoming a general assessment method used by companies and educational instructions in developed countried (Lu, 2023). The ability to handle testlet questions efficiently accounts for its acceptability (Istiyono et al., 2023). While the initial setup cost of CAT is substantial, it is shown to be economically beneficial. Furthermore, it reduces the necessity for physical infrastructure, such as examination venues and storage facilities (Murniati et al., 2023; Berg et al, 2024; Burgmanis et al., 2023. Benefits included in a CAT testlet system includes a reduction in test taker's anxiety levels, along with insights into how CAT influences exam-related stress among test takers (Ugwu and Ogbozor, 2021; Niragudi, 2021; Shuaibu Muhammad and Hayyo, 2023).

The approach of question selection in CAT has demonstrated a considerable perceived effect on test takers' performance based on individual differences (Aletan et al., 2022; Lync, 2022). An evaluation of test takers' performance indicates that CAT is more efficient and effective for knowledge-based assessments, as it is capable of identifying suspicious behavior among test takers during exams via surveillance in examination halls (Genemo, 2022; Rice, 2022). CAT is recognized as a convenient method for evaluating a test taker's knowledge and comprehension, offering advantages such as security, consistency, convenience, efficiency, innovation, and advanced management, along with lower costs and logistical challenges (Chatterjee et al., 2022). A descriptive and comparative study indicates that the existing plugins utilized for importing questions through parameterized queries, such as the FastTest Plugin, coupled with a large test question bank, are effective for selecting questions in CAT (Huerta et al. 2022).

Recognizing the hurdles encountered by test takers with limited or no ICT training is a valuable approach to evaluate preparedness for CAT among test takers (Ukwueze and Uzoagba, 2021). Providing training for test takers lacking proficiency in ICT is essential before engaging in CAT (Sheu and Evanero, 2022; El-Gamams et al., 2022). Introducing test takers to randomized questions and ICT prior to participating in CAT can enhance examination results (Dung et al., 2022; Usman and Olaleye, 2022; Ebimgbo et al., 2021). The randomization feature of CAT serves as an effective method for test takers to gauge their understanding and learning

progress, and it holds the potential to drive changes in the assessment procedures (Chun et al., 2023); (Campero et al., 2022).

Through both qualitative and quantitative evaluations of test takers' performances in CAT, it has been validated that CAT is a more effective examination system (Yu and Iwashita, 2021; Jalo et al., 2021). Neural Computerized Adaptive Testing (NCAT) represents a fully adaptable framework that employs a question selection algorithm that adjusts to the candidate's performance, implements a feedback mechanism, detects intricate relationships between test takers and questions (such as guessing and slipping factors), and accurately assesses test takers' proficiency, thereby decreasing test duration (Zhuang et al., 2023; Schurig et al., 2021).

CAT has also been utilized in creating instruments for assessing learning outcomes, which enhance efforts to integrate questions into the question banks (Agusta, 2022; Horita et al., 2021; Lin et al., 2022). Applying the Simple Additive Weighting Method (SAW) helps overcome test takers' doubts about the legitimacy of results produced by the CAT system and aids in analyzing and enhancing the result-checking process of the CAT (Negoro et al., 2021). Expanding coverage across various areas of the student curriculum can enhance question variety and improve test takers' reading skills when preparing for a CAT examination (Adigun et al., 2021).

Incorporating CAT for theoretical examinations necessitates a backup database structure, a scalable network system capable of handling numerous users, upgraded CAT software, a user-friendly interface, and an optimized approach (Addah et al., 2021). The implementation of CAT demands an innovative network meta-analysis method that enables comparison (i.e., ranking) of various feedback types regarding their impact on both lower- and higher-order performance, in addition to assessing feedback mechanisms and establishing feedback as a crucial element for effective learning (Mertens et al., 2022). An adaptive testing algorithm for optimization, coupled with a strategic method for item selection in a Computer Adaptive Test (CAT), indicates that the Efficiency Balanced Information (EBI) Method can serve as a viable technique for achieving more efficient scores, enhancing adaptive functionality testing, and maintaining consistency across CAT components and EBI backtracking (Joko and Pramono, 2023). The incorporation of biometric features in CAT offers a sufficient security measure, and with the use of Advanced Encryption Standard (AES), the security becomes more robust, as the system will incorporate biometric fingerprint verification, image capture, and processes for data encryption and decryption (Huh, 2022; Ndun, 2021). By utilizing TF-IDF, Multiple-Choice Questions (MCQs) can be widely applied across different assessment methods to introduce features such as feedback mechanisms. Feedback conditions during exams have been shown to enhance affective-motivational outcomes, attributed to Knowledge Result (KR) feedback incorporated into various delivery formats (Kuklick and Lindner, 2021). The Fisher-Yates Shuffle Algorithm (FYSA) effectively randomizes exam questions in CAT with quick execution speed while avoiding repetition and duplication. Their system employs the ADDIE model and is developed using a web framework (with PHP as the backend, JavaScript for client-side processing, and MySQL for the database). This approach was successful in establishing a database system for CAT and integrating functionalities related to accessibility, useful tools, and CAT content. The design and structure of the system were thoroughly detailed, and the user interface specifications were clearly outlined (Febriani et al., 2021; Abah et al., 2022).

The conventional method for selecting questions in CAT has been a fundamental aspect of assessment strategies within

educational contexts. Although these traditional approaches have fulfilled their roles, they fall short in terms of the efficiency and flexibility necessary for contemporary educational settings (Oladele, 2021; Kamaludin, 2023). Through different techniques employed by various researchers, there is significant potential in enhancing multiple elements of the CAT exam process, such as test generation, resource management, question selection, and decision-making (Paci and Kote, 2023). A variety of frameworks have been utilized to improve the question selection process in CAT. The Evidence-Centered Design (ECD) framework, created for the development of assessments, tests, and evaluations, focuses on collecting evidence that supports conclusions regarding an individual's knowledge, skills, and abilities (Clarke-Midura et al., 2021). The core principles of ECD revolve around designing with evidence in mind, gathering pertinent evidence related to assessment questions, representing student knowledge, skills, and abilities through student modeling, outlining tasks that generate specific evidence via task modeling, accumulating and integrating evidence from various tasks, and constructing validity arguments backed by empirical evidence (Farrell et al., 2024). ECD employs software tools such as the ECD Toolkit, assessment design studio, and evidence center (Fredriksson et al., 2021). Another prominent assessment framework is the Cognitive Diagnostic Assessment (CDA) approach, which seeks to assess students' cognitive abilities and knowledge in a detailed and sophisticated manner (Wang et al., 2024). Models used in CDA include the Rule-Space Model (RSM), Attribute Hierarchy Model (AHM), Cognitive Diagnostic Model (CDM), and Bayesian Network Model (BNM). The software tools affiliated with CDA consist of the CDA Toolkit, Assessment Design Studio, and the Cognitive Diagnostic Assessment Platform (CDAP) (Wu et al., 2023). In contrast, the Item Selection Model (ISM) is an exclusive psychometric framework for item selection designed to choose questions or tasks for assessments, tests, or evaluations. ISM seeks to enhance the item selection process to meet specific assessment objectives, which may include maximizing reliability, reducing bias, improving validity, and increasing efficiency. Its essential components encompass an item bank, a collection of items with established characteristics, item parameters that describe the statistical features of each item (such as difficulty and discrimination), a test specification that outlines assessment objectives, targets, and constraints, and an optimization algorithm involving a mathematical procedure for selecting optimal items (Panphet et al., 2024).

