Development and Usability Evaluation of an Intelligent Quiz Repository with AI-based Duplication and Similarity Detection for Instructional Support

Jayson A. Batoon *, Menirissa D. De Belen, Reylan M. Evale, and Digna S. Evale

College of Information and Communications Technology, Bulacan State University, Bulacan, Philippines
Email: jayson.batoon@bulsu.edu.ph (J.A.B.); menirissa.debelen@bulsu.edu.ph (M.D.D.B.); reylan.evale@bulsu.edu.ph (R.M.E.); digna.evale@bulsu.edu.ph (D.S.E.)
*Corresponding author

Manuscript received December 14, 2025; revised January 23, 2026; accepted February 18, 2026; published August 12, 2026

***Abstract*—This study presents the development and usability evaluation of an intelligent quiz repository system enhanced with an Artificial Intelligence (AI)-based duplication and similarity detection mechanism to support quiz creation and management in higher education. The administration of quizzes and the management of item banks will be positively affected by the system in higher education. The whole idea is to create an integrated platform for the instructors to keep their quiz items in an easy way: store, classify and generate. The system's major constituents are a quiz repository, a quiz generator, a dashboard that contains analytics, and an AI that has a cosine similarity method for tracking duplications and discovering repeating or almost identical questions. The study followed a developmental-descriptive path, concentrating on system architecture, user-interface design, and iterative refinements based on faculty input. The usability principles derived from Nielsen's heuristics and Software Usability Measurement Inventory (SUMI) were incorporated into the design to ensure clarity, uniformity, and user-friendliness. The findings reveal that the system's design has effectively addressed the issues of redundancy, disorganization, and manual workload, thus paving the way for new and expanded usability testing of the system in the near future.**

***Keywords*—quiz repository, Artificial Intelligence (AI) duplication tracking, system design, quiz generation, usability evaluation**

I. Introduction

The possibility of having such systems in place is emphasized more than ever in the uppermost educational institutions worldwide due to the digital era. The generation and management of traditional quizzes are considered to be the most or nearly the most time-consuming and tiring tasks for the teachers who have to cope with the whole process—from producing to handing out and collecting quizzes—and consequently much time wasting and [1]. In the case of many institutions, quiz item banks grow to such an extent that they can be freed from problems such as redundancy and duplicating only by proper organization and monitoring, which, however, are very hard to get rid of [2]. Such duplication, beside being a waste of storage and management resources, also diminishes the variety of the assessment items and hence jeopardizes the validity of item-bank use on a case-by-case basis.

Another area that is critical and in need of urgent measures is the automatic extraction of quizzes from learning materials. Question-generation systems have attracted a lot of research attention but still are far from being fully used and accepted in educational settings primarily due to the problems of matching the generated items to the learning objectives, keeping quality, and the like [3]. So, the potential for automation is not the same as the actual deployment of the repository systems that not only generate but also handle item banks and detect duplicates [4].

In the context of local schools, it is most likely the instructors that will reuse or repurpose quiz items since creating new ones is a labor-intensive process. But in the absence of a centralized and intelligently managed repository, the reuse of items can give rise to problems of injustice, deterioration of item quality, and dullness in item variety. Even if, at the end of the day, automated generation is available, instructors must still review, revise, and validate the generated items before using them in an assessment. Furthermore, item-writing quality continues to be a concern: research indicates that item-writing flaws significantly impact difficulty and discrimination in examinations [5]. Without intelligent support for item banking and reuse, educators risk amplifying these flaws through repeated or near-duplicate items, rather than mitigating them. Coupled with the challenge of automatically generating meaningful items, the gap becomes clearer: there are still no widely implemented platforms that integrate quiz item repository management, automated generation, and duplication tracking under one system tailored for academic assessment workflows [6].

The general objective of this study is to develop a smart quiz repository and an Artificial Intelligence (AI)-based quiz generation system integrated with duplication tracking that increases assessment management efficiency, integrity, and usability so that academic institutions will reap the benefits. Specifically, this study seeks to:

(1) Design an intelligent quiz repository and generation system that allows instructors to create, store, organize, and generate quiz items efficiently through a centralized platform.
(2) Integrate an AI-based duplication tracking mechanism that automatically detects and manages similar or redundant quiz items to ensure item bank integrity and reduce repetition.
(3) Evaluate the system's usability and effectiveness using a validated User Interface (UI) Evaluation Instrument based on Nielsen's Usability Heuristics and the Software Usability Measurement Inventory (SUMI).
(4) Analyze users' perception of the system in terms of efficiency, learnability, consistency, aesthetics, and satisfaction.

doi: 10.18178/ijiet.2026.16.8.2665

(5) Identify potential areas for improvement in the system's design and functionality based on the results of the evaluation.

While the term quiz generation is used in this study, its scope is defined within the context of repository-based assessment management. The proposed system does not utilize artificial intelligence to autonomously generate new quiz items from instructional content. Rather, quiz generation is implemented as a rule-based assembly process that compiles quizzes from existing items in the repository according to instructor-specified parameters such as topic, difficulty level, and number of questions. The application of artificial intelligence in the system is primarily focused on duplication and similarity detection, functioning as a quality-control mechanism to maintain item diversity and ensure the integrity of the quiz repository.

The study, therefore, seeks to tackle these connected issues through the creation and testing of a smart quiz repository and generation system integrated with AI-powered duplicate detection. The combination of repository management, AI-generated quiz items, and duplicate detection is expected not only to lighten the load of teachers but also to improve the quality of the item bank, increase the speed of quiz generation, and allow for continuous and diverse item repository. The primary goal is thus to improve assessment efficiency, increase item variety, and maintain academic integrity in quiz management and generation practices.

II. Literature Review

Assessment is one of the most vital and important parts in the whole cycle of teaching and learning, which enables teachers or educators to keep track of student success, find out how the students are having difficulty in learning, and gradually improve their teaching methods. At the same time, the majority of educational institutions are turning to automated assessment systems due to the growing integration of technology into education [6]. The digital examination tools are being favored over their traditional counterparts for their efficacy, quick feedback, and personalized teaching [7]. On the other hand, even though the trend is shifting to digital, many educators still prefer to manually prepare quizzes and this has been recognized as a labor-intensive, dull, and error-prone procedure [1]. The application of these methods restricts the amount of assessments that can be conducted and consequently affects their quality and credibility in a negative way.

The item duplication and redundancy problem in the repositories is one particular issue that quiz management has to deal with. Repositories of multiple-choice questions that contain a lot of content usually have some duplicate or semantically similar items. Thus, the diversity of the questions decreases and the size of the database increases [8]. In addition, the increasing number of poor quality items in the banks are one of the problems associated with the expansion of repositories. This is especially true when there are no systematic organization and duplicate detection mechanisms in place [9].

The manual item review and curation processes are the most ineffective solutions. Checking duplicate questions manually takes a huge amount of time and effort and is also very prone to errors resulting in the recommendation that intelligent automation is the only way to manage item quality effectively [2]. Furthermore, the tests based on item banks that are poorly managed and have redundant or damaged items are less valid and reliable [5]. Therefore, the automation in filtering and organizing quiz items becomes a necessity where the content quality and diversity are also being ensured.

