Question Generation Using Artificial Intelligence and Knowledge Augmented Generation

[#1] Dr. G. RAMESH, *PROFESSOR,*
[#2] P. RAMYA, *ASSISTANT PROFESSOR,*
[#3] T.R.V. VISVESH KHANNA, *B. Tech Students,*
[#1-3] Department of Information Technology
K.L.N. COLLEGE OF ENGINEERING (AUTONOMOUS), POTTAPALAYAM, SIVAGANGAI DISTRICT, TAMILNADU, INDIA.

---***---

Abstract - This project introduces an innovative, AI-driven platform tailored to streamline and enhance the process of question paper generation for educators at various academic levels. Utilizing cutting-edge technologies such as Natural Language Processing (NLP), Knowledge-Augmented Generation (KAG), and Large Language Models (LLMs), the system is capable of understanding and extracting key concepts from a variety of educational resources. These include user-uploaded materials such as textbooks, existing question banks, lecture notes, and previous year exam papers.
Educators are provided with a user-friendly interface where they can define specific parameters, including subject area, topic coverage, question complexity (difficulty level), and the desired number of questions. Based on these inputs, the platform intelligently generates a diverse and balanced set of questions, ensuring coverage across cognitive levels—such as knowledge recall, application, and analysis. Additionally, each question is accompanied by a corresponding answer key, which helps educators expedite the evaluation process.

Keywords - Automated Question Paper Generation, NLP-Based Assessment Design Tool, AI-Powered Educational Assessment System.

1.INTRODUCTION:

Assessment plays a vital role in the educational training process by offering critical insights into the effectiveness of language instruction. It helps identify difficult areas for learners and measure their progress in language courses (Pearson & Murphy-Judy, 2020; Voss, 2018). Through assessment, lecturers can evaluate students' language proficiency, recognize strengths and weaknesses, and suggest targeted improvements (Brown & Abeywickrama, 2010; Purpura, 2016). Common assessment tools—such as multiple-choice questions, reading comprehension exercises, and oral language evaluations (Hughes & Hughes, 2020)—require substantial time and effort to design, as they must reflect learners' current levels and educational goals. This time investment is crucial to ensure accuracy and effectiveness in language proficiency assessment at the university level.

Despite the acknowledged importance of language assessments, lecturers often face challenges when designing test items that align with institutional or national curriculum standards. The time and dedication required for this process are essential for promoting student success and maintaining high educational quality. However, technological advancements—particularly in artificial intelligence (AI)—have begun to reshape the educational landscape by simplifying time-consuming tasks (Çakmak, 2019; Kukulska-Hulme & Morgana, 2021; Selwyn et al., 2021). In the field of language education, technology has significantly improved both teaching and learning by enabling the automated and adaptive creation of assessments (Nazaretsky et al., 2022). While AI offers promising benefits for improving the efficiency and adaptability of assessments (García-Peñalvo et al., 2020; Gardner et al., 2021), there is limited research on the potential risks, hidden errors, and limitations of AI-generated language tests (Van Moere & Downey, 2016).

This study aims to address that gap by exploring the specific challenges and opportunities encountered by language institute lecturers when using AI to generate test questions and exam banks for English Language Teaching (ELT). A major contribution of this research is its exploration of AI's dual function—not only as a tool for automating test creation but also as a collaborator that can enhance test content, minimize human bias, and tailor assessments to accommodate diverse learner needs. The study offers new empirical insights into the practical benefits and potential drawbacks of using AI in ELT assessment, presenting a balanced view of its integration into the educational process.

2.LITERATURE REVIEW:

[1] Ayan Kumar Bhowmick, Ashish Jagmohan, Aditya Vempaty, Prasenjit Dey, Leigh Hall, Jeremy Hartman, Ravi Kokku, Hema Maheshwari, "Automating Question Generation from Educational Text," arXiv preprint arXiv:2309.15004, September 2024.

This study presents a modular framework employing transformer-based language models for automatic generation of multiple-choice questions (MCQs) from educational texts.