IRT models the connection between item responses and latent traits by focusing on simulating item parameters (difficulty, discrimination, and guessing) along with person parameters (Huda et al., 2024). IRT can be utilized in testing, item banking, and test equating, while Classical Test Theory (CTT) is utilized in test creation, validation, and assessment. CTT emphasizes reliability, validity, and item analysis, assuming that observed scores are a combination of true scores and error. It provides estimates for test reliability, validity, and item statistics (such as difficulty and discrimination) (Alghamdi et al., 2024). In contrast to IRT and CTT, Cognitive Load Theory (CLT) and Working Memory Theory (WMT) are relevant in instructional design and assessment. The key distinction is that CLT is primarily used in learning analytics, whereas WMT focuses on cognitive training. CLT analyzes the mental effort needed to process information, differentiating between intrinsic, extraneous, and germane load. It plays a role in instructional design and assessment (Pengelley et al., 2024). WMT investigates the function of working memory in information processing, identifying limitations and capacity constraints, and also informs instructional design and testing (Ngiam, 2024). Although CTT and

Information Processing Theory (IPT) share similar applications, they differ in their descriptions. IPT outlines how individuals process and store information, with an emphasis on attention, encoding, and retrieval, and provides insights for instructional design and assessment (Ali et al., 2023). Cognitive Diagnostic Models (CDM) have notable similarities with CLT, particularly in learning analytics and instructional design; however, while IPT is employed in instrumental testing, CDM is geared towards diagnostic assessments. CDM identifies the cognitive skills and knowledge necessary to complete tasks and offers in-depth information about test takers' strengths and weaknesses while assuming cognitive processes that underlie task performance (Zhai et al., 2023).

To fulfill some specific criteria, IRT has been expanded into the Testlet Response Theory (TRT) framework. TRT is a psychometric framework that models responses to testlets, which are groups of items that share common stimuli, contexts, or requirements. TRT consists of components such as testlet definitions and characteristics, testlet and item parameters, and response models (Aşiret and Ömür, 2024). TRT models include the Linear Logistic Testlet Response Model (LLTRM), Generalized Linear Mixed Model (GLMM), and Bayesian Hierarchical Model (BHM). The software tools used for TRT are Testlet Response Theory Software (TRTS), IRTPRO, and Mplus (Kocaoglu and Sahin, 2024). In the mechanism for question selection, Content Balancing guarantees the representation of various content areas, while Exposure Control (EC) restricts the access to sensitive or critical questions, serving as a security measure to manage questions (Lu et al., 2023).

3. QUESTION SELECTION MECHANISM

A successful question selection mechanism is essential for refining CAT systems. Factors such as alignment with curriculum standards, suitability of difficulty levels, variety of question formats, statistical validity and reliability, and control of item exposure are vital for guaranteeing that chosen questions effectively evaluate student understanding and improve the overall assessment experience.

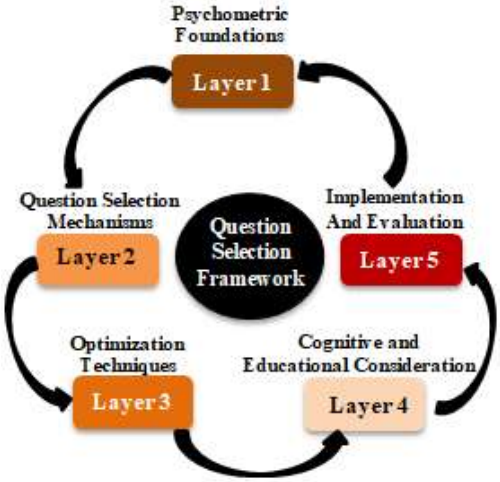

Figure 1: Question selection framework

Dynamic question selection strategies offer a method that adapts to question selection in real-time, based on various parameters, which enhances both the assessment process and accuracy in measuring test takers' knowledge. Critical components of dynamic question selection include adjusting difficulty adaptively, monitoring performance in real-time, creating personalized learning pathways, and utilizing data-driven insights.

Assessing the effectiveness of question selection methods in CAT is important for confirming that the evaluations accurately reflect the test takers' knowledge and abilities. When looking to optimize these methods through GA, various performance metrics can be applied, such as reliability, validity, discrimination index, item difficulty index, and response time analysis.

GA have become increasingly recognized as a strong optimization tool across numerous domains, including education. Their ability to

address complex issues and adjust to evolving environments makes them particularly valuable in educational contexts. Creating a GA requires establishing certain parameters that facilitate the optimization process to reach an optimal outcome.

The process of selecting questions using GA starts with the random generation of a population made up of various question subsets. Each subset can differ in its size and arrangement. These subsets are assessed using a fitness function that evaluates how well the questions correspond to learning objectives, maintains an even distribution of question types (such as theoretical and practical), and ensures that the overall difficulty level of the chosen questions aligns with the abilities of the intended audience. Those subsets that perform well are chosen for reproduction, and methods like Crossover and Mutation allow question subsets to merge and evolve, leading to new configurations that boost diversity and adaptability. The iteration continues until the most effective question subsets are determined. By pursuing optimization strategies, CAT can establish a more efficient, inclusive, and trustworthy system for administering fair and credible assessments to its test takers.

4. MATERIALS AND METHODS

The approach for the suggested CAT question selection system utilizes a Software Centric Methodology, emphasizing the essential principles for software quality through an Agile Software Development Model and Object-Oriented Analysis and Design (OOAD) Methodology. The programming languages employed primarily include Hypertext Markup Language (HTML 5), JavaScript (JS) and related libraries, and PHP (Hypertext Preprocessor) within an Item Selection Model (ISM) framework, taking advantage of flexible web development features to incorporate web-based capabilities.