Automatic Item Generation (AIG) has been the center of attention in recent studies, which utilizes Artificial Intelligence (AI) and Natural Language Processing (NLP) to produce fresh questions from the source material. A comprehensive review revealed that the use of AI-powered question generation, greatly lightens a teacher's burden but frequently the produced questions are not appropriate both academically and cognitively [1]. Other researchers stated that AIG systems are capable of generating large volumes of questions but the quality of the questions produced is the main concern [3, 7]. Despite these advances, most AIG systems operate independently of item repositories and do not include mechanisms for duplication detection or metadata tagging. This disconnect creates inefficiencies in educational contexts where faculty reuse question sets across multiple terms or subjects. Automation must be complemented by intelligent quality control features, such as AI-based similarity tracking to maintain the integrity of large question banks [10].

During the years that have passed, AI has gradually been incorporated into the applications of text similarity analysis and duplicate detection in educational datasets. The multimodal method, which combines text and image embeddings, is one such integration, which claims a significant increase in the precision and recall levels of detection [11]. Besides this, natural language processing and graph analysis were also utilized in the same context and showed that it was possible to have more or less constant monitoring of the relearning of concepts in educational repositories [8].

The findings not only confirm the necessity of AI in the manual and at the same time very tedious task of checking for redundancies but also even hinted at the possibility that AI techniques could be working silently in the quiz repository system, which would allow the teachers to have a small but high-quality database, thus ensuring fairness and efficiency. However, the area of investigation has still to reach the theoretical notion of AI duplication tracking opening up the co-requisite of repository management and usability evaluation in the academic sector; this is precisely where the present study is headed.

The usability factor has gradually shown itself to be a decisive quality criterion not only for the acceptance of educational technology systems but also for their effective use. The ISO 9241-210 standard for Ergonomics of Human-System Interaction provides an enlightening guide to the designers of human-centered systems on how to reap the benefits of improved effectiveness, efficiency, and user satisfaction [12]. Furthermore, the renowned usability heuristics are now not only giving the designers a theory but also making it a very easily appliable one [13]. To measure usability empirically, the Software Usability Measurement Inventory (SUMI) remains one of the most validated instruments in software evaluation [14]. It assesses five key

dimensions—efficiency, affect, helpfulness, control, and learnability, providing structured insights into user satisfaction. Studies applying these frameworks in academic settings consistently show that usability directly affects faculty acceptance and sustained use of new learning technologies [3, 7].

Quiz management systems that are used alongside learning management systems like Moodle or Canvas have been installed in numerous educational institutions around the globe, yet the systems lack an AI module which could perform the intelligent tracking of duplicates [1]. Moreover, in the Philippines, several universities still depend on very outdated question banks that are created manually, predominantly in the form of spreadsheets or on local drives with no version control. This situation goes hand in hand with the above-mentioned manuals for quiz management causing redundancy and inefficiency [2]. Thus, there is still a great need for a localized intelligent system that can manage quiz creation, duplicate detection, and user-centered design evaluation simultaneously.

The reviewed literature unanimously points to the necessity for intelligent, automated assessment systems that would be a great help to the teachers in the facilitation of the handling of the enormous quiz item repository. Modern research on Automatic Item Generation or AIG similarly asserts that AI can create assessment papers but the organizing of the repository and controlling the duplicates are not emphasized [1, 7, 11]. In contrast, papers on duplication detection grant AI the ability to correctly identify common questions, yet these algorithms are seldom accentuated in total quiz management systems and hence are not of widespread use [2, 8, 11].

In addition, Nielsen's heuristics and SUMI are excellent methods for assessing system usability [12], but very few studies have been conducted to evaluate the user-friendliness of AI-based quiz systems. Thus, it is clear that not much effort has been invested in the creation of a coupled system whose parts are the AI-based quiz system and its user support.

The issue that is being elucidated in the present study is resolved through the proposal and evaluation of an Intelligent Quiz Repository and Generation System that simplifies the item creation and the repository maintenance process while at the same time, ensuring the users' end experience is enjoyable and effective. The study integrates the AI-based duplication detection with the user-friendly interface design thus providing a novel means of improving educational assessment methods and facilitating the digital transition in schools and universities at a faster pace.

III. Materials and Methods

A. Research Design

For this research, a Developmental-Descriptive Research Design was adopted, which is suitable for projects involving the innovation of a new system along with its evaluation in actual conditions. The developmental part is the creation of the Intelligent Quiz Repository and Generation System with AI-Based Duplication Tracking, while the descriptive part is the evaluation of the system's user-friendliness and effectiveness among its intended users-the educators from the College of Information and Communications Technology (CICT) at Bulacan State University.

What developmental research tells is that a systematic formation and assessment of the new technologies aimed to solve the real educational issues is the main concern [15]. The project development stage was planned according to the agile Software Development Life Cycle (SDLC), which is an iterative and flexible method that allows fast prototyping, reception of feedback, and making minor changes. The use of Agile is very suitable for the development of educational technology since it emphasizes the partnership of the developers and the users as well as the adaptability to the users' feedback [16].

The usability and functionality of the system were quantitatively evaluated in the descriptive section of the study by means of the quantitative method. Quantitative descriptive designs are the best way to get measurable data that mirror perceptions and behaviors without altering any of the variables [17]. The employment of this dual design paradigm made it certain that the system developed was not only technically capable but also pedagogically acceptable from the user's standpoint.

B. System Development Process

The Agile SDLC was utilized in developing the system, and it was composed of five loops of iterations—Planning, Design, Development, Testing, and Deployment. The technique allowed each step to become more flawless by considering the user feedback and the tests carried out on the system as the reason of the change.

1) Planning phase

In this stage, the user requirements as well as their needs were gathered through informal interviews and consultations with CICT faculty members who frequently manage quizzes. The discussions revealed issues such as question redundancy, time-consuming quiz compilation, and variations in quality of questions. The researchers documented both functional requirements (quiz repository creation, AI duplication tracking, automatic quiz generation) and non-functional ones (usability, accessibility, security).

2) Design phase

The requirements were transformed by the researchers into design prototypes and workflow diagrams. The system's architecture, diagrams illustrating relationships between entities (Entity-Relationship Diagrams), and user interface mockups were created as per usability guidelines. Consequently, the interface was made non-confusing, easy to use, and accessible to people of different technical backgrounds.

3) Development phase

Initially, the interface was designed in Figma, which allowed the development team to make very similar prototypes, structured wireframes, and a cohesive interface that follows the modern User Interface/User Experience (UI/UX) standards. Later on, the front-end was built with React.js that offers the excellent virtual Document Object Model (DOM) rendering, component-based architecture, and high scalability for the migration of interactive and responsive sites and building of such sites. Recent research depicts React.js as the leading technology for future web

application development due to its reusability, performance, and maintainability [18]. The modules that are of major importance are:

- User Management: handles login, role assignment, and access levels;
- Quiz Repository: stores categorized quiz items with metadata such as subject, difficulty level, and topic.
- Quiz Generator: automatically compiles quizzes based on instructor-specified criteria.
- AI Duplication Tracker: detects and flags redundant or highly similar questions.
- Analytics Dashboard: visualizes repository statistics, duplication reports, and system usage data.