It includes distinct modules for question generation, correct answer prediction, and distractor formulation, enabling evaluation of different language models and generation techniques. The framework aims to reduce teachers' workload and facilitate personalized learning experiences.

[2] Jiayi Wang, Ruiwei Xiao, Ying-Jui Tseng, "Generating AI Literacy MCQs: A Multi-Agent LLM Approach," arXiv preprint arXiv:2412.00970, December 2024.

This research introduces a novel approach utilizing large language models (LLMs) to automatically generate scalable, high-quality multiple-choice questions (MCQs) for AI literacy assessments. The method aligns with user-provided learning objectives, grade levels, and Bloom's Taxonomy levels, incorporating an iterative workflow with LLM-powered critique agents to ensure pedagogical standards. Preliminary evaluations indicate strong interest from experts, suggesting potential to enrich AI literacy materials for K-12 educators.

[3] Subhankar Maity, Aniket Deroy, "Generative AI and Its Impact on Personalized Intelligent Tutoring Systems," arXiv preprint arXiv:2410.10650, October 2024.

This report explores the integration of generative AI, particularly large language models like GPT-4, into Intelligent Tutoring Systems (ITS) to enhance personalized education.

It discusses applications such as automated question generation, customized feedback mechanisms, and interactive dialogue systems tailored to individual learner needs. The study also addresses challenges including ensuring pedagogical accuracy, mitigating biases, and maintaining learner engagement, highlighting the transformative potential of generative AI in creating effective and engaging educational experiences.

[4] Tianshi Zheng, Weihan Li, Jiaxin Bai, Weiqi Wang, Yangqiu Song, "Assessing the Robustness of Retrieval-Augmented Generation Systems in K-12 Educational Question Answering with Knowledge Discrepancies," arXiv preprint arXiv:2412.08985, December 2024.

This study investigates the robustness of Retrieval-Augmented Generation (RAG) systems in K-12 educational question answering, particularly under knowledge discrepancies between authoritative textbooks and the parametric knowledge in Large Language Models (LLMs). The authors present EduKDQA, a dataset simulating knowledge discrepancies, and conduct extensive experiments revealing that most RAG systems experience substantial performance drops in such scenarios.

3.EXISTING SYSTEM:

In the current educational landscape, the process of creating question papers remains largely manual or, at best, semi-automated. Educators are responsible for curating questions by referencing physical or digital textbooks, pre-existing question banks, and previous examination papers. This manual approach is not only labor-intensive but also repetitive and susceptible to human errors such as duplications, inconsistencies in difficulty levels, or misalignment with the syllabus and learning objectives. Some institutions have attempted to modernize this process by implementing digital question bank management systems. However, these tools typically offer limited functionality. They often operate as static repositories where questions are pre-fed by users and then categorized based on basic parameters such as subject, topic, and difficulty level. While these systems may slightly improve efficiency by enabling search and filtering, they still require educators to manually select and compile the final set of questions.

Furthermore, these systems lack the ability to dynamically generate new questions based on evolving academic standards, recent curriculum changes, or personalized student learning outcomes. They do not leverage advanced technologies such as artificial intelligence or machine learning to understand context, generate diverse question types, or ensure coverage across Bloom's Taxonomy (e.g., remembering, understanding, applying, analysing, evaluating, and creating).

Additionally, the limited adaptability of these tools means they are often confined to specific subjects or educational levels, reducing their utility across a broad range of academic disciplines. The result is a workflow that continues to demand substantial time and effort from educators, especially during peak academic periods such as midterms and final examinations.

In summary, the existing systems fall short in offering intelligent automation, real-time adaptability, and contextual relevance, thereby highlighting the need for a more robust, AI-powered solution that can revolutionize the way assessments are designed and administered in modern education.

4.PROPOSED SYSTEM:

To address the limitations of manual and semi-automated question paper generation, this project proposes an AI-powered, intelligent question generation platform that leverages cutting-edge technologies such as Natural Language Processing (NLP), Knowledge-Augmented Generation (KAG), and Large Language Models (LLMs). The goal is to automate the creation of question papers in a way that is efficient, scalable, pedagogically sound, and contextually relevant to the curriculum.