The rationale for adopting this methodology encompasses (but is not limited to) the following points:

- While HTML is a markup language that does not inherently enforce OOAD principles, it can be effectively combined with JS and PHP technologies to facilitate OOAD. HTML is chosen for its platform independence, ensuring consistent display across various browsers and operating systems. Its benefits include quick rendering and adaptable layout, while the combination with CSS, PHP, and JS leads to visually engaging effects.
- JavaScript and its associated libraries represent a collection of multi-paradigm languages that support different programming styles such as Object-Oriented Programming (OOP), Functional Programming (FP), and Event-Driven Programming (EDP), all of which are vital for the question selection process. Although JavaScript does not strictly adhere to traditional OOAD principles, it accommodates many OOP concepts, including encapsulation, abstraction, inheritance, and polymorphism.
- PHP serves as a server-side scripting language that supports several programming paradigms, incorporating Object-Oriented Programming (OOP), Functional Programming (FP), and Procedural Programming (PP). The use of PHP 5+ in the programming process allows for full-fledged OOP features, including encapsulation, abstraction, inheritance, and polymorphism.

4.1 Analysis of the Existing System

The existing CAT System is proprietary software designed by Multilent. It is a distributed assessment management platform, crafted to be adaptable, highly secure, and capable of accommodating a significant number of simultaneous test takers. Its functionalities encompass content creation, data security, exam content delivery network, cloud-based distribution, reporting, and analytics. Figure 2 display the existing CAT model.

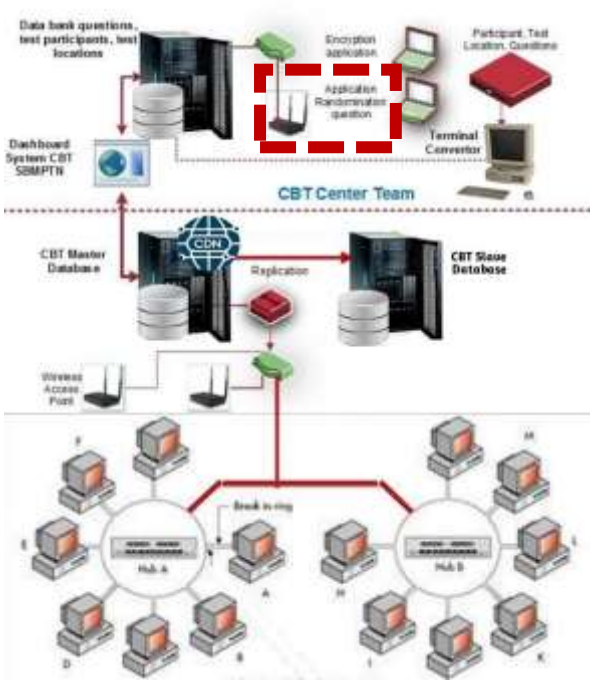

Figure 2: Existing CAT System Model using the Heuristic Randomizer (A Silent refactoring Design as described: *Rapid Test Multilent JAMB 2024 CAT Manual, 2024)*

The existing system architecture is illustrated below in Figure 3.

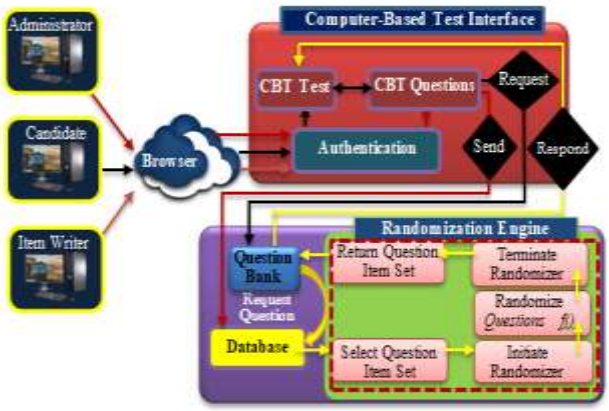

Figure 3: Existing CAT System Architecture using the Heuristic Randomizer (A Silent refactoring Design as described: *Rapid Test Multilent JAMB 2024 CAT Manual, 2024)*

The framework of the existing system consists of the CAT interface, question repository, and randomization engine. By tackling challenges and capitalizing on opportunities, the proposed system seeks to refine the question selection HR process in CAT examinations utilizing a Genetic Algorithm Randomizer.

4.2 Analysis of the Proposed CAT System

Enhancing the CAT examination system centers on refining the process of question selection within the CAT framework.

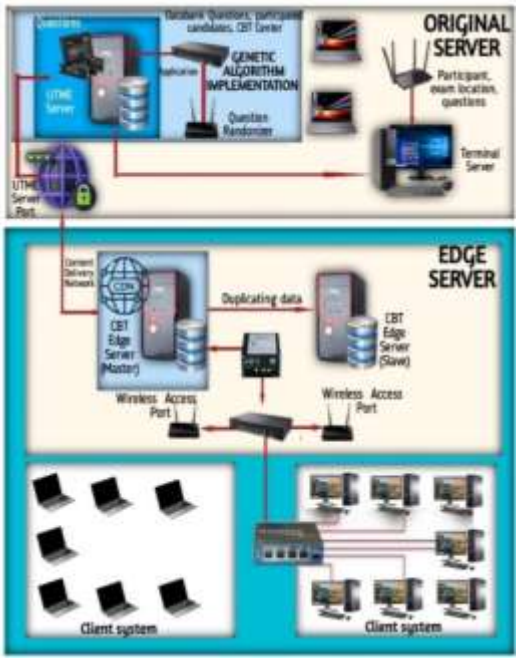

Figure 4: Model of the Proposed Optimization CAT System

The proposed model consists of multiple components working together to provide a seamless pre-examination experience. The architecture contains CAT Interface module, Question Bank module, and Randomization Engine module.