4) Testing phase

Functional and usability testing were performed using iterative cycles. Unit testing demonstrated that each module operated properly and concurrently integration testing verified that the system parts were able to operate together. The accuracy of the AI duplication tracker was validated by test datasets containing similar questions with the same intention. The feedback coming from the registered users turned out to be very important and therefore, the changes incorporated in the next version were mainly focusing on stability and ease of use.

5) Deployment phase

The evaluators of the CICT faculty could then access the local server where the system was installed right after successful testing. A brief orientation and demonstration session were held before the formal usability evaluation took place.

The whole development process led to a user-centered system that was in line with the top educational software engineering practices [19].

C. Respondents of the Study

The study primarily concentrated on the staff of Bulacan State University College of Information and Communications Technology (CICT). They were included in the study through purposive sampling because of their qualifications—mainly, their relevant experience in quiz creation and assessment design which made them the best candidates for the system evaluation.

Participants were recruited using a purposive sampling approach to ensure that respondents possessed direct experience in quiz creation and assessment management. Faculty members from the College of Information and Communications Technology (CICT) at Bulacan State University were formally invited through official institutional communication channels after securing approval from the college administration. Participation was voluntary, and informed consent was obtained prior to data collection. The recruitment strategy prioritized instructors who regularly design quizzes and examinations, as their domain expertise and familiarity with assessment workflows were essential for providing meaningful usability feedback. This targeted recruitment process ensured that the evaluation results reflected authentic instructional use cases rather than simulated or novice interactions.

Table 1 shows the Respondent Profile section (Part I of the instrument) furnished information from each participant, and it encompassed the descriptive details essential for the accurate data interpretation:

(1) Role in the system: Admin, Subject Expert, or Subject Member.
(2) Years of teaching/academic experience.
(3) Level of digital literacy: Beginner, Intermediate, or Advanced.

The factors mentioned supported the feedback regarding the usability. Research indicates that digital competencies and educators' backgrounds are the primary factors influencing users' contentment with e-learning systems [20]. The demographic data of the participants were analyzed through frequency and percentage distributions in order to both define the participants' characteristics and to check for any connections between the profile variables and the usability perceptions.

Table 1. Profile of respondents ($N = 40$)

Profile Category	Subcategory	Frequency (f)	Percentage (%)
Role in the System	Admin	5	12.5%
	Subject Expert	12	30%
	Subject Member	23	57.5%
Years of Teaching Experience	1–5 years	18	45%
	6–10 years	12	30%
	11 years and above	10	25%
Level of Digital Literacy	Beginner	6	15%
	Intermediate	20	50%
	Advanced	14	35%

D. Research Instrument

A User Interface (UI) Evaluation Instrument was applied to measure the degree of satisfaction of users with the system in terms of usability and effectiveness. The instrument is based on the following international standards:

- ISO 9241-210: Human-centered design principles for ensuring efficiency, effectiveness, and user satisfaction.
- Nielsen's Usability Heuristics: Core interface design principles, including visibility of system status, consistency, and error prevention.

Table 2. Nielsen's usability heuristics-based indicators (Focus: Interface design, navigation, consistency, error prevention, and aesthetics)

Item No.	Statement	Source Principle
1	The system clearly shows what is happening (e.g., loading indicators, confirmation messages).	Visibility of System Status
2	System feedback appears quickly after I perform an action.	Visibility of System Status
3	Buttons, icons, and menus are consistent across all screens.	Consistency and Standards
4	The system follows familiar interface conventions (e.g., placement of navigation and action buttons).	Consistency and Standards
5	I can easily undo or cancel an action when I make a mistake.	User Control and Freedom
6	The system prevents me from committing serious errors (e.g., deleting items accidentally).	Error Prevention
7	Error messages are clear and help me understand how to fix the problem.	Error Recovery
8	The design of the system is visually appealing and not cluttered.	Aesthetic and Minimalist Design
9	Texts, icons, and labels are readable and appropriately sized.	Aesthetic and Minimalist Design
10	Help or guidance (tooltips, prompts, or instructions) is easily available when needed.	Help and Documentation

Table 2 displays the indicators that are founded on Nielsen's usability principles and were employed to measure the system's interface design, navigation clarity, consistency, error prevention, and overall aesthetic quality.

Software Usability Measurement Inventory (SUMI): A tool that has become famous for evaluating the five main usability factors—efficiency, affect, helpfulness, control, and learnability. Table 3 shows the results from evaluations of the SUMI studies used to measure the usability dimensions in the study.

Table 3. Software Usability Measurement Inventory (SUMI)-based indicators (Focus: Efficiency, affect, helpfulness, control, and learnability)

Item No.	Statement	SUMI Dimension
1	I can complete tasks quickly using the system.	Efficiency
2	The system reduces the time I spend preparing or managing quizzes.	Efficiency
3	I feel comfortable and satisfied while using the system.	Affect
4	I enjoy working with the system because it feels intuitive.	Affect
5	The system's prompts and instructions are helpful when I am unsure what to do.	Helpfulness
6	The system provides sufficient information for me to accomplish my tasks.	Helpfulness
7	I feel in control while navigating through different functions.	Control
8	The system responds accurately to my inputs or selections.	Control
9	I can easily learn how to use the system even without formal training.	Learnability
10	The system's layout helps me remember where to find certain functions.	Learnability
11	The duplication tracking feature works intelligently and detects redundant quiz items effectively.	Efficiency
12	The quiz repository is organized in a logical and easy-to-navigate structure.	Helpfulness
13	I can generate quizzes easily according to my desired parameters.	Efficiency
14	Using the system makes quiz creation and management less stressful.	Affect
15	Overall, the system is user-friendly and meets my expectations.	Summary (Overall Usability)

The questionnaire contained two sections:

- Part I—Respondent Profile, which gathered background information for descriptive analysis.
- Part II—System Evaluation, which included 25 usability statements rated on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree).

The instrument, so to speak, was partly influenced by the assessments of three specialists belonging to different fields, namely, two IT professors and a UX specialist. This way, the content was validated which means the items were in line with the construct, therefore, the reliability of the outcomes of the construct being measured was increased [21]. The trial of the items with five faculty members who were not involved in the main study confirmed the clarity and internal consistency of the items.

E. Data Gathering Procedure

At first, the researchers professionally made their request and received a permit, then the Dean of the CICT gave the approval and the consent forms were made available for the people who would voluntarily participate in the study. After that, a system presentation was made, and an introduction was organized where the users got acquainted with the software and its main features such as the AI replication detector, repository development module, and quiz creation tool. Throughout the system usage period, the users were given login information and they were told to perform real tasks in the system, which would allow them to experience real teaching situations and use the main features of the system. The survey administration phase began following these activities where users completed the UI Evaluation Instrument through a Google Form link. The collection, mixing, and safe storage of all responses were done by an automated system running a secure database, after which Microsoft Excel was used for organization and the first validation through systematic tabulation. The entire process of data validation and analysis took place concurrently by checking the responses for completeness of data, numerical data was summarized through various statistical methods, and respondents' comments were sorted according to different topics in order to highlight the user-friendliness issues that occurred most often along with the system design ideas that were deemed most important. Utilizing strict data collection protocols not only augments the credibility of results but also their replicability [22].

