Beyond One-Size-Fits-All: An AI-Driven Approach for Personalized Quizzes Using Clustering and ChatGPT

Liyuan Liu[1,*], Meng Han[2], Seyedamin Pouriyeh[3], Yiyun Zhou[4], and Nasrin Dehbozorgi[5]

[1]Department of Decision and System Sciences, Saint Joseph's University, United States
[2]Intelligent Fusion Research Center, Zhejiang University, China
[3]Department of Information Technology, Kennesaw State University, United States
[4]The National Board of Medical Examiners, United States
[5]Department of Software Engineering, Kennesaw State University, United States
Email: lliu@sju.edu (L.L.); mhan@zju.edu.cn (M.H.); spouriye@kennesaw.edu (S.P.); yyzhou@nbme.org (Y.Z.); dnasrin@kennesaw.edu (N.D.)
*Corresponding author

Manuscript received March 11, 2025; revised April 27, 2025; accepted May 23, 2025; published September 8, 2025

***Abstract*—With rapid advancements in Generative AI (GenAI), educators have the opportunity to enhance student engagement and learning through personalized quizzes. Despite their potential, the adoption of customized learning assessments remains limited due to challenges in student grouping, difficulty calibration, and content fairness. This study proposes a structured, three-step framework leveraging AI to address these issues. Firstly, diverse student data—including academic performance, behavior, interaction with learning materials, demographics, psychological attributes, and feedback—is aggregated and normalized into multi-dimensional vectors. K-means clustering with Euclidean distance is then applied. Secondly, detailed profiles are created for each cluster by calculating their centroid, reflecting the unique characteristics and preferences of students. Finally, these profiles guide a GenAI system-ChatGPT to generate personalized quiz questions relevant to each group's learning style and field of study. Implementing this approach with 105 business statistics students at a university in the USA, statistical tests demonstrated significant improvement in student performance on customized quizzes compared to traditional assessments. The findings underscore the transformative potential of AI-driven personalization in educational settings, promoting more effective, tailored learning experiences.**

***Keywords*—artificial intelligence, generative artificial intelligence, customized learning, students clustering**

I. Introduction

Personalized learning is the educational experiences to meet the unique needs of students, adapting content, pacing, and instructional methods to align with student's individual preferences and abilities [1]. According to a study by RAND Corporation, personalized learning environments can result in about 3 percentile points gain in both mathematics and reading for students, demonstrating significant academic improvement [2]. Customized quizzes and exams represent a crucial component of personalized learning. Prior research has demonstrated that adaptive quizzes enhance student motivation and engagement. Furthermore, students often report that such quizzes effectively support their learning [3]. Unlike traditional "one size fits all" quizzes/exams that uniformly challenge all students with the same set of questions, customized quizzes/exams adapt to students' individual learning profiles. For example, in an undergraduate-level statistics class, customized quizzes adapt to each student's interests and future job plans; for instance, a student aiming for a career in healthcare might receive questions on biostatistics involving medical data analysis, while another planning to enter the finance sector would tackle statistics problems related to market trends and financial forecasting. Building on these personalized approaches, the advent of Generative AI (GenAI) offers new possibilities to further tailor educational experiences with greater efficiency and scalability.

With the advent of GenAI, there is a growing consensus among scholars [4–6] that these technologies can significantly enhance the development of personalized assessment materials. By automatically generating test questions tailored to the knowledge level and learning preferences of individual students, GenAI reduces the workload for educators and facilitates a more scalable and efficient implementation of personalized learning strategies in diverse educational environments. It can enhance student engagement, boost retention of course material, and increase learning efficiency [7]. However, despite the well-documented benefits of customized quizzes and exams, the adoption of GenAI to personalize students' learning materials remains relatively low. According to a 2023 report by Turnitin [8], over 75% of faculty members reported that they do not regularly use AI in their teaching. Fig. 1 illustrates the various ways educators currently use AI tools in education from their study, highlighting that most applications focus on prompt tuning and training students in AI tool use rather than personalized assessment design. Furthermore, most faculty members primarily utilize AI to check what students can see with GenAI platforms, and more than one-third of the faculty employ GenAI to educate students on the effective use of AI technologies.

To fully realize the potential of GenAI in education, it is beneficial to integrate it with data-driven methodologies such as clustering algorithms. For instance, Sharif *et al.* [9] demonstrated the application of clustering algorithms to improve personalized learning through recommendation models, emphasizing the role of data-driven approaches in tailoring educational content. Similarly, Salles *et al.* [10] introduced Interpret3C, a novel clustering pipeline that incorporates interpretable neural networks to better understand and adapt to diverse student needs in large-scale online environments. In the GenAI aspect, Oye [11] evaluated the impact of GenAI on personalized learning outcomes, highlighting its benefits in creating engaging and inclusive learning environments while also discussing challenges such as data privacy and the need for adequate teacher training. Additionally, Maity and Deroy [12] explored the integration of GenAI into Intelligent Tutoring Systems, focusing on dynamic content generation and adaptive learning pathways to enhance personalized education. Despite these promising developments, there remains a significant research gap in the practical implementation of integrated GenAI and clustering algorithm frameworks within real-world classroom settings. Most existing studies have been conducted in controlled environments or focus on theoretical models, lacking empirical validation in diverse educational contexts. Furthermore, challenges such as ensuring fairness, maintaining consistent difficulty levels across personalized assessments, and addressing the technical expertise required for implementation have not been thoroughly addressed. There are also several challenges to the widespread use of GenAI for creating customized quizzes/exams in higher education institutions. Implementing such technology requires a certain level of technical expertise, which many educators may lack. Additionally, there are concerns about the accuracy of AI generated questions, many existing GenAI tools are not specifically designed for educational purposes or are difficult to customize for specific educational needs and contexts, and there is still a pressing need to provide GenAI training for all educators. Therefore, it is essential to design a framework that can be easily adopted by all higher education institutions. To address these challenges and validate the efficacy of our innovative framework, we pose the following research questions:

- How can integrating GenAI (ChatGPT) with traditional AI algorithms (Clustering) enhance personalized assessments in higher education?
- Does the proposed AI-driven personalized assessment framework improve student academic performance compared to traditional quizzes?
- How can the proposed AI-driven personalized assessment framework be generalized and simplified to facilitate widespread adoption by educators without extensive technical expertise?

