Kashihara, A. et al. (Eds.) (2024). Proceedings of the 32nd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Cognicraft: Smart Exam Question Generation with AI and Bloom's Taxonomy

Christian SAGADRACA*, Zainal SANTOS, Danilo SIMON Jr., Marianne Jessica TOLENTINO & Reymar VENTURA
Mariano Marcos State University, Philippines
*casagadraca@mmsu.edu.phfa

Abstract: Conventional approaches of exam question creation are often characterized by time-consuming manual methods and challenges in aligning questions with specific educational objectives. This study addresses these challenges by developing Cognicraft, an innovative AI-powered exam question generation tool designed to streamline and improve the exam composition process. Cognicraft uses the Feature Driven Development-Agile Methodology and Bloom's Taxonomy to automatically generate exam questions that are academically acceptable. Iterative design cycles were used in the development process to identify, implement, and revise important features in response to continuous user feedback. The methodology included gathering requirements from educators, designing and implementing system features, and conducting User Acceptability Testing (UAT) to evaluate performance across four key dimensions: functionality, reliability, efficiency, and acceptability. The results from the UAT, involving 50 participants, demonstrated high ratings across all criteria. Cognicraft also achieved strong scores in generating relevant and varied exam questions, maintaining consistency and accuracy, and providing a user-friendly experience. The findings indicate that Cognicraft successfully addresses the limitations of traditional exam creation methods by significantly reducing the time and effort required for question generation while ensuring alignment with educational objectives. The system's ability to generate questions that adhere to Bloom's Taxonomy enhances the quality and diversity of assessments, offering a valuable tool for educators seeking to improve their exam creation processes. Cognicraft represents a significant advancement in educational technology, providing a practical and efficient solution to longstanding challenges in exam question development.

Keywords: AI in education, automated exam generation, Bloom's Taxonomy, cognitive assessment, educational technology

1. Introduction

Conventional methods of exam question creation present significant challenges for educators, including substantial time investment and difficulties in aligning questions with specific educational objectives (Das, Majumder, Phadikar, & Sekh, 2021). Conventional approaches often involve a laborious process that is prone to inconsistencies and inefficiencies. The present educational standard demands a more effective and efficient mode of assessment, ensuring exams could accurately reflect educational goals and fairly evaluate student learning outcomes (Shabatura, 2022).

Bloom's Taxonomy from the 1950s, offers a structured framework for categorizing educational goals and learning outcomes into hierarchical levels, ranging from basic knowledge recall to complex evaluation and synthesis (Shabatura, 2022). This taxonomy is widely used to design assessments that align with instructional objectives, but its effective application requires considerable expertise and effort, which can be challenging for educators with heavy workloads and tight deadlines (Das et al., 2021).

Recent advancements in artificial intelligence (AI) present promising solutions to these challenges. AI technologies can automate the generation of exam questions, thereby reducing the time and effort needed (Aldoseri, Al-Khalifa, & Hamouda, 2023). By leveraging data and

algorithms, AI-driven systems can create questions that align with Bloom's Taxonomy and offer diverse types and formats, addressing the need for comprehensive and varied assessments (Yan et al., 2023).

This study introduces Cognicraft, an innovative AI-powered system designed to enhance the exam question creation process. Cognicraft integrates AI with Bloom's Taxonomy to automate and optimize the generation of exam questions. The primary objective of this research is to evaluate Cognicraft's effectiveness in overcoming the challenges associated with traditional exam creation methods. By assessing the system's performance in terms of functionality, reliability, efficiency, and acceptability, this study aims to determine whether Cognicraft provides a viable solution to the issues faced by educators and contributes to the quality and efficiency of exam question generation.