F. Data Analysis

By means of descriptive statistics, the study analyzed the quantitative data that was obtained through the evaluation instrument. The application of the following analytical techniques took place:

- Frequency and Percentage—to describe respondent profiles.
- Mean and Standard Deviation—to determine the overall usability score for each dimension.
- Qualitative Descriptions—to interpret mean scores using the following scale:

An interpretation scale to aid in classifying classifications of computed mean scores into various descriptive categories is presented in Table 4.

Table 4. Interpretation of mean scores

Mean Range	Qualitative Description
4.50–5.00	Excellent / Very Highly Usable
3.50–4.49	Very Good / Highly Usable
2.50–3.49	Good / Moderately Usable
1.50–2.49	Fair / Slightly Usable
1.00–1.49	Poor / Not Usable

These techniques are widely recommended for usability studies to provide a clear interpretation of users' perception [23].

G. Ethical Considerations

The study was conducted following the strictest ethical standards throughout its whole course that were applied to human subjects. The potential participants received prior to the taking of part in the study a very comprehensive and detailed explanation of the whole process that included the purpose, duration, and methods used. They were also assured that the information they provide would be stored in a way that maintains confidentiality and would be solely used for research purposes. For instance, no personal identifiers were used, etc. Anyone could participate, and the respondents

could at any point in time withdraw their answers without any negative consequences.

The study respected AERA's ethical principles of respect for persons, beneficence, and justice [24]. Each phase of the study was made only after obtaining approval from the CICT Dean's office and only then could the research tools be used.

H. Summary of Methodological Alignment

The alignment of methodology in the research is presented in Table 5, which shows the links between the strategic objectives of the research and the corresponding methods, instruments, and analytical procedures. The alignment guarantees the consistency not only between the goals of the study, the methods of system development and evaluation applied, and the data gathering and interpretation tools used but also indicates that the technical innovations and user-experience components were both systematically handled in the research.

Table 5. Summary of methodological alignment

Research Objective	Method/Process	Instrument/Analysis
1. Design and develop the system	Agile SDLC, iterative prototyping	System documentation and functional testing
2. Integrate AI duplication tracking	NLP cosine similarity algorithm	Accuracy testing using test datasets
3. Evaluate system usability	Descriptive quantitative method	UI Evaluation Instrument (Likert scale)
4. Analyze users' perception	Survey and descriptive analysis	Mean, SD, qualitative interpretation
5. Identify improvement areas	Open-ended feedback and observation	Thematic analysis

The study proceeded to apply the chosen methodological framework to the two aspects of technical innovations and the evaluation of user's experience with the smart quiz repository system effectively and thoroughly at the same time.

IV. Result and Discussion

The Intelligent Quiz Repository and Generation System with AI-Based Duplication Tracking is a web-based platform that facilitates educator's academic activities on a daily basis. The system contains five different modules that perform various functions but are equally involved in the processes of quiz management. The first module is the Quiz Repository Module, which is the main storage of questions where the teachers can sort their items by subject, topic, type, and difficulty. The teachers can now forget about the time-consuming process of storing quizzes in different formats like Word documents and spreadsheets and so on. It is indeed an excellent organized storage place for quiz items but also a method of keeping them in good condition for a long time. Next, the Quiz Generator Module lets a user produce tests automatically just by adjusting such factors as the number of items, topics covered, or the mix of hard and easy ones. The usefulness of this tool minimizes the time needed for the amplification of manual quizzes to an extent while at the same time guaranteeing that the evaluations will not be biased.

The AI-Based Duplication Tracking Module, which is the most significant improvement in the system, employs the cosine similarity algorithm to measure similarity between new or current items in the repository. The system performs an automatic check for the identical or near-duplicate items and notifies the instructor if redundancy is detected, as soon as the questions are added. Hence, it not only minimizes the repetition but also guarantees that there are diverse items and that the quality of the item bank is improved.

A. Key Features of the System

Fig. 1 displays the Quiz Repository Interface which visually presents topics and their different teaching aids in an appealing card layout. The card implementation not only encourages navigation and locating items but also confirms the requirement for user-friendly and intuitive interfaces which has been voiced by several researchers [12]. In fact, the organization of the repository mirrors the argument that has been mentioned as the main reason for neat examination repositories that will allow easier access and lesser duplication in item management [25].

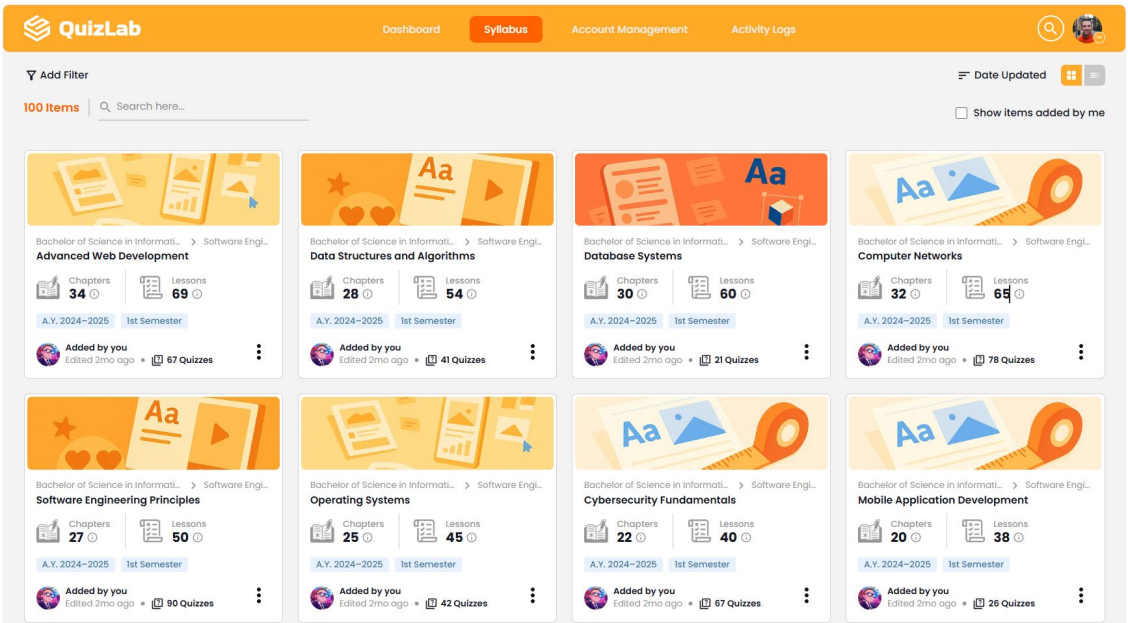

Fig. 1. Quiz repository interface.