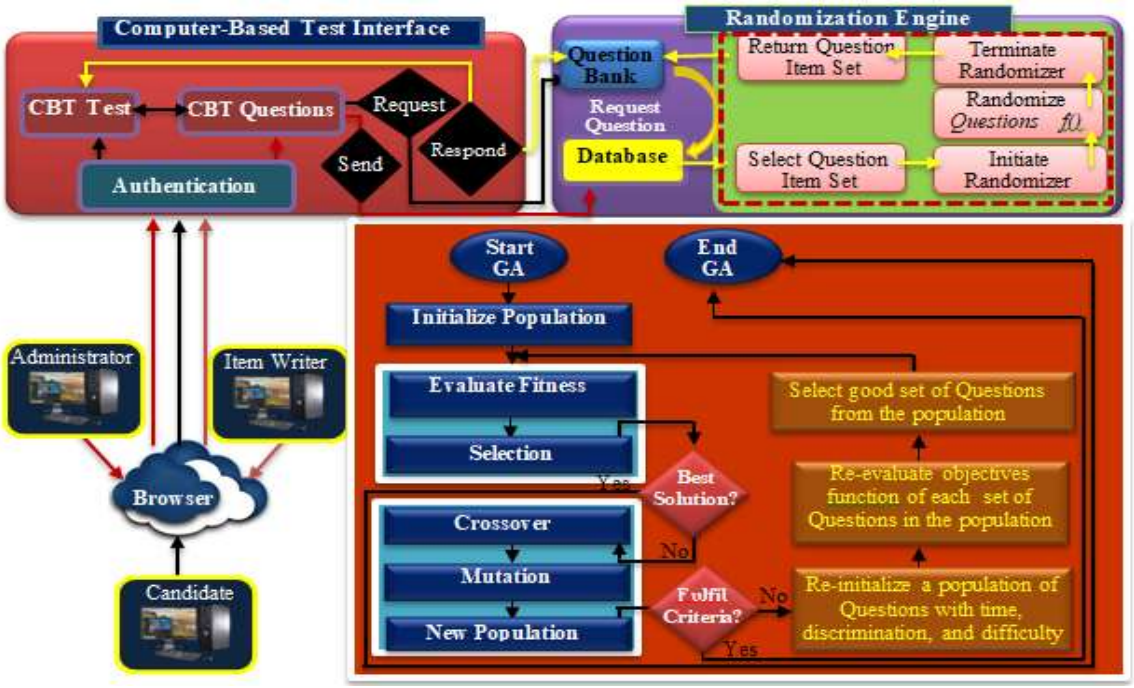

Figure 5: Architecture of the Proposed CAT System

The question selection process using GA starts with the selection of a random generation from a population made up of various question subsets. Each subset can differ in its size and arrangement. These subsets are assessed using a fitness function that evaluates how well the questions correspond to learning objectives, maintains an even distribution of question types, and ensures that the overall difficulty level of the chosen questions aligns with the abilities of the intended audience. The subsets that perform well are chosen for reproduction, and genetic operators (crossover and mutation) allow question subsets to merge and evolve, leading to new configurations that boost diversity and adaptability.

4.3 Genetic Algorithm Optimization

Question selection entails the process of picking questions from a broader set to construct customized assessments. This procedure can be affected by several factors, including the intended audience, educational aims, and the level of difficulty desired. GA can enhance question selection by pinpointing the most appropriate questions that satisfy established criteria. The Sequence flow utilizes an Object-Oriented Analysis and Design (OOAD) approach alongside an n-tier client-server Architecture within a distributed system. The interface of the CAT application is designed to be flexible, interactive, secure, and efficient for all test takers or users of the CAT system. It incorporates features that allow for real-time responses, and each interface will be developed using JavaScript, harnessing the principles of Object-Oriented programming and aiding in the application of GA. The n-tier client-server Architecture is divided into multiple layers, where the presentation layer functions as the CAT Interface, the Application layer contains the CAT logic, and the Database layer is present. Although these layers collaborate to form a single application, they also function independently of each other. This separation is intended to enhance efficiency, security, scalability, maintenance, and upgrades. An Application Programming Interface (API) is integrated to facilitate smooth data flow.

This framework aims to establish a CAT System that incorporates a GAR in its question selection process. CAT is an increasingly popular platform that evolves in

conjunction with technological advancements.

4.3.1 GAR Test Generation

Creating an assessments that accurately measure a test taker's knowledge and skills requires an optimized system.. HR test creation can be time-consuming and may lead to inconsistencies in difficulty levels and content coverage. GA offers a systematic approach to automate and optimize this process.

GA for test generation in CAT
Test Generation:

```
`// Define classes
class Question { constructor(id, d, r) { this.id = id; this.difficulty = d; this.relevance = r; } }
class Test { constructor(q) { this.questions = q; } fitness() {} }
// Set constants
const PS = 100, TL = 10;
// Initialize variables
let QB = [], P = generateInitialPopulation();
// Define functions
function generateInitialPopulation() {}
function selection(p) {}
function crossover(p1, p2) {}
function mutate(t) {}
// Main loop
for (let i = 0; i < iterations; i++) {
  P = selection(P);
  const child = crossover(P[0], P[1]);
  P[0].questions.set(index, mutate(child));
}
// Select best tests
const bestTests = selection(P);
```

` ` `

GA can be applied to test generation by encoding potential test configurations as chromosomes. Each chromosome represents a unique test, with genes corresponding to specific questions or topics. The process typically involves the following steps:

- Initialization: A population of test configurations is generated randomly, with each configuration representing a different combination of questions.
- Fitness Evaluation: Each test configuration is evaluated based on a fitness function that measures its quality.

4.3.2 GAR in Question Selection

Question selection involves choosing specific questions from a larger pool to create tailored assessments. This process can be influenced by various factors, such as the target audience, learning objectives, and the desired difficulty level. Genetic algorithms can optimize question selection by identifying the most suitable questions that meet predefined criteria. Question selection in the Genetic Algorithm (GA) for test generation in Computer-Based Testing (CAT) in JavaScript:

Question Selection:

```
// Question Selection
let questionBank = []; // Initialize question bank
questionBank.push(new Question(id, difficulty, relevance)); // Add questions

const testLength = 10; // Define test length
let selectedQuestions = [];

for (let i = 0; i < testLength; i++) {
  const randomIndex = Math.floor(Math.random() * questionBank.length);
  const randomQuestion = questionBank[randomIndex];
selectedQuestions.push(randomQuestion);
}
// Fitness Calculation
let fitness = 0;
selectedQuestions.forEach(question => {fitness += question.difficulty * question.relevance;
});

// Selection
population.sort((a, b) => b.fitness - a.fitness);
population = population.slice(0, populationSize / 2);
```

“”

Mutation Function:

```
function
mutate(selectedQuestions) {
  // Select a random question
index
  const index =
Math.floor(Math.random() *
selectedQuestions.length);

  // Select a random question
from the question bank
  const randomQuestion =
questionBank[Math.floor(Math.ra
ndom() * questionBank.length)];

  // Replace the question at the
selected index with the random
question
  selectedQuestions[index] =
randomQuestion;

  return selectedQuestions;
}
```

Crossover Function

```
function crossover(parent1,
parent2) {// Select a random
crossover point
  const crossoverPoint =
Math.floor(Math.random() *
parent1.length);

  // Create a new test with
questions from parent1 up to the
crossover point
  const child = parent1.slice(0,
crossoverPoint);

  // Add questions from parent2
after the crossover point

child.push(...parent2.slice(crossov
erPoint));

  return child;
}
```

In GA-based question selection, each chromosome represents a subset of questions from the available pool. The process follows a similar structure to test generation:

- Initialization: A population of question subsets is randomly generated. Each subset may vary in size and composition.
- Fitness Evaluation: Each subset is evaluated based on a fitness function.
- Mutation Function: This selects a random question in the `selectedQuestions` array and replaces it with a random question from the `questionBank`.
- Crossover Function: This creates a new test by combining questions from two parent tests. It selects a random crossover point and takes questions from the first parent up to that point, then adds questions from the second parent after that point.