In this study, we developed an AI-driven approach to create personalized quizzes and exams tailored to students' interests and profiles. This approach integrates traditional AI algorithms and the GenAI platform, specifically ChatGPT. The key contributions of this study are as follows:

- We introduced a novel framework integrating clustering algorithms with GenAI, specifically designed for higher education settings. Technically, our framework systematically employs k-means clustering combined with a clearly defined Elbow method to determine the optimal number of student clusters. This technical clarity facilitates easy implementation by educators, even those without extensive AI expertise, bridging advanced analytical methods with practical educational requirements.
- Leveraging a robust data-driven methodology, our approach technically utilizes multi-dimensional student data vectors—comprising academic performance, interaction patterns, behavioral metrics, demographic factors, psychological attributes, and feedback—to categorize students into distinct clusters. Each group's centroid is computationally derived and profiled, ensuring precision in capturing diverse learning preferences, thereby enabling accurate customization of educational content.
- Employing ChatGPT for dynamic quiz generation, our method technically ensures that each personalized assessment maintains a consistent difficulty level across different clusters. By integrating clearly defined prompts reflecting each group's educational background and learning objectives, our technical design guarantees fairness and balanced challenge, addressing a commonly overlooked aspect of personalized education.
- The empirical validation of our technical framework in a practical undergraduate classroom setting provides robust evidence of its effectiveness. Statistically significant improvements in student performance on customized quizzes, compared to traditional assessments, confirm the technical soundness and real-world applicability of our integrated clustering and GenAI methodology, highlighting its potential to substantially enhance educational outcomes.

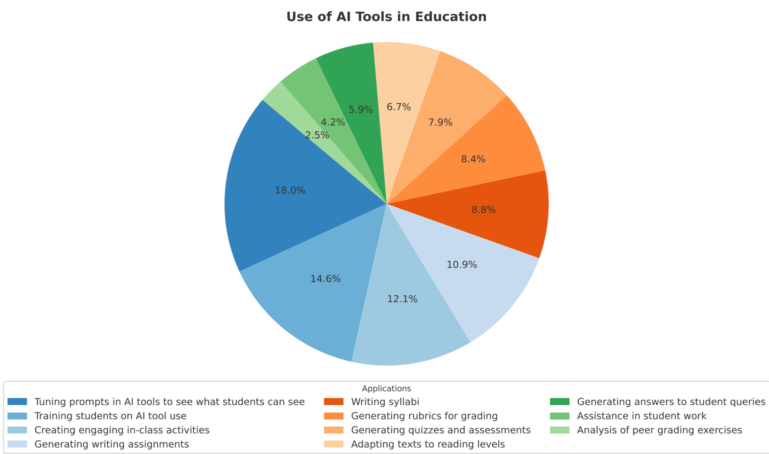

Fig. 1. Use of AI tools in education.

II. Literature Review

A. Personalized Learning and Its AI-Driven Foundations

Personalized learning has gained significant attention due to its capacity to cater to students' unique learning preferences, abilities, and experiences. Tenon *et al.* [13] demonstrated that personalized approaches not only enhance students' critical thinking and creativity but also better prepare them for the job market. Similarly, Rafiqovna [14] noted that customizing learning materials based on a student's personal experiences, knowledge, and learning habits can significantly improve their learning efficiency and speed. There are many contents can be personalized in learning. Heng *et al*. [15] suggested that develop the systems like Learning Management System (LMS) can recommend specific materials tailored to the learner's profile, including their prior knowledge and preferred format (e.g., text, video, or interactive content). Zhong [16] shown that personalized learning paths allow students to learn at their own pace and follow unique trajectories. Abedi *et al*. [17] advocated for the integration of customized assignments in higher education, by enabling students to select topics or project formats that align with their interests and strengths. They are not only encouraged personalizes learning but is also strongly recommended for its potential to enhance educational outcomes.

From a conceptual standpoint, personalized learning fundamentally aligns with the theories of constructivism and learner-centered pedagogy, emphasizing active student involvement and tailored educational experiences [18]. According to constructivist principles, Leeuwen *et al*. [19] illustrated that knowledge construction occurs most effectively when learners engage actively with content adapted to their individual learning processes and previous experiences. The rapid advancements in educational technology, particularly AI, have dramatically expanded the opportunities for practical implementation of these pedagogical theories.

AI-driven personalized learning leverages sophisticated algorithms that analyze large volumes of educational data to dynamically tailor learning experiences. Zawacki-Richter *et al*. [20] have highlighted in their study that AI's potential to significantly transform educational environments by enabling scalable, adaptive, and highly individualized instructional approaches. Liu *et al*. highlighted the value of using AI-driven approaches to customize students' learning experiences, demonstrating that tailoring support based on mid-term performance and learning behaviors can more effectively guide students toward academic success [21]. Holmes *et al*. stated that AI systems can continuously monitor students' academic performance, adapt content difficulty in real-time, and recommend resources that address individual learner needs and optimize educational outcomes [22]. Baker *et al*. [23] pointed the importance of ensuring that personalized learning systems powered by AI maintain fairness and transparency, avoiding potential biases that could disadvantage particular student groups.

Despite these significant advancements and conceptual alignments, practical, scalable frameworks that integrate personalized learning methods using AI—particularly in the form of personalized quizzes and assessments—remain limited. Many existing systems require substantial technical expertise, limiting their accessibility for broader educational settings and educators without extensive technological skills.

B. AI-Based Clustering in Personalized Learning

AI clustering algorithms have become instrumental in advancing personalized learning by enabling the grouping of students based on various academic and behavioral characteristics. These unsupervised machine learning techniques, such as k-means and hierarchical clustering, identify natural groupings within educational data, facilitating tailored instructional strategies.