2. **Research Method**

2.1 Research Design and Data-Gathering Procedure

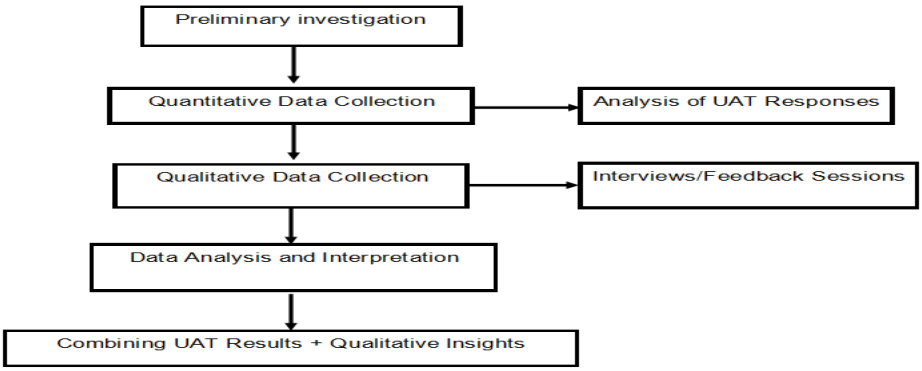

Figure 1. Flow of the Data-Gathering Process

Flowchart in Figure 1 represents the steps of this process, from the preliminary investigation to the combined interpretation, providing a holistic view of how users interacted with the system including both the quantitative and qualitative aspects of user experiences with CogniCraft. A preliminary investigation was conducted to gain an in-depth understanding of the specific needs and preferences of educators, also by reviewing relevant literature to understand the key challenges in traditional exam creation methods. Following this investigation a User Acceptance Form was developed to systematically capture feedback and assess the educator's readiness for the proposed changes. The target participants were identified consisting of faculty members from the College of Computing and Information Sciences. The researchers then sought ethical clearance from the University Research Ethics Review Board (URERB) to ensure that all ethical considerations were addressed. Upon receiving the Exempt Research Certificate from URERB the User Acceptance Test (UAT) Form was distributed to the selected respondents with clear instructions provided along with assurances of data privacy through confidentiality and anonymity. The data collected from the UAT Forms were analyzed and interpreted to evaluate the overall functionality and effectiveness of the system. Qualitative data was gathered through interviews or feedback sessions to capture user experiences, perceived challenges, and suggested improvements. The final step combined the quantitative UAT findings with qualitative feedback, ensuring that both numerical data and narrative insights were considered in evaluating the system's overall performance and user satisfaction.

2.3 Population and Sampling of the Study

The study's population is comprised of faculty members from the College of Computing and Information Sciences (CCIS) - MMSU who were the primary data source for both the

preliminary survey and User Acceptance Testing (UAT) phases. Participants were deliberately chosen based on their direct involvement in the educational process and their familiarity with contemporary teaching methods and technological tools.

2.4 Statistical Treatment

The researchers utilized weighted mean as the primary statistical analysis technique to evaluate the dataset collected during the User Acceptance Test (UAT). This method allowed for the calculation of average values while considering the relative importance or frequency of each data item, making it particularly advantageous when certain observations carried more weight or significance than others. In addition to weighted mean computations, data analysis also included frequency and percentage distributions. Frequency distribution was employed to illustrate the number of occurrences within each category of a variable, while percentage calculations provided a normalized representation of the raw data, facilitating clearer insights. The Likert Scale of Descriptive Equivalent, used to interpret the results. This scale was essential in categorizing user feedback into distinct levels of satisfaction and performance. The quantitative scale ranged from 1 to 5, with statistical limits assigned to each range. A score of 4.20 to 5.00 indicated "Excellence," 3.40 to 4.19 signified "Good," 2.60 to 3.39 was labeled as "Fair," 1.80 to 2.59 as "Poor," and 1.00 to 1.79 represented "Needs Improvement." This structured approach ensured that user responses were systematically analyzed and categorized for more effective interpretation and decision-making.

2.5 System Development Methodology

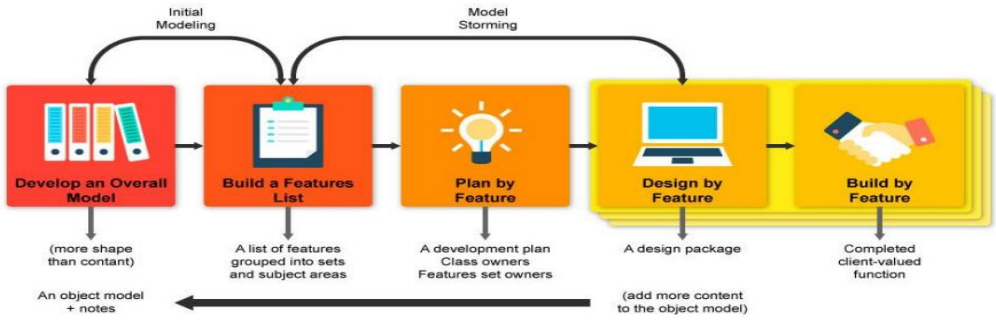

Figure 2. Feature Driven Development-Agile Methodology

The researchers adopted Feature-Driven Development (FDD) as shown in Figure 2 to guide the development of the system application. This structured approach ensured a user-centric focus and facilitated the efficient delivery of functionalities throughout the development process. Each phase of FDD contributed significantly to the application's development. In the initial phase, researchers defined the core purpose and key functionalities of the application by conducting interviews to gather user requirements, understand workflows, and identify pain points. This led to the development of an overall model that guided the alignment of the application's features with the thesis objectives. A comprehensive list of user stories was created, prioritizing features based on their importance and contribution to the objectives. Critical functionalities, such as automatic question generation based on Bloom’s Taxonomy using the Unofficial Hugging Chat API, database integration for storing questions, and a user-friendly interface, were developed first to gather early feedback. During the planning phase, detailed user stories with acceptance criteria and estimated development efforts were outlined to ensure clear communication among developers, the thesis adviser, panel members, and users. In the design phase, UI mockups, data structures, and algorithms were created to ensure a cohesive user experience and reduce rework. Finally, the features were developed iteratively, undergoing testing and deployment, allowing for early bug detection and continuous improvement based on real-world use.