5. RESULTS

The optimized CAT system features modules that serve as logical containers for organizing objects related to specific tasks, along with optional programming logic. Specifically, these modules encapsulate the models associated with each task and the code used to execute these tasks.

As illustrated in Figure 6, the control panel provides a variety of features that empower test item writers to effectively administer the question bank. Through this platform, test item writers can generate and submit new questions to the question bank, which are subsequently preserved in the database. It is designed to facilitate easy editing of existing questions, allowing for seamless updates or alterations as required. Questions that have already been saved can be accessed from the database, enabling Item Writers to search, filter, and retrieve particular questions, while also offering tools for organization, including categorization and assigning question difficulty level, discrimination, and time.

The GAR Configuration Panel is designed for administrators to manage and configure the Genetic Algorithm Randomizer (GAR)

parameters to ensure optimal question selection. This panel offers a platform for administrators to activate the GAR, guaranteeing that questions are chosen in a manner that optimizes the validity and reliability of the evaluation. Administrators can adjust the GAR by modifying parameter setups such as population size, mutation rate, crossover rate, etc. This capability allows the customization of the GAR to meet the specific requirements of their evaluation. Figure 7 illustrates the GAR Configuration Panel for the CAT system.

This information aids administrators in refining the GAR and optimizing the overall quality of the evaluation.

Figure 6: Question Control Panel

Through the GAR Configuration Panel, randomized questions in the question bank can be managed. This includes actions like altering, inserting, and removing questions, as well as organizing them by topic. The GAR Configuration Panel allows administrators to set assessment parameters, such as the total number of questions, assessment duration, instructions, and scoring guidelines. This ensures that the GAR selects questions that align with the objective of the assessment. The panel offers analytical insights into question selection, including data on question frequency, level of difficulty, and discrimination index.

Figure 7: GAR Configuration Panel

By providing a comprehensive platform for configuring the GAR, managing the question bank, setting assessment criteria, and analyzing data, the GAR Configuration Panel contributes to improved assessment validity, increased efficiency, optimized security, and better decision-making. This ultimately results in a more effective and dependable evaluation process. The CAT module in Figure 8 displays the required assessment of the test-takers. During registration, test-takers are expected to

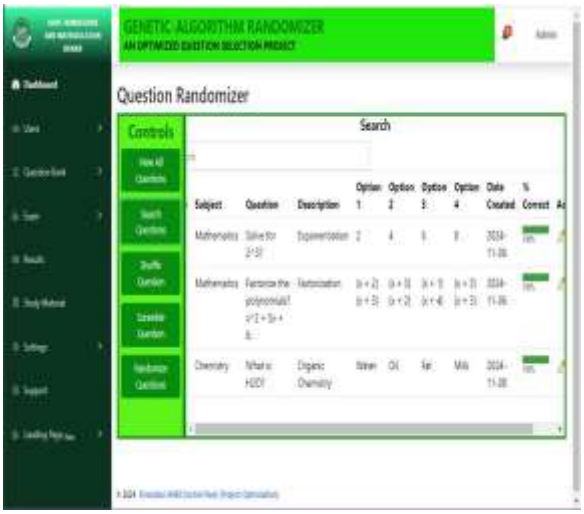

Figure 8: List of CAT Assessment Subjects

The CAT Assessment Module serves as the platform through which candidates interact with assessment questions. The CAT Assessment questions are dynamically created by the GAR, guaranteeing that every candidate receives a distinct set of questions

customized to their assessment needs.

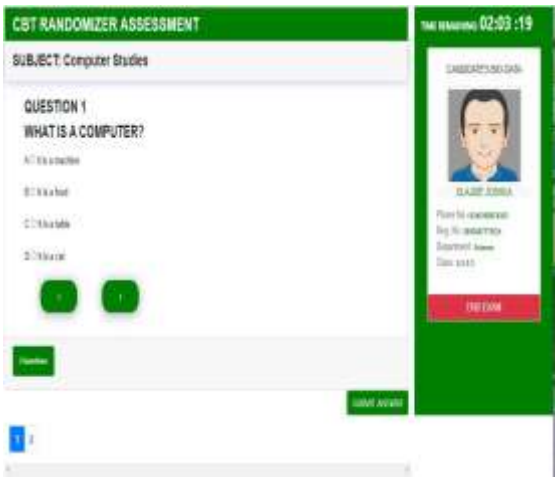

Figure 9: GAR CAT Assessment Platform

register 4 subjects each. The list of subjects are displayed before the commencement of the CAT assessment to ensure accuracy. On starting the assessment, the test-takers confirms that the assigned subjects were selected for assessment.

The CAT Assessment interface presents candidates with a user-friendly layout that offers clear instructions and guidelines for completing the assessment. This page generally contains a question display section where the GAR-selected questions are shown to the candidate. Figure 9 displays the GAR CAT platform.

This section includes functionalities such as scrolling and text formatting to ensure that questions are presented clearly and succinctly. The timer or countdown that displays the remaining time the candidate has to finish the assessment, and the progress tracker that indicates the candidate's advancement through the assessment, detailing the number of questions answered and the total number of questions in the assessment. The CAT assessment page is crafted to be very interactive, enabling candidates to engage with the assessment questions dynamically. The page features multiple-choice questions, offering options for selecting answers. By integrating these features and functionalities, the CAT assessment page delivers a comprehensive and inclusive platform for candidates to demonstrate their knowledge and skills.

6. CONCLUSION

CAT has revolutionized the education and assessment sphere, providing a globally accepted approach for evaluating candidates. The combination of GA with HR has yielded promising advancements in question selection optimization. GA's capability to assess question sets based on factors like difficulty, discrimination, and estimated timing allows the system to intelligently choose questions and improve resource allocation. The introduction of the GAR presents a notable enhancement to the existing systems, boosting scalability, performance, and user satisfaction. By utilizing the optimization strengths of GA along with the randomization aspects of HR, GAR guarantees that question selection is no longer a hindrance within CAT systems. Essentially, this research illustrates the effectiveness of optimization methods in enhancing the efficiency and quality of CAT systems. The outcomes of this study can be utilized across various assessment contexts, ultimately resulting in more trustworthy, valid, and precise evaluation outcomes. As CAT maintains its critical role in education and assessment, incorporating advanced optimization techniques like GA will be vital in providing high-quality assessments that accurately reflect a candidate's understanding.