Recent studies have demonstrated the efficacy of AI clustering in educational contexts. For instance, Park [24] developed a recommender system utilizing collaborative filtering, effectively clustering students based on past performance to generate personalized practice assessments. Du *et al*. [25] introduced a reinforcement learning-based examination system that leverages clustering algorithms to dynamically adapt learning paths, continuously optimizing the personalized assessment process. Mihaescu *et al*. [26] employed Singular Value Decomposition (SVD)-based clustering techniques to adjust quiz difficulty according to individual student performance, enhancing assessment personalization. Additionally, Almazroi *et al*. [27] utilized clustering methodologies to personalize assessments based on students' personality traits derived from their social media interactions.

Beyond these applications, clustering algorithms have been used to analyze student engagement patterns. A study by Xu *et al*. [28] applied AI optimization algorithms in higher education management and personalized teaching, demonstrating significant improvements in students' learning outcomes, engagement, satisfaction, and efficiency when using AI-driven personalized teaching compared to traditional approaches. Similarly, research by Zheng *et al*. [29] explored the impacts of AI empowerment on students' personalized learning, highlighting the role of AI in enhancing learning effectiveness through intelligent recommendation and automated feedback.

Despite these advancements, challenges remain in the practical implementation of AI clustering in personalized learning. Issues such as data privacy, algorithmic bias, and the need for human-AI interaction must be addressed to ensure equitable and effective educational outcomes. Moreover, the development of scalable frameworks that can be easily adopted by educators without extensive technical expertise is crucial for the broader application of AI clustering in personalized learning environments.

C. GenAI in Personalized Education

In recent years, GenAI, particularly Large Language Models (LLMs) such as ChatGPT, has further revolutionized the educational domain by automating the generation of personalized educational content. These AI systems are designed to process and generate language-based outputs that are tailored to the specific needs and queries of learners, making personalized learning more dynamic and accessible.

The application of GenAI in education is grounded in the principles of adaptive learning and constructivism, which advocate for educational experiences that adapt to the learning pace and style of individual students. By leveraging GenAI, educational platforms can dynamically generate content that adjusts to the evolving understanding and interest levels of each student, thereby supporting a more personalized learning journey. Shemshack *et al*. [30] emphasized that AI integration, particularly within big data contexts, significantly enhances the scalability and efficiency of personalized learning. These systems can analyze vast amounts of educational data to provide insights that help tailor the learning process to individual needs more effectively. Liu *et al.* [31, 32] introduced innovative frameworks utilizing generative artificial intelligence to enhance syllabus clarity and curriculum alignment in higher education, demonstrating that AI-driven analysis of educational content can significantly improve instructional efficacy, course preparation consistency, and collaborative learning environments. VillegasCh *et al.* [33] offered practical frameworks for AI-driven personalized educational systems, showcasing how GenAI can be integrated into existing educational infrastructures to support a range of personalized learning activities, from adaptive course content to automated feedback systems. Similarly, Pataranutaporn *et al.* [34] and Tselepatiotis *et al.* [35] utilized AI-driven methods to dynamically adjust educational content based on learners' skills and interests. These studies demonstrate the capability of GenAI to not only understand and generate content but also to anticipate and react to the needs of learners in real-time.

Despite these advancements, research specifically focused on integrating GenAI into personalized assessment remains relatively limited. Pesovski *et al.* [36] introduced generative AI methods within a Learning Management System (LMS) to automate customized quiz generation. However, their implementation was constrained to small-scale scenarios and lacked scalability for broader educational use. This highlights a significant gap in the current research—developing scalable, effective GenAI solutions that can be widely implemented across various educational settings without extensive customization. There is a growing need to explore how GenAI can be further developed to support large-scale educational applications without losing the personal touch that characterizes effective teaching. Future research should focus on overcoming the scalability challenges and ensuring that GenAI applications in education remain inclusive, equitable, and capable of providing high-quality, personalized educational experiences to a broad spectrum of learners.

III. METHODS

In this study, we developed a three-step framework utilizing AI and GenAI to deliver personalized quizzes and exams, aimed at enhancing student engagement and improving their learning experience. As depicted in Fig. 2, the initial phase of Dynamic Student Clustering involves educators aggregating and analyzing a wide range of student data to inform the clustering process. The types of data include, but are not limited to: (1) Academic Performance Data (x_i^{acad}): test scores, and progress indicators, etc. (2) Behavioral Data (x_i^{behav}): Data on participation in class activities, discussion forums, class attendance, or group projects. (3) Interaction Data (x_i^{interact}): Students' responses to Previous quizzes or exams, and data collected from learning management systems (LMS) on course material accesses, submissions, and interactions. (4) Demographic Information (x_i^{demo}): Students' age, gender, socioeconomic status and their cultural or linguistic background. (4) Psychological Factors (x_i^{psych}): Students' motivation levels or personality that can be collected from some psychological assessment. (6) Feedback and Self-Assessments (x_i^{feedback}): Information from surveys about students' personal educational challenges and objectives, or insights into what students feel about their educational experiences. Each student i is represented by a multi-dimensional data vector composed of the normalized values from each category, ready for analysis using K-means for clustering:

$$v_i = \left(x_i^{\text{acad}}, x_i^{\text{behav}}, x_i^{\text{interact}}, x_i^{\text{demo}}, x_i^{\text{psych}}, x_i^{\text{feedback}}\right)$$

The Euclidean distance between any two students i and j, represented by their vectors v_i and v_j, is calculated as follows:

$$d(v_i, v_j) = \sqrt{\sum_{k=1}^{n} \left(x_{i,k} - x_{j,k}\right)^2}$$

where $x_{i,k}$ and $x_{j,k}$ are the components of the vectors v_i and v_j respectively. This distance measure is used in the K-means clustering algorithm to assign each student to the closest cluster center.