The constant use of colors, icons, and metadata display across the entire system is indicative of the principles of

consistency and aesthetics of Nielsen together with minimalist design [13]. The interface is at the same time a clarifying and a slowing down tool for the user's thinking process thus enabling teachers to quickly find the educational resources that had been the subject of the manual assessment workflow complaints frequently [1].

Although the system includes a quiz generation feature, this function operates as a supporting mechanism rather than a core AI-driven process. Quiz generation is implemented through predefined selection rules that assemble quizzes from the existing repository, ensuring alignment with instructional requirements. In contrast, the AI-based duplication and similarity detection module represents the main technical contribution of the system. By identifying exact, near-duplicate, and semantically similar questions during quiz preparation, the AI component enhances item quality control and prevents redundancy, thereby strengthening the overall effectiveness of quiz management.

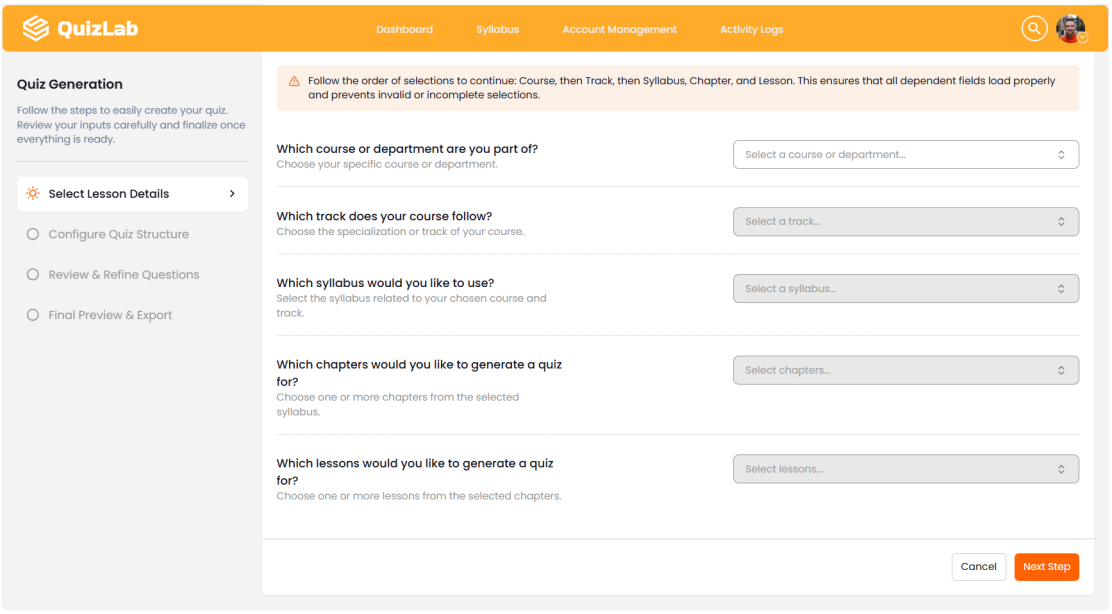

Fig. 2. Quiz generation interface.

The guided quiz generation sequence is shown in Fig. 2, where the instructor first selects the Course, Track, Syllabus, Chapters, and Lessons to create a quiz. This process ensures that the quiz items are aligned with the instructional content, which is one of the main problems in automated question generation systems pointed out by the study [7].

The wizard-style interface reduces the chances of users making mistakes and ensures that the task is done in the same manner every time, so it is in line with the established heuristics for user control, error prevention, and system status visibility [13]. The extensive hierarchical dropdown filtering of options acts as an additional measure to prevent mismatches or incomplete selections, thus solving the issue of alignment in traditional quiz preparation [1].

The interface presented in Fig. 3, which is referred to as the AI Duplication Tracking Analysis Interface, illustrates the system's detection of identical and very similar multiple-choice questions during the quiz development process. The interface marks the questioned pairs with color-coded indicators: red signifies the closest or perfectly identical duplicates and blue is for possible semantic duplicates. By this technique, items are shown next to one another, and the educators can then confirm the AI-generated similarity ratings by actually checking before the quiz content is done.

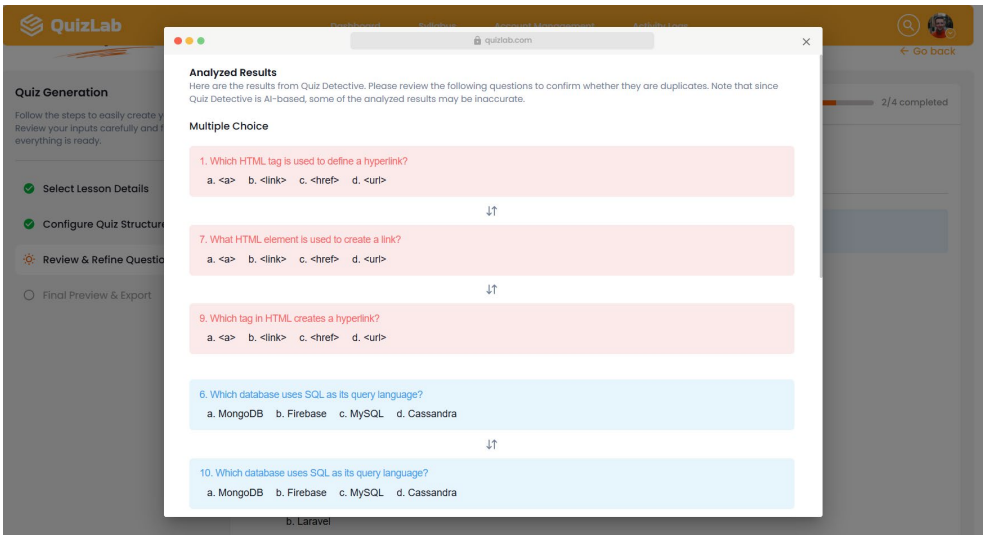

Fig. 3. AI duplication tracking analysis interface.

The system's capability to spot repetitions is demonstrated by the creation of unique clusters. For example, there are three questions: "*Which HTML tag is used to define a hyperlink?*", "*What HTML element is used to create a link?*", and "*Which tag in HTML creates a hyperlink?*" that are under the same category of identity inquiries because they come from the same intent, learning goal, and answer set. Likewise, the question "Which database uses SQL as its query language?" is repeated and thus automatically marked to prevent the production of quizzes containing the same content.

The interface presented shows the capability of AI duplication tracker to bring down the count of duplicate items present in the repository to a considerable extent. The researchers in the area have already mentioned this problem earlier. The process of manual duplication detection takes a lot of time and may even miss a few cases, which is a common scenario in large question banks [2]. One of the approaches through which the system gets around this restriction is the automatic clustering of the same items. Also, the successful recognition of semantically similar questions is in line with the findings that AI and NLP-based similarity models (e.g., cosine similarity) are able to detect overlapping content accurately even when the questions are differently worded [8, 11].