Key Features and Functional Overview:

1. Content Ingestion and Understanding: The system allows educators to upload a wide range of academic resources, including textbooks, curriculum guides, lecture notes, previous year question papers, and question banks. These documents are processed using NLP techniques to extract meaningful information such as key concepts, terminologies, learning objectives, and topic hierarchies.

2. Smart Question Generation: Leveraging Large Language Models (like GPT or similar), the platform is capable of dynamically generating a diverse set of questions—including multiple-choice, short-answer, long-answer, fill-in-the-blank, and application-based items. These questions are generated with an understanding of:

   - Curriculum alignment
   - Subject relevance
   - Bloom's Taxonomy levels (e.g., Remembering, Understanding, Applying, Analysing, Evaluating, Creating)
   - Grammatical and contextual correctness

3. Customizable Parameters: Educators can define parameters such as:

   - Subject and Topic
   - Difficulty level (easy, medium, hard)
   - Number of questions
   - Question type
   - Mark distribution and time constraints

Based on these inputs, the system tailors the question paper output accordingly.

4. Quality Control and Redundancy Check: The AI engine includes built-in mechanisms to prevent repetition of questions, content overlap, or inclusion of outdated or irrelevant material. It evaluates the coherence, clarity, and uniqueness of each question before finalizing the paper.

5. Scalability and Multi-Subject Support: The platform is designed to support a wide range of academic disciplines, from languages and humanities to science, technology, engineering, and mathematics (STEM). This makes it suitable for use across schools, colleges, universities, and coaching institutions.

6. User Feedback Loop: Educators can provide feedback on the generated questions, helping the system learn and improve over time through reinforcement learning techniques. This ensures continuous enhancement of quality and contextual relevance.

5. ADVANTAGE-

1. Dramatically reduces time and manual effort Automates the entire question paper creation process, saving hours of manual work.

2. Improves question diversity and depth Generates a wide variety of question types that target multiple levels of understanding.

3. Ensures curriculum alignment and appropriate difficulty balance Tailors questions based on syllabus and difficulty parameters set by educators.

4. Minimizes human error, duplication, and bias Uses AI algorithms to detect and eliminate redundant, inaccurate, or biased content.

5. Provides flexibility to educators while maintaining academic rigor offers customizable settings without compromising on educational standards and quality.

6. METHODOLOGY:

a. Study Design- This research employed a qualitative case study approach, utilizing semi-structured interviews to gather in-depth, open-ended responses. The primary aim was to explore lecturers' experiences, perspectives, and attitudes toward using AI-driven assessment tools in exam development. This method enabled the capture of detailed and often unexpected insights, focusing on individual emotions, experiences, and perceptions to analyze complex social dynamics. By collecting data from multiple sources, the case study approach provided a comprehensive understanding of individuals, groups, and specific events. The researcher's involvement in collaborative group activities and meticulous analysis of collected data played a critical role in shaping the study's design.

b.Participants-

The study involved seven lecturer groups from the Institute of Language at Van Lang University, located at Campus 3 (Main Campus). These groups were responsible for teaching English courses ranging from Level 1 to Level 7, including specific classes (courses 27, 28, and 29) during the fall semester of the 2023–2024 academic year. Participants were selected using purposive sampling to ensure alignment with the study's objectives.

Strict inclusion criteria were applied to ensure the validity and reliability of the findings. Participants were required to hold at least a master's degree, have over five years of teaching experience, and serve as active members of the English Language Teaching (ELT) examination board. These requirements were grounded in the assumption that experienced educators provide deeper insights into contemporary teaching and assessment practices (Smith & Fletcher, 2020; Johnson & Johnson, 2020).

Purposive sampling was deemed appropriate due to its ability to identify participants with specialized knowledge directly relevant to the study's focus (Palinkas et al., 2015). This was especially important for analys ing AI-supported assessment practices and comparing them with conventional and alternative evaluation methods. Participant anonymity was preserved in accordance with ethical research standards (Bryman, 2016), ensuring data privacy throughout the study.