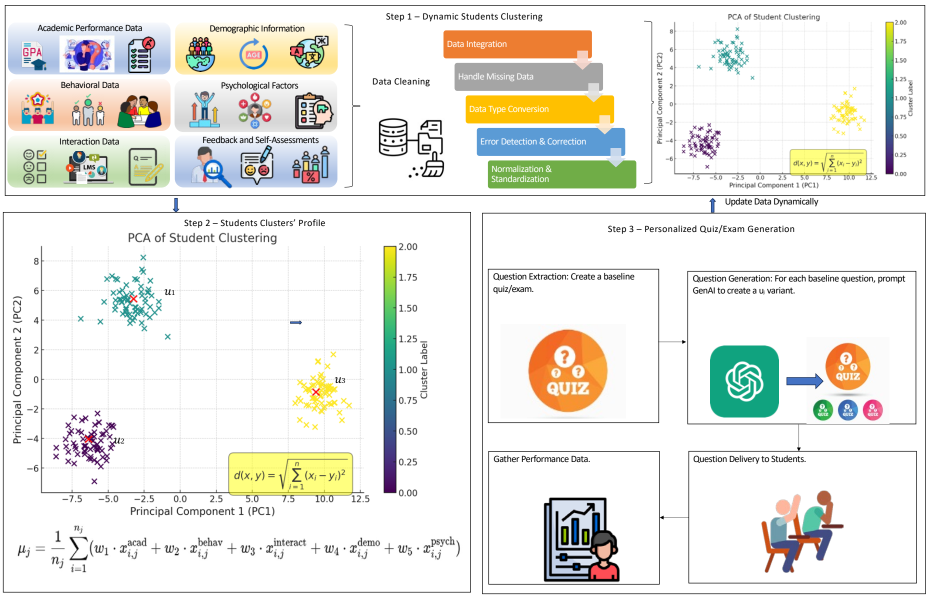

Fig. 2. Framework overview.

Algorithm 1 Elbow Method for Optimal k in K-mean Clustering

Given: Data vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$
for $k = k_{\min}$ **to** $k_{\max}$ **do**
  Initialize k cluster centers randomly
K-means objective:
$$\min_{C_j, u_j} \sum_{i=1}^{k} \sum_{v_j \epsilon C_j} \left\| v_j - u_j \right\|^2$$
Compute loss (SSE):
$$SSE_k = \sum_{i=1}^{k} \sum_{v_j \epsilon C_j} \left\| v_j - u_j \right\|^2$$
end for
Plot (k, SSE_k) and choose k^* at the "elbow."

The challenge in K-means is to find the best number of clusters k. In this study, we utilize the "Elbow" method to determine the optimal k, as depicted in Algorithm 1, which illustrates the process of finding the best k for K-means clustering. The output of Step 1 in our framework is each student's data, annotated with their respective cluster labels. Following the clustering of student data, the second stage involves defining and refining the profiles for each identified student cluster. These profiles, denoted as u_1, u_2, and u_n, represent the centroids of the clusters and encapsulate the unique characteristics and learning preferences of the students within each group. These profiles are critical for the subsequent personalization of educational content. The centroid of each cluster is calculated as the mean of all data vectors belonging to that cluster. Mathematically, the centroid u_j for each cluster j is defined as:

$$u_j = \frac{1}{N_j} \sum_{i=1}^{N_j} v_{i,j}$$

where N_j is the number of students in cluster j, and $v_{i,j}$ is the data vector of the *i-th* student in cluster j. This average is computed across all dimensions represented in the data vectors, thus providing a multi-dimensional mean that characterizes the typical attributes of the cluster. The centroids $u_1, u_2, \ldots, u_n$, therefore reflect the aggregated characteristics of their respective clusters. The final step in our framework utilizes the profiles from Step 2 to generate personalized quizzes and exams. Initially, a baseline quiz or exam is established with standardized questions. An example of a baseline question in statistics might be: "Calculate the mean and standard deviation for the following data set: 5, 20, 40, 65, 90." To adapt this question for students in Cluster 1, who major in finance and prefer active learning, we would use a Generative AI tool with the following prompt:

"Cluster 1 consists of finance majors adept in quantitative analysis, now seeking to enhance their real-world application skills. For this cluster, create a quiz question that revises the baseline statistical task but includes a novel dataset reflective of real financial scenarios they might face in their careers. The task should involve calculating the mean and standard deviation, similar to the baseline question, which asks students to analyze the data set: 5, 20, 40, 65, 90. Adapt this task by using a different set of data points relevant to financial risk analysis or investment portfolio management."

Following the generation of these tailored questions, the customized quizzes or exams are administered to the appropriate student clusters. Detailed performance data, such as scores and time to completion, is collected to evaluate the effectiveness of the personalized educational content and further refine teaching strategies. The students' data will be dynamically updated from Step 1.

IV. Results and Discussion

A. Dataset Description

In the Fall of 2024 and Spring of 2025, a dataset was compiled from a Business Statistics course, a core requirement for all undergraduate business students at a university in Philadelphia, USA. The course enrolled 105 freshmen, sophomores, juniors and seniors from various majors including Finance, Accounting, Marketing, and Business Intelligence, among others. The diversity of the student body was reflected in the wide range of business disciplines represented, with the most common being Finance and Accounting. In this study, we employed a convenience sampling technique due to practical considerations. Specifically, we did not actively select or randomize samples but utilized naturally occurring class groupings. The observations are gathered in a naturally occurring educational setting, reflecting "real-world" learning conditions. Because the university applies uniform admission standards, both cohorts are expected to possess comparable academic preparation, a premise confirmed by statistical tests showing no significant differences in GPA or class distribution between the two groups in Section IV B. The experimental class comprised students who enrolled in the Business Statistics course during Fall 2024 (n = 70), whereas the control class included students who enrolled in the same course during Spring 2025 (n = 35). Both classes were taught by the same instructor, thereby controlling for instructional consistency.