3. Results and Discussion

3.1 Review on the Challenges of Traditional Examination Creation

Educators have faced numerous challenges in creating exams, including concerns related to security, difficulties in printing and distributing physical copies, and the need to accommodate diverse learning styles and preferences. To address these challenges, various strategies have been recommended, such as leveraging technology for managing question banks, implementing secure online question paper delivery systems, and ensuring that instructions provided to students are clear and unambiguous. Furthermore, aligning exams with course objectives, writing questions with clarity, and considering the time required for students to complete the exam have been identified as critical factors in enhancing the examination process's overall effectiveness. Users have also been advised to clarify exam formats and scoring criteria while providing guidance on effective study techniques to help students prepare more efficiently. Emphasizing higher-order thinking skills, engaging students in active learning, and offering study strategies like creating crib sheets or portfolios have been suggested as methods to help students synthesize course materials and perform better in exams.

3.2 Cognicraft's Response to Addressing Traditional Exam Creation Challenges

Cognicraft effectively addresses the key challenges identified in traditional exam creation through its AI-powered exam generation system. By automating the process of question creation, the system alleviates the time-consuming task of manually crafting exams, allowing educators to generate questions that are aligned with Bloom's Taxonomy and course objectives. This ensures that the questions are not only consistent in quality but also cater to diverse cognitive levels, promoting a more comprehensive assessment of students' knowledge and skills. Cognicraft's digital approach also eliminates the logistical difficulties related to printing and distributing physical copies of exams, offering a secure online platform for managing question banks and delivering assessments. Additionally, the system allows for the customization of exam formats, making it easier for educators to align the exams with specific learning outcomes and course requirements. By providing varied question types and difficulty levels, Cognicraft ensures that assessments are clear, structured, and tailored to student needs, addressing the issue of diverse learning preferences. Its ability to quickly generate exams enables educators to focus more on guiding students with effective study techniques and engaging them in higher-order thinking tasks, thus improving both exam quality and the overall learning experience.

3.3 User Acceptance Testing

The testing procedure for the CogniCraft system followed its presentation and demonstration phases. A User Acceptance Test (UAT) was conducted to evaluate the system's functionality, efficiency, reliability, and acceptability, categorizing these aspects as excellent, good, fair, poor, or in need of improvement. The researchers employed the ISO 25010:2011 standard to ensure a methodical evaluation of software quality.

The system's functionality was assessed with a composite mean score of 4.29, indicating "Excellent" performance across all evaluated criteria. Reflecting the system's ability to effectively generate exams with varying difficulty levels,as well as question types.

In terms of reliability, the system achieved a 4.5, also interpreted as "Excellent." Indicating the system's accuracy in generating exams aligned with Bloom's Taxonomy levels and its robustness in avoiding bugs and errors, showing exceptional performance, with reliable saving and retrieval of exam data, and lesser to no crashes or unexpected errors.

Regarding efficiency, the system earned a composite mean score of 4.5, classified as "Excellent." This result highlighted the system's ability to facilitate quick exam creation

compared to traditional methods, deliver prompt responses even for complex question structures, and provide an intuitive user interface that minimized learning time.

The system's acceptability was evaluated with a composite mean score of 4.275, reflecting "Excellent" usability and user experience. Indicating its user-friendliness, easy to navigate, and consistently dependable for exam creation tasks.

The grand mean score across all criteria was 4.39, considered as "Excellent", affirmed the system's high performance, reliability, efficiency, and user satisfaction, showcasing its effectiveness in meeting the needs of its users and enhancing the exam creation process.

3.4 System Result

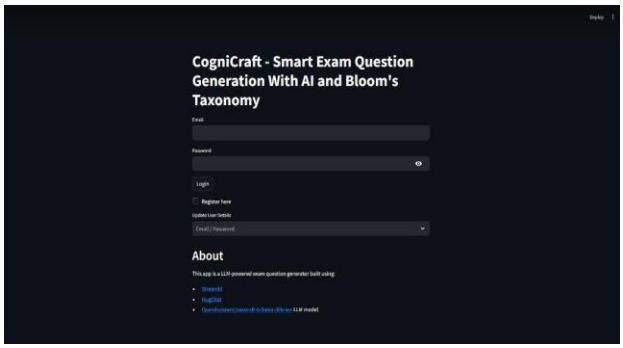

Figure 5. Cognicraft's User Login

The CogniCraft Smart Exam Generator's login page, depicted in Figure 5, functioned as the main entry point for users. This page utilized standard username and password authentication. Future enhancements are planned to address design improvements and common issues identified during testing. The system employed Streamlit for front-end development, and usability tests were conducted with 20 participants from CCIS.