7. RECOMMENDATIONS

Recommendations Based on research and findings, it is highly advised that educational institutions and testing organizations integrate GAR to optimize the CAT question selection mechanism. This approach offers benefits that improve the validity, reliability,

efficiency, and effectiveness of test development and administration. To ensure a seamless adoption, researchers are encouraged to investigate the application of GA for optimizing question selection mechanisms in diverse assessment types. Also, researchers are encouraged to advocate for the adoption of GAR in educational assessment and testing through policy initiatives and stakeholder engagement. Comprehensive testing and validation of the GAR system should also be conducted and regular assessment and evaluation of system's performance should be made. Continuous updates and enhancement of GAR algorithm to maintain peak performance. By adopting GAR, educational institutions and assessment organizations can significantly improve the quality and efficacy of their CAT systems, ultimately resulting in improved assessment outcomes and a better learning experience for candidates.

References

Abah, J. A., HONMANE, O., Age, T. J., and OGBULE, S. O. (2022). Design of Single-User-Mode Computer-Based Examination System for Senior Secondary. *VillageMath Educational Review (VER)*, *3*(1), 107–134. https://doi.org/10.2139/ssrn.4061818

Abiola, L., and Okegbemiro, J. O. (2021). Candidates ' Perceptions Towards Computer - Based Testing : A Case Study Of Distance Learning Programme University Of Maiduguri , Federal College Of Education (Technical) Gombe Campus. *Gombe Technical Education Journal*, *13*(2), 45–54.

Addah, E., Ovuworie, G., and Ariavie, G. (2021). Implementing computer-based test for theory exams in tertiary institutions in Nigeria. *Tropical Journal of Science and Technology*, *2(1)*, *2*(1), 14–28. https://doi.org/10.47524/tjst.21.4

Adigun, K. A., Adarabioyo, M. I., Adejuwon, S. O., Adeyemo, O. A., And Babalola, B. T. (2021). SCHOOL Examination Methods And Candidates ' Performances In Nigeria (Computerand Paper Based Examination In Focus). *Turkish Journal of Computer and Mathematics Education*, *12*(14), 2449–2462.

Agusta, A. R. (2022). Development of Learning Outcomes Assessment Instruments Using Computer Based Test (CBT). *Jurnal Pendidikan*, *6*(1), 255–267.

Aletan, M. S., Falade, E. O., Erondu, N. C., and Ogunsola, R. 2(2022). Undergraduates' Views and Attitudes Towards the Use of Computer-Based Tests for Examination at the University of Lagos, Nigeria. *International Journal of Advanced Research in Social Sciences, Environmental Studies and Technology*, *7*(1), 62–74. https://doi.org/10.48028/iiprds/ijarssest.v7.i1.06

Alghamdi, A., Alghuthayr, K., Alqahtani, S., Alshahrani, Z., Asiri, A., Ghazzawi, H., Helmy, M., Trabelsi, K., Husni, M., and Jahrami, H. (2022). The translation and validation of the surgical anxiety questionnaire into the modern standard Arabic language: results from classical test theory and item response theory analyses. *BCM Psychiatry*, *24:694*, 1–13. https://doi.org/10.1186/s12888-024-06142-y

Ali, A., Wang, H. and Boekhorst, J.A. (2023) A moderated mediation examination of shared leadership and team creativity: a social information processing perspective. Asia Pac J Manag 40, 295–327. https://doi.org/10.1007/s10490-021-09786-6

Alsufayan, R., and El-dakhs, D. A. S. (2023). Achievement Emotions in Paper-Based Exams vs . Computer-Based Exams : The Case of a Private Saudi University. *International Journal of Online Pedagogy and Course Design Volume 13 • Issue 1*, *13*(1), 1–21.

https://doi.org/10.4018/IJOPCD.322084

Aşiret, S., and Ömür, S. (2024). Investigating The Performance of Item Selection Algorithms in Cognitive Diagnosis Computerized Adaptive Testing. Journal of Measurement and Evaluation in Education and Psychology, 15(2), 148-165. https://doi.org/10.21031/epod.1456094

Auwalu, S. M., and Hayyo, D. (2023). Candidates' Perception on the Role of Computer Based Test in Curbing Examination Malpractice. *International Journal of Academic Research in Progressive Education and Development*, *12*(1), 1643–1651. https://doi.org/10.6007/ijarped/v12-i1/16317

Berg, C., Devathasan, K., and Craig, M. (2024). Test Anxiety , Self-Efficacy and Prior Experience. *In Proceedings Ofthe 2024 Innovation and Technology in Computer Science Education V.*, *1*, 2999557. https://doi.org/10.1145/3649405.3659487

Burgmanis, Ģ., Mikite, M., France, I., and Namsone, D. (2023). Development of Computer-Based Diagnostic Assessment System : case study of equivalence of paper-and-pencil and computer-based testing. *Proceedings of the International Scientific Conference*, *1*, 613–622.

Campero, A., Vaccaro, M., Song, J., Wen, H., Almaatouq, A., and Malone, T. (2022). a test for evaluating performance in human-computer systems. *Human-Computer Interaction*, *1*.

Chatterjee, P., Bose, R., Banerjee, S., and Roy, S. (2022). Cloud based LMS Security and Exam Proctoring Solution. *Journal of Positive School Psychology*, *6*(3), 3929–3941.

Chen, F., Lu, C., Cui, Y., and Gao, Y. (2023). Learning Outcome Modeling in Computer-Based Assessments for Learning : A Sequential Deep Collaborative Filtering Approach. *IEEE Transactions on Learning Technologies*, *16*(2), 243–255. https://doi.org/10.1109/TLT.2022.3224075

Chun, K., Jin, H. K., Yoon, J., Kim, M. G., Choi, K. H., Kim, E., Kim, H., Kim, J., Kim, G., Kim, K., Lee, J., Chung, E. K., Lee, Y. S., and Rhie, S. J. (2023). Novel innovative computer-based test (Inno-CBT) item types for national licensing examinations for health care professionals. *BMC Medical Education*, *23*(560), 1–7.

Clarke-Midura, J., Silvis, D., Shumway, J. F., Lee, V. R. and Kozlowski J. S., (2021) Developing a kindergarten computational thinking assessment using evidence-centered design: the case of algorithmic thinking, Computer Science Education, 31:2, 117-140, DOI: 10.1080/08993408.2021.1877988

Devathasan, K., Craig, M., Damian, D., and Berg, C. (2024). Test Anxiety and Self-Efficacy in a Computer-Based Test Environment. *Conference: WCCCE "24": The 26th Western Canadian Conference on Computing Education*, 1–7. https://doi.org/10.1145/3649405.3659487

Dung, C. J., Adakole, B. A., Gwani, E. M., and Yagwom, M. B. (2022). Design and Development of an Online Cbt Software for Assessment and Examination in the Department of Science and Technology. *International Journal of Innovative Research and Advanced Studies*, *10*(1), 2248–4376. https://www.bwjournal.org/index.php/bsjournal/article/view/575

Ebimgbo, S. O., Igwe, N. J., and Okafor, A. E. (2021). Perceived effectiveness of computer-based test examination mode for large classes among undergraduates of Nigerian Universities: Implications for social work. *Journal of Social Work in Developing Societies*, *3*(1), 62–77.