The dataset includes scores from four quizzes, with uniquely customized to align with each backgroud and historical performance, and the one-size-fits-all quizzes. To ensure uniform difficulty across all quizzes, we selected an equal number of 'hard,' 'easy,' and 'medium' questions for each quiz. The difficulty levels are defined by the Connect platform from McGraw-Hill [37]. Connect is a widely adopted digital learning and assessment platform in higher education. Connect's item bank is developed and reviewed by subject matter experts to ensure alignment with educational standards and learning objectives, thereby supporting content validity. While McGraw-Hill does not publicly provide specific reliability coefficients such as Cronbach's alpha for its assessments, the platform's design incorporates best practices in educational measurement to promote reliable and valid evaluations of student performance. The average score on this customized quiz was 98.20%, slightly higher than the average score of 97.32% for the non-customized quizzes, suggesting a potential benefit of personalized educational approaches in enhancing student learning outcomes. All student data in this study have been fully anonymized to ensure confidentiality and compliance with privacy standards. Identifiable information has been removed, and only academic performance and broad major categories are included in this case.

B. Experiment Results

Before we employed k-means to cluster students into different groups, we first evaluate the baseline equivalence of the two cohorts in terms of their academic backgrounds, we conducted two statistical tests: Welch's t-test for GPA comparison, and the Chi-square test for class-standing composition. For GPA, the null hypothesis H_0 was that the mean GPA of the experimental group (μ_{exp}) equals the mean GPA of the control group (μ_{ctrl}).

$$H_0: \mu_{\text{exp}} = \mu_{\text{ctrl}}$$
$$H_a: \mu_{\text{exp}} \neq \mu_{\text{ctrl}}$$

Welch's t-test formula used was:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $\overline{x_1}$, $\overline{x_2}$ are the sample means; s12, s_1^2, s_2^2 are the sample variances; and n_1, n_2 are the sample sizes. The GPA comparison yielded a non-significant p-value of 0.063 as Fig. 3 shows, indicating no statistically significant difference between the two groups' mean GPA.

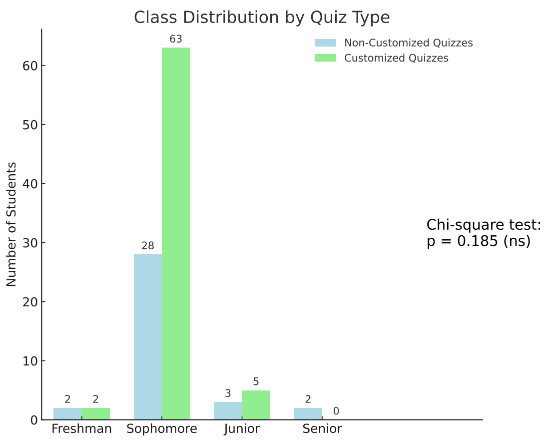

Fig. 3. Class distribution between experimental and control groups.

For class-standing, the Chi-square test (χ^2) was applied with the null hypothesis (H_0) that the class-standing composition distributions are equal across both groups:

H_0:The distributions of class standing are equal
H_a:The distributions of class standing are not equal

The Chi-square formula is given by:

$$\chi^2 = \sum \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

where O_{ij} is the observed frequency, and E_{ij} is the expected frequency for each category. This analysis also resulted in a non-significant p-value of 0.185 as Fig. 4 shows, confirming that no significant differences existed between groups regarding class-standing composition. Thus, the baseline equivalence of both GPA and class-standing was statistically supported.

In the initial phase of our experimental framework, we utilized students' majors and previous academic performance data to cluster them into distinct groups. The scope of data dimensions was intentionally limited due to compliance requirements with human subject data usage policies, which we anticipate expanding upon receiving the necessary approvals. To determine the optimal number of clusters, we employ the Elbow Method. It gauges how much additional explanatory power is gained by adding more clusters. It plots the within-cluster sum of squares (WCSS) on the y-axis against the number of clusters k on the x-axis. As k increases, WCSS naturally falls because each cluster becomes tighter; however, the rate of improvement drops after a certain point. The "elbow" is the value of k where this decline first begins to flatten out—signifying that further splits yield only marginal benefit. In our analysis (Fig. 5), that inflection occurs at $k = 16$, so we selected 16 clusters as the best trade-off between compactness and model simplicity. After that, we run a K-means algoritm to cluster the students' data into 16 groups, as Fig. 6 shows. In this study, majors are the primary criteria for customizing student quizzes. We present average scores by major to compare performance on customized versus non-customized quizzes. Fig. 7 displays the results for the top five majors with the highest enrollment in the dataset. Fig. 7 shows the students generally perform better on customized quizzes compared to non-customized ones across all examined majors. Notably, the most significant improvement is observed in Business Intelligence & Analytics, where the difference in average scores reaches 6.67 points, suggesting a particularly effective customization of quizzes in engaging students and enhancing their performance in this field. In contrast, Sports Marketing shows the smallest improvement with a difference of 1.90 points, indicating a less pronounced but still positive effect of quiz customization. These findings underline the potential benefits of tailoring educational content to the specific academic contexts and learning preferences of different student groups.

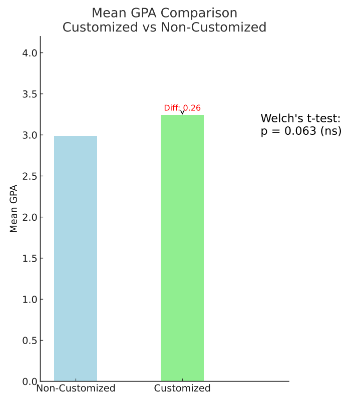

Fig. 4. Students GPA between experimental and control groups.

Building on these results, our findings both resonate with and extend recent work in adaptive learning and personalized assessment. For instance, Hwang *et al.* [38] discussed the integration of AI in education, highlighting its role in facilitating personalized learning experiences. Critically, while prior research has largely focused on content sequencing or feedback timing, our work is among the first to integrate multi-dimensional AI algorithms, such as k-means clustering with GenAI-driven question generation. Rouzegar and Makrehchi's comparison of GPT-3.5 versus GPT-4 for crafting customized test items [39] validated that advanced LLM prompts can produce high-quality, contextually relevant questions when seeded with expert templates—an insight we applied to our domain-specific prompts. Samala *et al.*'s comprehensive taxonomy of generative AI applications in education [40] underscores the pedagogical potential of prompt engineering, which we harnessed to target core analytical skills in each cluster.