The research design adhered to three guiding principles:

- Connectivity: Survey and interview questions were designed in alignment with the study's goals, specifically targeting variables related to ELT assessment practices.
- Generalizability: The findings aimed to reflect broader demographic trends by capturing a range of perspectives.
- Feasibility: The study was conducted within the limitations of available resources and data access, ensuring a realistic and practical research scope.

c. Data Collection and Analysis- The data collection process followed a multi-phase approach to gain rich insights into the use of AI in ELT assessment at Van Lang University. Semi-structured interviews were conducted with lecturers, each lasting between 45 and 80 minutes. Most interviews were audio recorded, transcribed, and subsequently reviewed for accuracy. A thematic summary of each interview was developed, categorized by key themes and sub-themes using the frameworks of Gioia et al. (2013) and Jaskiewicz et al. (2015). To validate the findings, each participant received their interview summary and a copy of the consent form for review and confirmation (Appendix 1).

Open-ended interview questions were carefully crafted to explore AI's influence on test design, grading, and feedback mechanisms. The iterative analysis process consisted of the following stages (Braun & Clarke, 2006; Saldana, 2015):

1. Familiarization: Researchers engaged in repeated reading of the transcripts to gain an in-depth understanding of key themes and emerging patterns.
2. Initial Coding: Relevant segments of text were coded based on significance to the research questions. Tags such as "AI efficiency," "bias reduction," and "content quality" were used to categorize relevant data.

3. Theme Development: Related codes were grouped into overarching themes. For example, codes linked to “AI efficiency” and “time-saving” were merged to highlight AI's effectiveness in enhancing productivity.

4. Review and Refinement: Themes were continuously reviewed to ensure they accurately reflected the data. Adjustments were made as necessary to capture nuanced perspectives.

5. Theme Definition: Final themes and sub-themes were clearly defined and labeled. For instance, under the theme “AI Efficiency,” sub-themes like “automated grading” and “rapid test generation” were identified to illustrate specific benefits.

6. Interpretation: Themes were synthesized into a narrative report, connecting findings to the research objectives. Direct quotes from participants were included to illustrate key insights and ensure authenticity.

To strengthen the validity of the results, data triangulation was employed by analysing both traditional and AI-generated test materials (Noble & Heale, 2019). This approach allowed for a broader understanding of AI’s impact on ELT assessment practices. Member checking, where participants verified their interview summaries, further enhanced the credibility of the findings (Lincoln & Guba, 1988).

Through a transparent and systematic analytical process, this study uncovered meaningful insights into the role of AI in language assessments, highlighting both its advantages and limitations within the educational context.

7.RESULT AND DISCUSSION:

Fig:7.1 Dashboard

This is the main dashboard of the application, where users are welcomed after login. It provides a clean and organized overview of the platform, showing quick-access modules

Fig 7.2 Dashboard (Contd..)

In fig 7.2 This is an extended view of the dashboard showing options for profile management, subject selection, and shortcut widgets for quick task initiation. An extended dashboard view provides role-based controls for educators, allowing subject-specific settings, quick links to recently used documents, and customization tools for a more personalized workflow

Fig 7.3 Resource Library

In fig 7.3 The Resource Library allows users to upload and manage educational content, which is then processed by AI. Files can be categorized, tagged, and previewed to ensure structured and accessible storage before question generation begins. This screen displays the uploaded academic materials, organized by file type, subject, or year. Materials include PDFs, DOCX, and TXT files like textbooks, question banks, and previous year papers.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Fig 7.4 Project Builder

In fig 7.4 The Paper Builder module empowers teachers to assemble question papers with precision. It supports drag-and-drop organization, section-wise formatting, and total mark calculation—enabling full customization according to exam requirements. This module displays a visual interface where educators can drag and drop questions to form a custom question paper. It also offers options for adding sections, setting marks, and previewing the layout.