El-Hamamsy, L., Zapata-C´aceres, M.,

Barroso, E. M., Mondada, F., Zufferey, J. D., and Bruno, B. (2022). The Competent Computational Thinking Test : Development and Validation of an Unplugged Computational Thinking Test for Upper Primary School. *Journal of Educational Computing Research*, *60*(7), 1818–1866. https://doi.org/10.1177/07356331221081753

Farrell, C. C., Penuel, W. R., Arce-Trigatti, P., Soland, J., Singleton, C., Resnick, A. F., Stamatis, K., Riedy, R., Henrick, E., Sexton, S., Wellberg, S., and Schmidt, D., (2024) Designing measures of complex collaborations with participatory, evidence-centered design. *Front. Res. Metr. Anal.* 9:1210547. doi: 10.3389/frma.2024.1210547

Fredriksson, T., Mattos, D. I., Bosch, J., and Olsson, H. H., (2021) "Assessing the Suitability of Semi-Supervised Learning Datasets using Item Response Theory," **2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)**, Palermo, Italy, 2021, pp. 326-333, doi: 10.1109/SEAA53835.2021.00049

Genemo, M. D. (2022). Suspicious activity recognition for monitoring cheating in exams. *Proceedings of the Indian National Science Academy*, *88*(1), 1–10. https://doi.org/10.1007/s43538-022-00069-2

Horita, S., Park, Y. S., Son, D., and Eto, M. (2021). Computer-based test (CBT) and OSCE scores predict residency matching and National Board assessment results in Japan. *BMC Medical Education*, *21*(1), 1–8. https://doi.org/10.1186/s12909-021-02520-2

Huda, A., Firdaus, F., Irfan, D.,Hendriyani, Y., Almasri, A., and Sukmawati, M. (2024). Optimizing Educational Assessment: The Practicality of Computer Adaptive Testing (CAT) with an Item Response Theory (IRT) Approach. *International Journal on Informatics Visual*, *8*(1), 473–480. https://dx.doi.org/10.62527/joiv.8.1.2217

Huerta, M., Caballero-Hernández, J. A., and Fernández-Ruiz, M. A. (2022). Comparative Study of Moodle Plugins to Facilitate the Adoption of Computer-Based Assessments. *Applied Science*.

Huh, S. (2022). Application of computer-based testing in the Korean Medical Licensing Examination , the emergence of the metaverse in medical education , journal metrics and statistics , and appreciation to reviewers and volunteers. *Journal of Educational Evaluation for Health Professions*, *19*(2), 1–6.

Istiyono, E., Sunu, W., Dwandaru, B., Rai, M., Ayub, S., and Saepuzaman, D. (2023). A study of the readiness of post-pandemic computer-based four-tier diagnostic test (CBFTDT): A review of economic level , school grades , and device accessibility. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *27*(1), 92–104.

Jalo, H. S., Adamu, D., and Abdullahi, U. A. (2021). Comparative Study of Computer Based Test and Paper Based Test in Assessing Candidates in General Subject. *International Journal on Perceptive and Cognitive Computing*, *7*(1), 125–131.

Kamaludin, A. (2023). Development of Computer Based Test (CBT) Website-based Chemistry National Exam for Candidates of Islamic High School. *Jurnal Penelitian Pendidikan IPA*, *9*(6), 4164–4171. https://doi.org/10.29303/jppipa.v9i6.1540

Kara, H., DOĞAN, N., and Kara, B. E. (2024). Robustness of Computer Adaptive Tests to the Presence of Item Preknowledge : A Simulation Study. *Journal of Measurement and Evaluation in Education and Psychology Research*, *15*(2), 138–147.

Kuklick, L., and Lindner, M. A. (2021). Computer-based knowledge of results feedback in different delivery modes : Effects on performance , motivation , and achievement emotions. *Contemporary Educational Psychology*, *67*, 1–17. https://doi.org/10.1016/j.cedpsych.2021.102001

Lin, J., Zhang, L., Wei, W., Cheng, P.-H., and Chang, C.-Y. (2022). Interactive computer assessment and analysis of Candidates ' ability in scientific modeling. *EURASIA Journal of Mathematics, Science and Technology Education*, *18*(12), 1–13.

Lindner, M. A., and Greiff, S. (2023). Editorial Process Data in Computer-Based Assessment Challenges and Opportunities in Opening the Black Box. *European Journal of Psychological Assessment (2023)*, *39*(4), 241–251.

Lu, W. (2023). Research on test takers ' access to a new computer-assisted test. *ICEEL '23: Proceedings of the 2023 7th International Conference on Education and E-Learning*, 189–196.

Lu, S., Liu, M., Yin, L., Yin, Z., Liu, X., Zheng, W. (2023). The multi-modal fusion in visual question answering: a review of attention mechanisms. *PeerJ Computer Science* 9:e1400 https://doi.org/10.7717/peerj-cs.1400

Lynch, S. (2022). Adapting Paper-Based Tests for Computer Administration: Lessons Learned from 30 Years of Mode Effects Studies in Education. *Practical Assessment, Research and Evaluation*, *27*(Article 22), 1–16.

Mertens, U., Finn, B., and Lindner, M. A. (2022). Effects of Computer-Based Feedback on Lower- and Higher-Order Learning Outcomes : A Network Meta-Analysis. *Journal of Educational Psychology*, *114*(8), 1743–1772.

Murniati, Sofia, Sudirman, and Aji, D. P. (2023). Development of Computer-Based Assessment Instruments Nuclear Physics Introduction Course as An Effort to Grade Authentically. *Jurnal Penelitian Pendidikan IPA*, *9*(7), 5103–5110. https://doi.org/10.29303/jppipa.v9i7.4416

Negoro, W. S., Utama, U. P., Agustin, F., Utama, U. P., Nasution, M. I. A., Utama, U. P., Muhatrii, M., Utama, U. P., Ramadhanu, A., Amri, S., and Utama, U. P. (2021). *Participants Evaluation of Computer-Based Writing Exam Eligibility in High School Using Simple Additive Weighting Method.*

Ngiam, W.X.Q. (2024) Mapping visual working memory models to a theoretical framework. *Psychon Bull Rev* **31**, 442–459. https://doi.org/10.3758/s13423-023-02356-5.