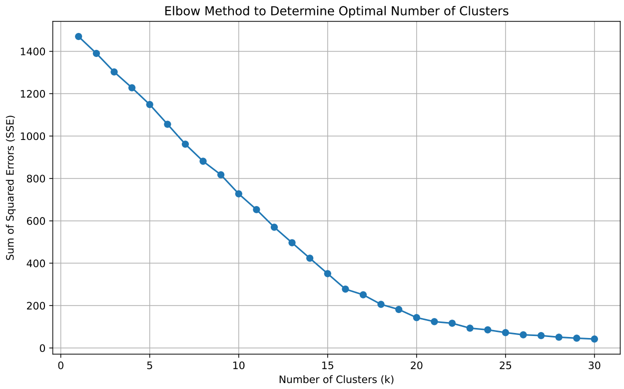

Fig. 5. Elbow method to determine optimal number of clusters.

Fig. 6. Visualization of students clustering.

In Step 3, we employ GenAI tools, specifically ChatGPT, to generate customized quizzes tailored to the different clusters identified in the study. To ensure fairness and provide an equitable learning environment, we defined a baseline quiz by selecting questions from the question pool available on McGraw-Hill's Connect platform. The profiles of each cluster were derived based on the K-means clustering results following the steps outlined in Section III. Using these profiles, quizzes were customized with ChatGPT to align with the specific characteristics of each cluster. For example, Cluster 7 comprises students majoring in Business Intelligence & Analytics who demonstrate strong quiz performance, moderate midterm scores, and high engagement in class practice, representing 85.7% of all the Business Intelligence & Analytics major students. Below is the prompt that we used to generate the quiz questions for Cluster 7 students:

Cluster 7 consists of Business Intelligence & Analytics majors who exhibit strong quiz performance, moderate midterm scores, and high engagement in class practice. Using the following baseline question, generate a similar question that incorporates statistical analysis relevant to real-world business scenarios and challenges students' quantitative skills:

Baseline Question: "A simple random sample of 20 observations is derived from a normally distributed population with a population standard deviation of 3.2. Please construct the 95% confidence interval." And here is the generated question: A business intelligence analyst at a retail company wants to examine the average customer spend during holiday seasons to better forecast sales and manage inventory. From previous years' data, the standard deviation of customer spend is known to be \$120. A random sample of 60 transactions from the last holiday season shows an average spend of \$300, calculate a 95% confidence interval for the true average customer spend during holiday seasons.

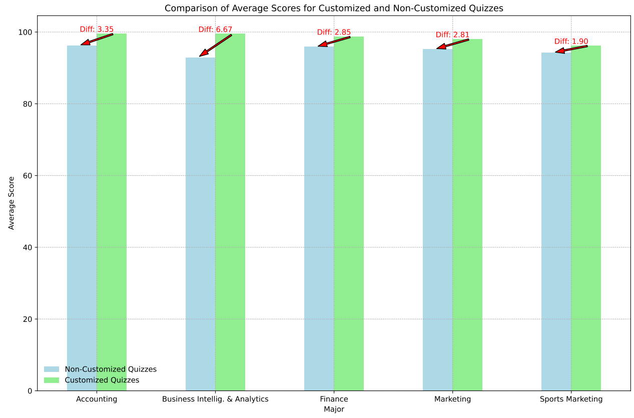

Fig. 7. Comparison of average scores for customized and non-customized quizzes.

To evaluate whether customized quizzes significantly improve student performance, we conducted two complementary statistical analyses. First, a paired sample *t*-test was performed using data from a single cohort of 70 students enrolled in Fall 2024. This within-subject analysis compared each student's performance on non-customized and customized versions of the same quiz, allowing us to isolate the impact of quiz customization while controlling for individual differences. Second, we conducted a Welch's *t*-test to compare quiz scores between two independent groups (105 students): a control group that received non-customized quizzes and an experimental group that received customized quizzes. This between-group analysis accounted for unequal variances and sample sizes, offering additional validation of the effect of quiz customization on student outcomes. Together, these analyses provide robust evidence for the effectiveness of personalized quiz design in enhancing academic performance.

For the paired sample t test, we propose the following hypotheses for our one-tailed test:

$$H_0\text{: } \mu_{\text{Customized}} \leq \mu_{\text{Non-Customized}}$$

Null Hypothesis: The mean score of the customized quizzes is less than or equal to the average score of the non-customized quizzes.

$$H_1\text{: } \mu_{\text{Customized}} > \mu_{\text{Non-Customized}}$$

Alternative Hypothesis: The mean score of the customized quizzes is greater than the average score of the non-customized quizzes.

The test statistics in the paired sample t test is:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

where $\bar{d}$ is the mean of the differences between paired observations, s_d is the standard deviation of the differences, and n is the number of paired observations.

A paired sample t-test was employed to compare the mean scores: (1) The t-statistic was calculated based on the differences between the scores of customized quiz and the non-customized quiz for each student. (2) The resulting t-statistic was 3.1692 with a p-value of 0.0023, indicating a significant difference at the 0.05 significance level.

To evaluate whether customized quizzes improve student performance, we conducted a one-tailed independent samples *t*-test using Welch's correction for unequal variances. The hypotheses were defined as follows:

- **Null hypothesis (H_0):** There is no difference or the customized quizzes do not lead to higher scores.

$$H_0: \mu_{\text{customized}} \leq \mu_{\text{non-customized}}$$

- **Alternative hypothesis (H_a):** Students who received customized quizzes achieved significantly higher scores than those who received non-customized quizzes.

$$H_a: \mu_{\text{customized}} > \mu_{\text{non-customized}}$$

The test statistics is calculated by:

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

And the Degrees of Freedom of (Welch's Test) is calculated by:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

where $\overline{X_1}$, X_2 are the sample means, s_1^2, s_2^2 are the sample variances, n_1 and n_2 are the sample sizes of group 1 and group 2 respectively.