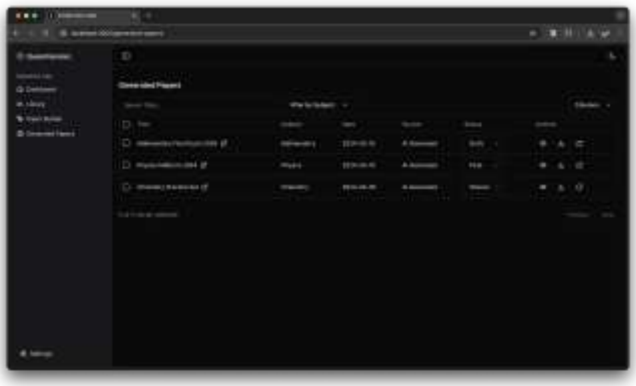

Fig 7.5 Genarated paper

In fig 7.5 Here, educators can view all previously created question papers. The interface enables seamless access to past papers, supports editing, and allows quick reusability—saving time and effort for repeated assessments.

Fig 7.6 Question Paper Gen

In fig 7.6 This is a sample output of the system's question generation engine. Questions are well-structured, diverse, and aligned with selected parameters like difficulty and topic. Answer keys are auto-generated to assist in quick evaluation.

8.CONCLUSION:

This proposed system represents a transformative shift in how academic assessments are prepared. By intelligently automating the question paper creation process, the platform empowers educators to focus more on teaching and mentoring, rather than administrative and repetitive tasks. With its AI-driven backbone, the system offers not only efficiency and accuracy but also personalization, scalability, and quality assurance—all essential in today's fast-evolving educational environment. The proposed system addresses the limitations of traditional question paper generation by introducing an AI-powered platform that is efficient, scalable, and intelligent. By leveraging advanced techniques such as Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs), the platform enables educators to automatically generate high-quality, topic-aligned questions and answer keys with minimal effort. This not only saves valuable time for teachers but also ensures consistency, diversity, and academic relevance in assessments.

9.REFERENCES:

[1] Ayan Kumar Bhowmick, Ashish Jagmohan, Aditya Vempaty, Prasenjit Dey, Leigh Hall, Jeremy Hartman, Ravi Kokku, Hema Maheshwari, "Automating Question Generation from Educational Text," arXiv preprint arXiv:2309.15004, September 2023.

[2] Jiayi Wang, Ruiwei Xiao, Ying-Jui Tseng, "Generating AI Literacy MCQs: A Multi-Agent LLM Approach," arXiv preprint arXiv:2412.00970, December 2024.

[3] Subhankar Maity, Aniket Deroy, "Generative AI and Its Impact on Personalized Intelligent Tutoring Systems," arXiv preprint arXiv:2410.10650, October 2024.

[4] Eaton, S.E., Mindzak, M. and Morrison, R. (2021), *Artificial Intelligence, Algorithmic Writing & Educational Ethics*, Canadian Society for the Study of Education. doi: 10.11575/PRISM/38967.

[5] Garcia, M.P., Smith, J., Johnson, L. and Taylor, R. (2023), "Personalized adaptive testing: a framework for enhancing language assessment through artificial intelligence", *Language, Learning and Technology*, Vol. 27 No. 1, pp. 45-62.

[6] Garc□ıa-Pe~nalvo, F.J., Corell, A., Abella-Garc□ıa, V. and Grande-de-Prado, M. (2020), "Recommendations for mandatory online assessment in higher education during the COVID-19 pandemic", in *In Radical Solutions for Education in a Crisis Context: COVID-19 as an Opportunity for Global Learning*, Springer Singapore, Singapore, pp. 85-98, doi: 10.1007/978- 981-15-7869-4_6.

[7] Gardner, J., O'Leary, M. and Yuan, L. (2021), "Artificial intelligence in educational assessment:'Breakthrough? Or buncombe and ballyhoo?", *Journal of Computer Assisted Learning*, Vol. 37 No. 5, pp. 1207-1216, doi: 10.1111/jcal.12577.

[8] Gioia, D.A., Corley, K.G. and Hamilton, A.L. (2013), "Seeking qualitative rigor in inductive research: notes on the Gioia methodology", *Organizational Research Methods*, Vol. 16 No. 1, pp. 15-31, doi: 10.1177/1094428112452151.