B. Performance of the AI-based Duplication Tracking

The AI-Based Duplication Tracking Module of the system underwent a thorough test of its accuracy, which was based on the identification of quiz items as identical, near duplicates and semantically similar respectively. The module used the cosine similarity algorithm, a method for determining the proximity of the text vectors, to conclude that the two questions are interchangeable in meaning. The system during the testing period very accurately selected the exact duplicates and the distinguishing of the items that were only slightly different in wording, punctuation, or phrasing was done consistently.

This aspect of the study was looked into and it became apparent that the merging of the NLP-based similarity calculation with preprocessing and graph-based clustering [8] is a really strong technique for dealing with multiple-choice item banks. The researchers in their paper stated that cosine similarity can be very effective in precisely capturing even the most nuanced semantic relations between questions, which is in a way a confirmation of the present study's accuracy claims. Moreover, the system's capacity to detect near-duplicates has also constituted one of the reasons for the overall conclusion that AI-based duplication detection models are much better than manual checks, particularly for large-scale archives where duplications are regularly found and often unnoticeable [11]. Table 6 displays the accuracy results of the AI-based duplication tracking module.

The statistics presented in Table 1 reveal that the AI-Based Duplication Tracking Module has a very high overall accuracy of 91.25% when different item similarity conditions are applied. The perfect identification of exact duplicates is considered as one of the main factors that contributes to the validity of the algorithm in processing text structures that are similar, which corroborates the prior works on similarity detection [8]. The high accuracy rate in the detection of near duplicates (90%) and semantically similar questions (80%) is in line with the argument that the cosine similarity can capture fine semantic distinctions [11]. The very low frequency of false positives in dissimilar questions (95% accuracy) shows that the system is careful not to limit item variety inadvertently, thus, the redundancy problem is viewed as less serious [2, 9].

Table 6. Accuracy results of the AI-based duplication tracking module

Type of Item Pair	Description	No. of Test Pairs	Correctly Detected	Accuracy (%)
Exact Duplicates	Identical wording, structure, and meaning	20	20	100%
Near-Duplicates	Slight wording changes but same meaning	20	18	90%
Semantically Similar Questions	Different structure but similar learning objective	20	16	80%
Dissimilar Questions	Different content and intent	20	19	95%
Overall System Accuracy	Combined performance across all categories	80	73	91.25%

C. Usability Evaluation Based on Nielsen's Heuristics

The usability of the system was assessed very positively by many of Nielsen's usability heuristics indicators which are represented in Table 7. The mean ratings of the indicators fluctuated from 4.43 to 4.65 with a Standard Deviation (SD) from 0.48 to 0.58, which means that the agreement among the raters was very strong and hence the variation in their evaluations was very small. The findings are in line with the literature that describes similar patterns in user studies.

Table 7. Results of Nielsen's usability heuristics-based evaluation (N = 40)

Statement (Summary)	Mean	SD	Interpretation
System clearly shows what is happening (visibility of status)	4.65	0.48	Very Highly Usable
System provides quick and timely feedback	4.55	0.50	Very Highly Usable
Buttons, icons, and menus are consistent	4.60	0.49	Very Highly Usable
System follows familiar interface conventions	4.52	0.51	Very Highly Usable
Users can easily undo/cancel actions	4.43	0.58	Highly Usable
System prevents serious errors (error prevention)	4.48	0.54	Highly Usable
Error messages are clear and helpful	4.50	0.52	Very Highly Usable
System design is visually appealing and uncluttered	4.58	0.50	Very Highly Usable
Texts, icons, and labels are readable	4.63	0.49	Very Highly Usable
Help or guidance is easily available	4.47	0.55	Highly Usable
Overall Mean	**4.54**	**0.52**	**Very Highly Usable**

The items about the system status visibility (Items 1 and 2) achieved the highest mean ratings ($M = 4.65$, $SD = 0.48$; $M = 4.55$, $SD = 0.50$) which support the idea that user feedback mechanisms are the most effective way to overcome user uncertainty and hence interactions can take place seamlessly [13]. Moreover, providing system feedback and being transparent are considered crucial when creating user-friendly interfaces [12]. The current study's ranking of

visibility not only confirms the validity of these principles but also matches the findings from other educational systems where prompt feedback has resulted in increased task accuracy and user satisfaction with AI-assisted grading tools [7].

Items regarding consistency (Items 3 and 4) also got high scores (M = 4.60 − 4.52; SD = 0.49 − 0.51), thereby supporting the claim that the uniformity of interface elements caused a tremendous increase in usability of intelligent assessment systems by making user interactions more easily predictable and the mental load less [3]. The paper coincides with the findings of another research that digital assessment systems need to exhibit layout and navigation consistency so as not to confuse and inefficiently teach the instructor [1].

Items related to aesthetic and minimalist design (Items 8 and 9) received one of the highest ratings (M = 4.58 − 4.63; SD = 0.49 − 0.50). This aligns with the common concept that neat and simple visuals not only lessen user frustration but also make the user feel that the system is easy to use [14]. Similarly, researchers have claimed that good visual hierarchy and readability are among the factors that attract users to AI-powered education technology [7]. Thus, the present research findings on aesthetics may be viewed as a close following of the already-established full loop of best practice.

The indicators related to error reduction, erroneous recovery, and user control (Items 5, 6, and 7) produced mean scores that were lower than the main indicators but were still good (M = 4.43 − 4.50; SD = 0.52 − 0.58). This is similar to the case of systems that apply automatic error prevention techniques such as duplication detection and task validation, still requiring users' control support but at a lesser extent [2]. It can also be argued that error content can be reduced through AI-based systems, but users might be needing very strong control features for working with complex automated processes [10]. The larger SD values observed in the study signify that the users had different views about the control and recovery mechanisms which was also the case in other studies.

The average score of 4.54 (SD = 0.52) categorizes the system within the "Very Highly Usable" group, thus supporting the user's adherence to the specified usability heuristics. Other researchers who developed AI-based question generation systems where user-oriented design played a decisive role in high acceptance and favorable user ratings reported similar usability outcomes [3, 7]. The consistency and proximity between the SD values also indicate that users with different levels of digital literacy and teaching experience had equally positive interactions, which is in line with the claim that user-friendly system design can help to close the digital skill gap [20].

D. Usability Evaluation Based on SUMI Indicators

The findings in Table 8 indicate that the system has a high usability rating through the five background factors of SUMI, which are Effectiveness, Affect, Helpfulness, Control, and Learnability. The mean scores varied between 4.47 and 4.63, and the SD values were also quite low (0.48–0.56), so there is a lot of user agreement and a generally good experience with the system.

1) Efficiency dimension

The highest rated efficiency item was "I can complete tasks quickly" (M = 4.58, SD = 0.50) and "Quiz generation is easy" (M = 4.59, SD = 0.50). These results agree with the earlier ones that showed that teaching staff worked less and manual tasks took less time when educational systems were AI-assisted or automated [1].