The results of the Welch's *t*-test indicated a statistically significant difference in quiz performance between the customized and non-customized groups. Specifically, the test yielded a *t*-value of -2.217 with a corresponding *p*-value of 0.032, which is below the conventional significance threshold of 0.05. This result suggests that students who received customized quizzes scored significantly higher than those who received non-customized quizzes, providing empirical support for the effectiveness of quiz customization in enhancing academic performance. These results align closely with prior research showing that personalization increases student engagement and effort. Whereas earlier studies relied on manual content tagging or rule-based adaptation, our study confirms that automated, AI-driven customization yields similarly robust gains in performance. By leveraging GenAI for quiz item generation within a k-means clustering framework, we eliminate the technical complexity and time-intensive authoring processes that have constrained earlier implementations. As a result, our solution not only upholds the established link between personalization and engagement but also offers a fully scalable platform that instructors can deploy without specialized programming skills—paving the way for widespread, sustainable adoption of adaptive assessments in future courses.

V. Conclusion

This research has successfully demonstrated how integrating Generative AI, specifically ChatGPT, with traditional AI clustering techniques can dramatically enhance personalized assessments in higher education. Our findings indicate that this sophisticated AI-driven framework significantly boosts student academic performance, showcasing the clear advantages of tailored quizzes over traditional methods. We introduced a groundbreaking framework designed specifically for the higher education context, which merges clustering algorithms with GenAI technologies. This novel framework not only simplifies the adoption of personalized learning methodologies but also significantly expands their applicability across a diverse range of educational institutions. Importantly, it is crafted to be user-friendly, requiring minimal technical expertise, which facilitates its adoption by educators across disciplines, enabling its practical deployment even by those with limited tech experience. By employing clustering algorithms, our study has pushed the boundaries of educational data analysis, enabling the segmentation of students into distinct groups. This segmentation is based on a thorough analysis of their academic performance and learning behaviors, which allows for the highly specific customization of educational content. Additionally, using ChatGPT to generate customized quizzes ensures that assessments are not only personalized but also maintain a consistent level of challenge across different student groups. Implemented in an actual classroom setting with undergraduate students, the effectiveness of this framework was thoroughly validated. The results clearly demonstrated improvements in student performance, offering solid evidence of the framework's efficacy. These results not only highlight the practical viability of our approach but also illustrate its potential to significantly improve educational outcomes. In conclusion, this study represents a significant advancement in educational technology, setting a new standard for personalized education. It provides a robust foundation for future innovations in the field, suggesting promising directions for further research and development.

Moving forward, we plan to develop an application tool that will allow educators to easily create customized quizzes. This application will be straightforward to use, with the complex algorithms operating seamlessly behind the scenes. Our focus will be on ensuring that the quizzes are fair and adapted to diverse learning needs, incorporating insights from educational psychology to address fairness and bias. This tool aims to simplify the personalization of learning, making effective and equitable educational practices more accessible to all faculty members and protect students' private data at the same time.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

L.L. was primarily responsible for writing the manuscript and collecting the data. M.H. and Y.Z. supervised the study, provided technical research ideas, and guided the experimental design. N.D. and S.P. contributed to editing the manuscript and offered fundamental research ideas for the experiments. All authors reviewed and approved the final version of the manuscript.