[9] Godwin-Jones, R. (2022), "Partnering with AI: intelligent writing assistance and instructed language learning", *Language, Learning and Technology*, Vol. 26 No. 2, pp. 5-24.

[10] Heaton, J.B. (1988), *Writing English Language Tests*, Longman, New York.

[11] Helm, J.M., Swiergosz, A.M., Haeberle, H.S., Karnuta, J.M., Schaffer, J.L., Krebs, V.E. and Ramkumar, P.N. (2020), "Machine learning and artificial intelligence: definitions, applications, and future directions", *Current reviews in musculoskeletal medicine*, Vol. 13 No. 1, pp. 69-76, doi: 10.1007/s12178-020-09600-8.

[11] Hinkelman, D. (2018), *Blending Technologies in Second Language Classrooms*, Springer, London.

[12] Hughes, A. and Hughes, J. (2020), *Testing for Language Teachers*, 3rd ed., Cambridge University Press, Cambridge.

[13] Jaskiewicz, P., Combs, J.G. and Rau, S.B. (2015), "Entrepreneurial legacy: toward a theory of how some family firms nurture transgenerational entrepreneurship", *Journal of Business Venturing*, Vol. 30 No. 1, pp. 29-49, doi: 10.1016/j.jbusvent.2014.07.001.

[14] Ji, H., Han, I. and Ko, Y. (2023), "A systematic review of conversational AI in language education: focusing on the collaboration with human teachers", *Journal of Research on Technology in Education*, Vol. 55 No. 1, pp. 48-63, doi: 10.1080/15391523.2022.2142873.

[15] Johnson, A., Smith, B. and Lee, C. (2019), "Leveraging AI for automated test generation in higher education", *Journal of Educational Technology*, Vol. 45 No. 2, pp. 213-228.

[16] Johnson, R.L., Allen, T. and Green, M. (2021), "The role of expertise in AI-assisted assessment: implications for higher education", *Computers and Education*, Vol. 173, 104268.

[17] Johnson, D. and Johnson, R. (2020), "What Is cooperative learning?—Cooperative Learning Institute", available at: http://www.co-operation.org/what-is-cooperative-learning (accessed 22 December 2024).

[18] Johnson, K.L. and Lee, H.S. (2022), "Natural language processing for automated item generation in language assessment", *Journal of Educational Technology Research*, Vol. 40 No. 2, pp. 201-218.

[20] Killawala, A., Khokhlov, I. and Reznik, L. (2018), "Computational intelligence framework for automatic quiz question generation", *2018 IEEE International*

Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, pp. 1-8, doi: 10.1109/FUZZ-IEEE.2018.8491624.

[21] Kukulska-Hulme, A. and Morgana, V. (2021), *Mobile-Assisted Language Learning across Educational Contexts*, Routledge, New York.

[22] Lam, R. (2015), "Language assessment training in Hong Kong: implications for language assessment literacy", *Language Testing*, Vol. 32 No. 2, pp. 169-197, doi: 10.1177/0265532214554321.

[23] Lan, C. and Fan, S. (2019), "Developing classroom-based language assessment literacy for in- service EFL teachers: the gaps", *Studies In Educational Evaluation*, Vol. 61, pp. 112-122, doi: 10.1016/ j.stueduc.2019.03.003.

[24] Lee, H. and Kim, S. (2020), "Enhancing test item creation using AI: a case study in higher education", *International Journal of Artificial Intelligence in Education*, Vol. 30 No. 4, pp. 607-625.

[25] Lincoln, Y.S. and Guba, E.G. (1988), "Criteria for assessing naturalistic inquiries as reports".

[26] McCarthy, J., Minsky, M.L., Rochester, N. and Shannon, C.E. (2006), "A proposal for the dartmouth summer research project on artificial intelligence, August 31, 1955", *AI Magazine*, Vol. 27 No. 4, p. 12, doi: 10.1609/aimag.v27i4.1904.