Niragudi, S. A. H. (2021). B . A . Degree St Udents ' Attitude Towards Computer-Based Testing. *IJCRT*, *9*(3), 6196–6203.

Nurhikmah, H., Febriati, F., and Ervianti. (2021). The Impact of Computer- based Test and Candidates ' Ability in Computer Self - Efficacy On Mathematics Learning Outcomes. *Journal of Education Technology*, *5*(4), 603–610.

Odo, J. S. (2019). *Candidates ' Perception Of Computer Based Examination In Federal College Of Education Eha-Amufu*. *15*(1), 1–6.

Oladele, J. I. (2021). Computer-Adaptive-Testing Performance for Postgraduate Certification in Education as Innovative Assessment. *International Journal of Innovation, Creativity and Change. Www.Ijicc.Net*, *15*(9), 87-. www.ijicc.net

Paci, H., and Kote, N. (2023). Identification of possible cheating on computer-based exams . 2 . Technics , Advantages and Disadvantages of. *INTERNATIONAL SCIENTIFIC JOURNAL*, *8*(4), 140–143.

Panphet, N., Junpeng, P., and Intharah, T., (2024) "Utilizing Decision Tree

Analysis for Optimizing Test Item Selection in the Development of Mathematical Procedures in Measurements and Geometry," 2024 International Technical Conference on Circuits/Systems, Computers, and Communications **(ITC-CSCC)**, Okinawa, Japan, 2024, pp. 1-6, doi: 10.1109/ITC-CSCC62988.2024.10628122

Pengelley, J., Whipp, P. R., and Malpique, A. (2024). A testing load: a review of cognitive load in computer and paper-based learning and assessment. *Technology, Pedagogy and Education*, *1-17*. https://doi.org/10.1080/1475939x.2024.2367517

Pramono, A. J. B. (2023). COMPUTERIZED ADAPTIVE TEST (CAT) OPTIMIZATION USING THE EFFICIENCY BALANCED INFORMATION (EBI) METHOD. *Proceeding of International Conference on Education, Society and Humanity*, *01*(01), 320–330.

Rajiha, N., and Rini, S. (2023). COMPUTER-BASED STANDARDIZED TESTING (CBT) SUMMATIVE ASSESSMENT : ONE INSTITUTION EXPERIENCE. *JISAE (Journal of Indonesian Candidate Assessment and Evaluation)*, *9*(1), 1–7.

Rice, N., Pêgo, J. M., Collares, C. F., Kisielewska, J., and Gale, T. (2022). The development and implementation of a computer adaptive progress test across European countries. *Computers and Education: Artificial Intelligence*, *3*, 1–11. https://doi.org/10.1016/j.caeai.2022.100083

Schurig, M., Jungjohann, J., and Gebhardt, M. (2021). Minimization of a Short Computer-Based Test in Reading. *Front. Educ*, *6*(84595), 1–12. https://doi.org/10.3389/feduc.2021.684595

Senja, P., Ha, S., Astuti, I., and Enawaty, E. (2024). Candidate Perceptions Regarding the Use of a Computer-Based Test (CBT) System at State Junior High School 3 Sungai Betung. *Journal of Educational Learning and Innovation*, *4*(1), 120–129. https://doi.org/10.46229/elia.V4i1

Sheu, L. A., and Evanero, V. O. (2022). Comparability of computer based and paper and pencil test on Candidates' performance in educational assessment course at the Federal University Gusau, Nigeria. *Journal of Educational Research in Developing Areas*, *3*(2), 114–126. https://doi.org/10.47434/jereda.3.2.2022.114

Shin, J., Chen, F., Lu, C., and Bulut, O. (2022). Analyzing Candidates ' performance in computerized formative assessments to optimize teachers ' test administration decisions using deep learning frameworks. *Journal of Computers in Education*, *9*, 71–91.

Sinha, A. (2024). Computer-based test , close supervision : How IITs Invest in democracy . *The Indian Express Journalism of Courage*.

Supriyati, Y., Iriyadi, D., and Falani, I. (2021). THE DEVELOPMENT OF EQUATING APPLICATION FOR COMPUTER BASED TEST IN PHYSICS HOTS CATEGORY. *Journal of Technology and Science*, *11*(1), 117–128.

Ugwu, K. T., and Ogbozor, E. J. (2021). INTERNATIONAL JOURNAL OF THE ARTS AND SCIENCES, (IJOTAS) VOL 4 No 1 2021 E-ISSN: 2616-096X. *INTERNATIONAL JOURNAL OF THE ARTS AND SCIENCES*, *4*(1), 14–24.

Ukwueze, C. A., and Uzoagba, O. N. (2021). ICT Literacy and Readiness for Computer Based Test Among Public Secondary School Candidates in Anambra State. *New Media and Mass Communication*, *97*(2012), 1–14. https://doi.org/10.7176/nmmc/97-01

Usman, K. O., and Olaleye, S. B. (2022). Effect of Computer Based Test (CBT) Examination on Learning Outcome of Colleges of Education Candidates in Nigeria. *Mathematics and Computer Science*, *7*(3), 53–58. https://doi.org/10.11648/j.mcs.20220703.14

Wang, D., Ma, W., Cai, Y. (2024) A general nonparametric classification method for multiple strategies in cognitive diagnostic assessment. Behav Res 56, 723–735. https://doi.org/10.3758/s13428-023-02075-8

Wu, X., Xu, T. and Zhang, Y. (2023) Research on the data analysis knowledge assessment of pre-service teachers from China based on cognitive diagnostic assessment. *Curr Psychol* **42**, 4885–4899 . https://doi.org/10.1007/s12144-021-01836-y

Yilmaz, F. N. (2021). Comparing Paper-Pencil and Computer-Based Tests: A Meta-Analysis Study in The Sample of Turkey. *Eurasian Journal of Educational Research*, *93*, 279–300. https://doi.org/10.14689/ejer.2021.93.13

Yu, W., and Iwashita, N. (2021). Comparison of test performance on paper-based testing (PBT) and computer-based testing (CBT) by English-majored undergraduate Candidates in China. *Language Testing in Asia*, *11*(1). https://doi.org/10.1186/s40468-021-00147-0

Zhuang, Y., Sun, J., and Liu, J. (2023). Retracted: Diagnosis of Chronic Kidney Disease by Three- Dimensional Contrast-Enhanced Ultrasound Combined with Augmented Reality Medical Technology. *Journal of Healthcare Engineering*, *2021*(5542822), 1–12. https://doi.org/10.1155/2021/5542822

Zhai, X., Haudek, K.C. and Ma, W. (2023) Assessing Argumentation Using Machine Learning and Cognitive Diagnostic Modeling. Res Sci Educ 53, 405–424. https://doi.org/10.1007/s11165-022-10062-w