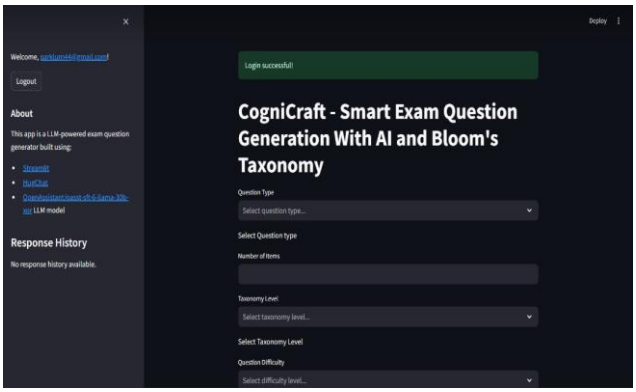

Figure 6. Cognicraft Exam Creation Interface

Figure 6 illustrates the exam creation interface of CogniCraft, an innovative online platform aimed at improving academic result management for educational institutions. The system was designed to offer teachers efficient access to academic results while overcoming the limitations of traditional exam creation methods. Users could select various parameters, including question type, taxonomy level, and difficulty. The system's user acceptability testing received positive feedback, though users recommended further enhancements.

The comparative analysis of various exam generation systems, including CogniCraft. It reveals that CogniCraft stands out for its incorporation of Bloom's Taxonomy, a feature that lacks in other systems. While all systems except ClassMarker support multiple question types, CogniCraft, along with ClassMarker, Testportal, and Quizgecko, offers options for adjusting question difficulty. Additionally, systems like ClassMarker, Testportal, Quizgecko, QuizGenerator, and CogniCraft support contextual entry, which is crucial for creating detailed assessments. Quizalize and Kahoot lack this feature, possibly due to their focus on interactive,

gamified learning experiences. CogniCraft's unique integration of Bloom’s Taxonomy with comprehensive features positions it as a valuable tool for educational institutions seeking a structured approach to assessment.

4. Scope and Limitations of the Study

The creation of Cognicraft was done using the Feature Driven Development-Agile Methodology and its integration with Bloom's Taxonomy to automate the generation of exam questions. The study assesses the system's effectiveness based on four key dimensions: functionality, reliability, efficiency, and acceptability. A User Acceptance Testing (UAT) was conducted with 50 faculty members from various educational institutions to gather feedback and evaluate the system's performance in real-world educational settings. The analysis includes quantitative and qualitative assessments of the system's ability to generate relevant, varied, and high-quality exam questions aligned with educational objectives.

5. Conclusion

The study led to several key conclusions. First, CogniCraft effectively addressed the challenges of traditional exam creation methods, providing a viable solution to issues such as time consumption and difficulty in aligning questions with educational objectives. Second, the analysis identified that while traditional exam creation methods had room for improvement, they were not insurmountable obstacles. The system’s AI-driven approach and incorporation of Bloom’s Taxonomy streamlined exam creation, offered a broader range of question types, and enhanced assessment fairness. Third, the deployment of CogniCraft proved to be a practical solution, successfully simplifying exam creation and improving assessment equity. Fourth, the evaluation results were highly favorable, reflecting the system’s robust functionality, consistent reliability, efficient performance, and high user satisfaction. Lastly, CogniCraft was established as a valuable tool for educators seeking to enhance the exam creation process, with ongoing improvements facilitated by the Feature-Driven Development-Agile Methodology. This study underscores the transformative potential of technology in advancing educational assessment and fostering more effective learning outcomes.

References

Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-thinking data strategy and integration for artificial intelligence: Concepts, opportunities, and challenges. *Applied Sciences, 13*(12), 7082. https://doi.org/10.3390/app13127082

Das, B., Majumder, M., Phadikar, S., & Sekh, A. A. (2021). Automatic question generation and answer assessment: A survey. *Research and Practice in Technology Enhanced Learning, 16*(1). https://doi.org/10.1186/s41039-021-00151-1

Shabatura, J. (2022) “Using Bloom’s taxonomy to write effective learning outcomes”. *Teaching Innovation and Pedagogical Support*. Retrieved from https://tips.uark.edu/using-blooms-taxonomy/

Yan, L., et al. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology, 55*(1), 90–112. https://doi.org/10.1111/bjet.13370