2) Affect dimension

The user comfort and enjoyment items (M = 4.60 and 4.52) were the ones rated the most highly, meaning that the users experienced a very nice and even delightful interaction with the platform. This goes along with the interpretations that said user satisfaction was increased to a great extent by the user-friendly system design in case of intelligent assessment tools [3]. These very high affective scores are strong confirmations of SUMI's hypothesis that user acceptance is very much connected with positive emotional responses [14].

3) Helpfulness dimension

Helpfulness indicators got the highest ratings (M = 4.48 − 4.60), signaling that the overall user's opinion was that the system's prompts, instructions, and framework were easy to understand and helpful. These results are in line with the principles that highlight the need for proper guidance from the system to do the tasks quickly and efficiently [12]. The clear rating for the repository (M = 4.60) reinforces the principles regarding the need for item banks that are well structured and easily accessible [9].

Table 8. Results of Software Usability Measurement Inventory (SUMI)-based evaluation (N = 40)

Statement (Summary)	SUMI Dimension	Mean	SD	Interpretation
I can complete tasks quickly using the system.	Efficiency	4.58	0.50	Very Highly Usable
The system reduces the time I spend preparing quizzes.	Efficiency	4.55	0.51	Very Highly Usable
I feel comfortable and satisfied while using the system.	Affect	4.60	0.49	Very Highly Usable
I enjoy working with the system because it feels intuitive.	Affect	4.52	0.52	Very Highly Usable
Prompts/instructions are helpful when unsure.	Helpfulness	4.48	0.54	Highly Usable
System provides enough information to accomplish tasks.	Helpfulness	4.50	0.53	Very Highly Usable
I feel in control while navigating the system.	Control	4.47	0.56	Highly Usable
System responds accurately to inputs.	Control	4.55	0.50	Very Highly Usable
I can learn the system even without training.	Learnability	4.63	0.49	Very Highly Usable
Layout helps me remember where functions are.	Learnability	4.57	0.51	Very Highly Usable
Duplication tracking works effectively.	Efficiency	4.53	0.52	Very Highly Usable
Repository is logical and easy to navigate.	Helpfulness	4.60	0.49	Very Highly Usable
I can generate quizzes easily based on parameters.	Efficiency	4.59	0.50	Very Highly Usable
System reduces stress in quiz preparation.	Affect	4.55	0.51	Very Highly Usable
System is user-friendly and meets expectations.	Summary	4.62	0.48	Very Highly Usable
Overall Mean	—	**4.56**	**0.51**	**Very Highly Usable**

4) Control dimension

The control items ($M = 4.47 - 4.55$) indicated that the users thought of themselves as possessing the capability to control the system and also that their actions were being performed correctly. The scores are still within the border of "Usable" and "Very Highly Usable" levels though somewhat lower. Besides, these scores accord with the usability principle which states that users should be granted a powerful feeling of control over the system and be allowed to act freely [13].

5) Learnability dimension

The dimension of learnability was the one with the highest rating ($M = 4.63$, $SD = 0.49$), which suggests that the users could learn the system even without any prior training. This result confirms the study which states that by utilizing intuitive AI-supported systems, the learning curve for teachers can be significantly shortened [7].

6) Duplication tracking feature

The assessment of the duplication monitoring function ($M = 4.53$, $SD = 0.52$) shows that users fully trust the AI capabilities of the system. This aligns with the results that similarity-based methods are effective for detecting overlapping or semantically similar items [8, 11].

7) Overall usability

The system's average score of 4.56 ($SD = 0.51$) puts it in the "Very Highly Usable" category, which not only verifies the usability standards but also aligns with previous studies. This degree of user-friendliness is a common finding in the evaluation of educational systems where fast and simple systems are always rated the highest [14].

E. Comparison with Existing Quiz Management Systems

Existing quiz management and item bank systems are commonly integrated within learning management systems such as Moodle and Canvas, which provide functionalities for storing questions, categorizing items, and assembling quizzes based on predefined criteria [26]. These platforms are widely adopted in higher education and are effective for routine assessment administration; however, duplication control in these systems largely depends on manual review or basic keyword-based searches. As question repositories grow over time, instructors often rely on personal judgment or memory to avoid reusing identical or highly similar items, which can be inefficient and error-prone [9].

In contrast, the proposed system explicitly integrates an AI-based duplication and similarity detection mechanism using cosine similarity to identify exact, near-duplicate, and semantically similar quiz items. This functionality is automatically triggered during quiz preparation, providing real-time feedback that supports instructor decision-making prior to quiz finalization. Similarity-based approaches have been shown to be effective in detecting overlapping assessment items even when questions are phrased differently [8, 11].

While Moodle and Canvas primarily emphasize content delivery and assessment deployment, the present system focuses on repository quality control and item diversity through intelligent similarity analysis. Moreover, many standalone quiz repositories and item bank tools reported in the literature prioritize storage and retrieval efficiency but do not integrate AI-driven similarity detection with systematic usability evaluation [2]. The proposed system differentiates itself by combining AI-assisted duplication tracking with a usability-centered design framework, empirically validated through Nielsen's usability heuristics and the Software Usability Measurement Inventory (SUMI).

V. Conclusion

The study presents the idea and evaluation of the Intelligent Quiz Repository and Generation System with AI-Based Duplication Tracking which helped mainly in resolving the long-standing problems of quiz creation, repository management, and quality assessment in the higher education area. The system was a single solution to all the difficulties like slow manual quiz making, redundancy or similarity in the questions of item banks, and the absence of complete tools that not only assisted in automated quiz generation but also monitored duplication.

Incorporating Developmental-Descriptive Research Design, the study developed a common and easily accessed platform made up of a quiz repository, an auto quiz generator, analytics dashboard, and an AI-based tracking system for duplicate detection. The trustworthiness of the cosine similarity algorithm in the discrimination among exact duplicates, near-duplicates, and semantically similar items was tested and it received an overall accuracy of 91.25%. This result implies that the AI-powered similarity detection might be of great assistance in item quality enhancement and at the same time keeping the quiz repository protected from any unauthorized access.

The usability tests constituted one more proof of the system's efficiency. The system was graded as "Very Highly Usable" according to the Nielsen Usability Heuristics evaluation, while it got the most points in the areas like visibility of system status, consistency of the interface, error prevention, and aesthetic design. The SUMI assessment reached a comparable finding as it also labeled the system "Very Highly Usable" regarding all five factors—efficiency, affect, helpfulness, control, and learnability. The rating of such a high level indicates that the users will have no difficulties with the system at all; instead they shall be very well supported and what is more, the system will take them a very short time to learn.

When compared with existing quiz management solutions, the system demonstrates that AI-assisted duplication detection, when embedded within a user-centered design, can enhance both assessment quality and instructor experience without increasing system complexity.

The system fulfilled every requirement of the special study purposes. It was found that the incorporation of AI plagiarism detection, well-structured database, and automatic quiz generation could lead to a fourfold increase in efficiency, accuracy and quality of quiz management. The system achieves this by elimination of question repetition, enlargement of item diversity and finally, lightening the instructor's workload. Thus, the system is a part of the process leading to more credible and fair assessment methods. Moreover, the very good usability results suggest that the platform could be accepted by various faculty members with differing levels of comfort with technology.