REFERENCES

[1] G. S. Partnership, "What is personalized learning?" *Glossary of Education Reform*, 2020.
[2] R. Corporation, "How does personalized learning affect student achievement?" RAND Reports, 2015.
[3] B. Ross, A.-M. Chase, D. Robbie, G. Oates, and Y. Absalom, "Adaptive quizzes to increase motivation, engagement and learning outcomes in a accounting unit," *International Journal of Educational Technology in Higher Education*, vol. 15, no. 1, pp. 1–14, 2018.
[4] M. Guettala, S. Bourekkache, O. Kazar, S. Harous *et al.*, "Generative artificial intelligence in education: Advancing adaptive and personalized learning," *Acta Informatica Pragensia*, vol. 13, no. 3, pp. 460–489, 2024.
[5] P. K. Dubey, "Personalized learning and pedagogy: Harnessing AI for tailored education," *Integrating AI Into Pedagogical Education*, IGI Global Scientific Publishing, 2025, pp. 81–102.
[6] A. D. Samala, S. Rawas, T. Wang, J. M. Reed, J. Kim, N.-J. Howard, and M. Ertz, "Unveiling the landscape of generative artificial intelligence in education: A comprehensive taxonomy of applications, challenges, and future prospects," *Education and Information Technologies*, pp. 1–40, 2024.
[7] H. Rouzegar and M. Makrehchi, "Generative AI for enhancing active learning in education: A comparative study of GPT-3.5 and GPT-4 in crafting customized test questions," arXiv preprint arXiv:2406.13903, 2024.
[8] C. Shaw, L. Yuan, D. Brennan, S. Martin, N. Janson, K. Fox, and G. Bryant. (October 2023). Generative AI in higher education: Fall 2023 update of time for class study. [Online]. Available: https://tytonpartners.com/time-for-class-2023/GenAI-Update
[9] S. Sharif, M. Theeng, and D. Mani, "Application of clustering algorithms to enhance personalized learning through recommendation model," in *Proc. the 2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 2024.
[10] I. Salles, P. Mejia-Domenzain, V. Swamy, J. Blackwell, and T. Käser, "Interpret3C: Interpretable student clustering through individualized feature selection," in *Proc. the International Conference on Artificial Intelligence in Education*, pp. 382–390, Cham: Springer Nature Switzerland, 2024.
[11] E. Oye, *Evaluating the Impact of Generative AI on Personalized Learning Outcomes*, 2024.
[12] S. Maity and A. Deroy, "Generative AI and its impact on personalized intelligent tutoring systems," arXiv preprint arXiv:2410.10650, 2024.
[13] S. R. Tenon and P. Epler, "Evaluation of principles and best practices in personalized learning," *IGI Global*, 2020.
[14] A. N. Rafiqovna, "The importance of personalized learning technologies and some recommendations," *American Journal of Philological Sciences*, vol. 4, no. 06, pp. 91–93, 2024.
[15] L. Ean Heng, W. Pei Voon, N. A. Jalil, C. Lee Kwun, T. Chee Chieh, and N. Fatiha Subri, "Personalization of learning content in learning management system," in *Proc. the 2021 10th International Conference on Software and Computer Applications*, 2021, pp. 219–223.
[16] L. Zhong, "A systematic review of personalized learning in higher education: Learning content structure, learning materials sequence, and learning readiness support," *Interactive Learning Environments*, vol. 31, no. 10, pp. 7053–7073, 2023.
[17] R. Abedi, M. R. Nili Ahmadabadi, F. Taghiyareh, K. Aliabadi, and S. Pourroustaei Ardakani, "The effects of personalized learning on achieving meaningful learning outcomes," *Interdisciplinary Journal of Virtual Learning in Medical Sciences*, vol. 12, no. 3, pp. 177–187, 2021.
[18] K. C. Li and B. T. M. Wong, "Constructivist learning in higher education supported by learning analytics," *Interactive Learning Environments*, vol. 29, no. 8, pp. 1383–1399, 2021.
[19] A. Leeuwen and J. Janssen, "A systematic review of teacher guidance during collaborative learning in primary and secondary education," *Educational Research Review*, vol. 27, pp. 71–89, 2019.
[20] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education—where are the educators?" *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, p. 39, 2019.
[21] L. Liu and M. Han, "Deciphering the influence of mid-term examinations on student learning outcomes: A comprehensive investigation employing statistical and machine learning approaches," *International Journal of Learning and Teaching*, 2024.
[22] W. Holmes, M. Bialik, and C. Fadel, "Artificial intelligence in education: Promises and implications for teaching and learning," *Journal of Educational Technology & Society*, vol. 25, no. 1, pp. 1–13, 2022.
[23] R. S. Baker and A. Hawn, "Algorithmic bias in education," *International Journal of Artificial Intelligence in Education*, pp. 1–41, 2022.
[24] Y. Park, "Personalized practice exam recommendation for helping students prepare for course assessment," in *Proc. 2022 International Conference on Advanced Learning Technologies (ICALT)*, pp. 390–391, 2022.
[25] Y. Du, W. Song, Y. Coa, X. Chen, and G. Huang, "Design of personalized AI examination system based on reinforcement learning," in *Proc. 2024 6th International Conference on Computer Science and Technologies in Education (CSTE)*, IEEE, 2024, pp. 119–124.
[26] C. Mihaescu, O. M. Teodorescu, P.-S. Popescu, and M. Mocanu, "Learning analytics solution for building personalized quiz sessions," in *Proc. 2017 18th International Carpathian Control Conference (ICCC)*, pp. 140–145, 2017.
[27] A. A. Almazroi, A. M. Idrees *et al.*, "Intelligent framework for enhancing the quality of online exams based on students' personalization," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, 2022.
[28] X. Xu, L. Zhang, J. Wang, and Y. Liu, "AI optimization algorithms enhance higher education management and personalized teaching through empirical analysis," *Scientific Reports*, vol. 15, Nature Publishing Group, 2025, Art. no. 10157.
[29] S. Zheng and M. Han, "The impact of AI enablement on students' personalized learning and countermeasures—A dialectical approach to thinking," *Journal of Infrastructure, Policy and Development*, vol. 8, no. 14, EnPress Publisher, 2024, Art. no. 10274.
[30] A. Shemshack and J. M. Spector, "A systematic literature review of personalized learning terms," *Smart Learning Environments*, vol. 7, no. 1, p. 33, 2020.
[31] L. Liu, R. A. Mendoza, T. R. Martin, and V. M. Miori, "Generative AI-powered educational alignment: A framework for matching syllabus course topics with web description," in *Proc. 2024 9th International Conference on Distance Education and Learning (ICDEL)*, IEEE, 2024, pp. 340–346.
[32] L. Liu, V. Miori, and M. Han, "TopicShare: An AI and ChatGPT-based data sharing framework on teaching content in higher education," in *Proc. 2024 5th International Conference on Information Technology and Education Technology (ITET)*, IEEE, 2024, pp. 56–61.
[33] W. Villegas-Ch, J. García-Ortiz, and S. Sánchez-Viteri, "Personalization of learning: Machine learning models for adapting educational content to individual learning styles," *IEEE Access*, 2024.
[34] P. Pataranutaporn, V. Danry, J. Leong, P. Punpongsanon, D. Novy, P. Maes, and M. Sra, "AI-generated characters for supporting personalized learning and well-being," *Nature Machine Intelligence*, vol. 3, no. 12, pp. 1013–1022, 2021.
[35] M. Tselepatiotis and E. Alepis, "Adaptive learning in mobile serious games: A personalized approach using AI for general knowledge quizzes," in *Proc. 2024 15th International Conference on Information, Intelligence, Systems & Applications (IISA)*, IEEE, 2024, pp. 1–8.
[36] I. Pesovski, R. Santos, R. Henriques, and V. Trajkovik, "Generative AI for customizable learning experiences," *Sustainability*, vol. 16, no. 7, p. 3034, 2024.
[37] McGraw-Hill Education, *Connect Platform: Question Difficulty Classification*, 2025.
[38] G. J. Hwang, H. Xie, B. W. Wah, and D. Gašević, "Vision, challenges, roles and research issues of Artificial Intelligence in Education," *Computers and Education: Artificial Intelligence*, vol. 1, art. 100001, 2020.
[39] H. Rouzegar and M. Makrehchi, "Generative AI for enhancing active learning in education: A comparative study of GPT-3.5 and GPT-4 in crafting customized test questions," arXiv preprint arXiv:2406.13903, 2024.
[40] A. D. Samala, S. Rawas, T. Wang, J. M. Reed, J. Kim, N.-J. Howard, and M. Ertz, "Unveiling the landscape of generative artificial intelligence in education: A comprehensive taxonomy of applications, challenges, and future prospects," *Education and Information Technologies*, vol. 29, pp. 1–40, 2024.