Conflict of Interest

The authors declare that there is no conflict of interest

regarding the conduct of this research and the publication of this article.

AUTHOR CONTRIBUTIONS

Menirissa D. De Belen (M.D.D.B.) led the conceptualization of the study, system design, and overall project coordination, and took the lead in writing the initial draft of the manuscript. Jayson A. Batoon (J.A.B.) implemented the intelligent quiz repository and generation system, developed the AI-based duplication tracking module, and assisted in data collection and results interpretation. Reylan M. Evale (R.M.E.) contributed to the usability evaluation process, assisted in data analysis and interpretation, and supported the refinement of the Results and Discussion sections, particularly in relation to usability and workflow integration. Digna S. Evale (D.S.E.) supervised the research design and evaluation procedures, validated the instruments, contributed to data analysis and interpretation, and provided critical revisions to improve the final manuscript. All authors reviewed and approved the final version of the paper

ACKNOWLEDGMENT

The authors wish to thank the College of Information and Communications Technology of Bulacan State University for its support in the conduct of this research. They are also grateful to the faculty members who participated in the evaluation of the system and to the administrators and staff who provided administrative and technical assistance throughout the project.

REFERENCES

[1] B. Das, M. Majumder, and A. Sekh, "Automatic question generation and answer assessment: A survey," *Research and Practice in Technology Enhanced Learning*, vol. 16, 5, 2021. doi: 10.1186/s41039-021-00151-1

[2] M. Chowdhary, V. Goyal, and V. Mohania, "Unsupervised question duplicate and related questions detection in e-learning platforms," arXiv Print, arXiv:2301.05150, 2022. https://arxiv.org/abs/2301.05150

[3] K. Mouri, "Intelligent assessment systems and automatic question generation: Current trends and educational implications," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 23, pp. 59–74, 2021. doi: 10.3991/ijet.v16i23.27849

[4] M. J. Gierl, J. Shin, T. Firoozi, and H. Lai, "Using content coding and automatic item generation to improve test security," *Frontiers in Education*, vol. 7, 853578, 2022. doi: 10.3389/feduc.2022.853578

[5] B. R. Rush, D. C. Rankin, and B. J. White, "The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value," *BMC Medical Education*, vol. 16, no. 1, 250, 2016. doi: 10.1186/s12909-016-0773-3

[6] R. Circi, J. Hicks, and E. Sikali, "Automatic item generation: Foundations and machine-learning-based approaches for assessments," *Frontiers in Education*, vol. 8, 858273, 2023. doi: 10.3389/feduc.2023.858273

[7] C. Ling *et al.*, "Enhancing educational assessment through AI-assisted question generation: A systematic review," *Computers & Education: Artificial Intelligence*, vol. 7, 100205, 2024. doi: 10.1016/j.caeai.2024.100205

[8] V. J. Albano, M. D. L. Higueras, and R. Smith, "NLP-based management of large multiple-choice test item banks via preprocessing, similarity computation and graph analysis," *Journal of Learning Analytics*, 2023.

[9] M. J. Gierl *et al.*, "Building and managing large-scale item banks for educational assessment: Practical and psychometric considerations," *Frontiers in Education*, vol. 7, 853578, 2022. doi: 10.3389/feduc.2022.853578

[10] Z. Sayın, "AI-driven automatic question generation and validation using cosine-similarity and SME review," *Turkish Online Journal of Distance Education*, vol. 24, no. 2, pp. 45–58, 2023. doi: 10.17718/tojde.1304785

[11] P. T. Manh, H. N. Nguyen, and D. T. Tran, "Duplicate multiple-choice question detection in educational repositories using multimodal deep learning," *Education and Information Technologies*, vol. 28, pp. 6549–6568, 2023. doi: 10.1007/s10639-023-11763-5

[12] International Organization for Standardization, ISO 9241-210:2019. Ergonomics of human-system interaction—Human-centred design for interactive systems. [Online]. Available: https://www.iso.org/standard/77520.html

[13] J. Nielsen, *Usability Engineering*, Morgan Kaufmann, 1994.

[14] J. Kirakowski, "The Software Usability Measurement Inventory (SUMI)," *British Journal of Educational Technology*, vol. 24, no. 3, pp. 210–212, 1993. doi: 10.1111/j.1467-8535.1993.tb00661.x

[15] S. McKenney and T. C. Reeves, "Educational design research: Portraying, conducting, and enhancing productive scholarship," *Medical Education*, vol. 55, no. 1, pp. 82–92, 2021. doi: 10.1111/medu.14280

[16] A. E. Artyukhov, I. I. Volk, and T. A. Vasylieva, "Agile methodology in higher education quality assurance system for SDGs 4, 8 and 9 achievement: National experience," *CTE Workshop Proceedings*, vol. 9, pp. 81–94, 2022. doi: 10.55056/cte.105

[17] P. Miksza, K. Elpus, and K. Tracey, "Quantitative descriptive and correlational research," in *Music Education Research: An Introduction*, 2nd ed. Oxford University Press, 2023, pp. 127–148.

[18] M. Sharma and R. Singh, "Modern web application development using React.js: A performance and scalability analysis," *International Journal of Research Publication and Reviews*, vol. 5, no. 3, pp. 123–130, 2024.

[19] R. S. Pressman and B. R. Maxim, *Software Engineering: A Practitioner's Approach*, 9th ed. McGraw Hill India, 2023. ISBN: 978-9355325044

[20] L. Ilomäki *et al.*, "Digital competence—An emergent boundary concept for policy and educational research," *Education and Information Technologies*, vol. 21, no. 3, pp. 655–679, 2016. doi: 10.1007/s10639-014-9345-4

[21] A. A. Adewale, S. Ghavifekr, and A. Z. A. Razak, "Validity and reliability of a questionnaire developed to explore quality assurance components for teaching and learning in vocational and technical education," *Humanities and Social Sciences Communications*, vol. 9, no. 1, 303, 2022. doi: 10.1057/s41599-022-01306-1

[22] R. F. McCloud *et al.*, "Evidence for decision-making: The importance of systematic data collection as an essential component of responsive feedback," *Global Health: Science and Practice*, vol. 11, no. 2, e2200246, 2023. doi: 10.9745/GHSP-D-22-00246

[23] B. Albert and T. Tullis, *Measuring the User Experience: Collecting, Analyzing, and Presenting UX Metrics*, 3rd ed. Morgan Kaufmann, 2022. ISBN: 978-0-12-818080-8

[24] American Educational Research Association, "Code of ethics," *Educational Researcher*, vol. 40, no. 3, pp. 145–156, 2011. doi: 10.3102/0013189X11410403

[25] G. A. Picón, "Perspective of the institutional digital repository in the protection and diffusion of academic activity," *Ciencia Latina Revista Científica Multidisciplinar*, May 6, 2024. doi: 10.37811/cl_rcm.v8i2.10831

[26] A. Oussous, F. Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "An evaluation of open-source adaptive learning solutions," *Information*, vol. 14, no. 2, 57, 2023. doi: 10.3390/info14020057